

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
TELECOMUNICAÇÕES**

LAYSA DE FREITAS MARQUES HOLANDA

**ANÁLISE ACÚSTICA VOCAL E CLASSIFICAÇÃO DE VOZES
SAUDÁVEIS E COM PATOLOGIAS LARÍNGEAS**

Fortaleza, Ceará
2021

LAYSA DE FREITAS MARQUES HOLANDA

**ANÁLISE ACÚSTICA VOCAL E CLASSIFICAÇÃO DE VOZES
SAUDÁVEIS E COM PATOLOGIAS LARÍNGEAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Telecomunicações do Instituto Federal de Educação, Ciência e Tecnologia do Ceará como requisito parcial para obtenção do título de mestre em Engenharia de Telecomunicações. Área de concentração: Sistemas de telecomunicações

Orientador: Prof. Dr. Francisco José Alves de Aquino

Fortaleza, Ceará
2021

Dados Internacionais de Catalogação na Publicação Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

H722a Holanda, Laysa de Freitas Marques.
 Análise Acústica Vocal e Classificação de vozes Saudáveis e com Patologias Laríngeas /
 Laysa de Freitas Marques Holanda. - 2021.
 74 f. : il.

 Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Engenharia de
Telecomunicações, Campus Fortaleza, 2021.
 Orientação: Prof. Dr. Francisco José Alves de Aquino.

 1. Sistemas de Telecomunicações. 2. Processamento de Sinais. 3. Reconhecimento de
Padrões. 4. Machine Learning. 5. SVM. I. Título.

CDD 621.382

LAYSA DE FREITAS MARQUES HOLANDA

**ANÁLISE ACÚSTICA VOCAL E CLASSIFICAÇÃO DE VOZES
SAUDÁVEIS E COM PATOLOGIAS LARÍNGEAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Telecomunicações do Instituto Federal de Educação, Ciência e Tecnologia do Ceará como requisito parcial para obtenção do título de mestre em Engenharia de Telecomunicações. Área de concentração: Sistemas de telecomunicações.

Aprovada em 18/10/2021

BANCA EXAMINADORA

Orientador Prof. Dr. Francisco José Alves de Aquino

Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Prof. Dr. Francisco Nivando Bezerra

Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Prof. Dr. Francisco Rafael Vasconcelos Guimarães

Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Prof. Dr. Ricardo Silva The Pontes

Universidade Federal do Ceará - UFC



Documento assinado eletronicamente por **Francisco Jose Alves de Aquino, Professor do Ensino Básico, Técnico e Tecnológico**, em 18/10/2021, às 16:00, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Francisco Rafael Vasconcelos Guimaraes, Professor do Ensino Básico, Técnico e Tecnológico**, em 18/10/2021, às 16:37, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Francisco Nivando Bezerra, Professor do Ensino Básico, Técnico e Tecnológico**, em 23/10/2021, às 15:32, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Ricardo Silva Thé Pontes, Usuário Externo**, em 08/11/2021, às 16:14, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade do documento pode ser conferida no site https://sei.ifce.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 informando o código verificador **3067033** e o código CRC **7B26D5F0**.

Dedico este trabalho ao meu marido Diego que sempre me inspira e apoia, aos meus pais e irmã, Dirlei, Adriana e Bárbara, pela educação e incentivo a alcançar meus sonhos

AGRADECIMENTOS

Meus agradecimentos maiores a Deus, que iluminou minha caminhada acadêmica e minha vida. O cultivo da fé permitiu ultrapassar as dificuldades e alcançar a realização pessoal e o sonho profissional.

Ao meu marido Diego por ser o maior incentivador, apoiador e desta jornada.

Ao Prof. Dr. Francisco José Alves de Aquino, pela disponibilidade e excelente orientação neste período de estudos.

Ao Corpo Docente do Programa de Pós Graduação de Engenharia de Telecomunicações pelo aprendizado recebido.

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP), pelo apoio financeiro, possibilitando que este trabalho fosse realizado.

Aos membros da banca

A todos que de alguma forma contribuíram para a realização deste trabalho.

RESUMO

Como forma de minimizar o desconforto de pacientes que necessitam submeter-se a exames invasivos, como laringoscopia, este trabalho propõem um método não invasivo que relaciona um conjunto de medidas acústicas retiradas a partir do estudo do sinal de voz falado e sustentado a vogal ‘a’ e um método de reconhecimento de padrões utilizando *Support Vector Machine* (SVM) como forma de classificação de patologias para hipótese de diagnóstico de patologias laríngeas. Utilizou-se a base de dados Alemã *Saarbrücken Voice Database* (SVD), que se encontra disponível *online* de forma gratuita pelo Instituto de Fonética da Universidade de *Saarland*, nesta base de dados é possível encontrar sinais de voz, entre saudáveis e patológicos de mais de 2000 sujeitos. Neste trabalho utilizou-se 786 amostras de sinais entre vozes masculinas e femininas, saudáveis e as patologias Edema de Reinke e Disfonia. Como medidas acústicas foram utilizadas a frequência fundamental, formantes, energia e entropia como vetor de característica de entrada para as máquinas de vetor suporte. Um dos objetivos do trabalho era a extração dessas características para a aplicação de *machine learning* para a identificação de patologias laríngeas, através da ferramenta SVM. Este classificador foi capaz de classificar vozes saudáveis de patológicas e as patologias Edema de Reinke e Disfonia.

Palavras-chave: Análise acústica, patologias laríngeas, *machine learning*, SVM.

ABSTRACT

As a way of minimizing the discomfort of patients who need to undergo invasive exams, such as laryngoscopy, this work proposes a non-invasive method that relates a set of acoustic measurements taken from the study of the spoken and sustained voice signal to a vowel 'a' and a pattern recognition method using SVM as a way of classifying pathologies for the hypothesis of diagnosis of laryngeal pathologies. We used the German Saarbrücken Voice Database (SVD), which is available online free of charge by the Institute of Phonetics of the University of Saarland, in this database it is possible to find voice signals, between healthy and pathological of more than 2000 subjects. In this work, 786 samples of signals were used among male and female voices, healthy and the pathologies Edema of Reinke and Dysphonia. As acoustic measures, the fundamental frequency, formants, energy and entropy were used as input characteristic vector for the algorithms. One of the objectives of the work was to extract these characteristics for the application of machine learning for the identification of laryngeal pathologies, through the SVM tool. This classifier was able to classify healthy voices as pathological and as pathologies Reinke's Edema and Dysphonia.

Keywords: Acoustic analysis, laryngeal pathologies, *Machine Learning*, SVM.

LISTA DE FIGURAS

Figura 2.1 - Posição da laringe na garganta.....	19
Figura 2.2 - Músculos intrínsecos (a) Músculo Cricoaritenóideo Posterior (CAP). (b) Músculo Cricoaritenóideo Lateral. (c) Músculo Aritenoideo (A). (d) Músculo Tireoaritenóideo (TA)	19
Figura 2.3 - Pregas vocais em: (a) adução e (b) abdução - visão endoscópica	20
Figura 2.4 - Trato Vocal	21
Figura 2.5 - Diferentes Graus do Edema de Reinke.....	27
Figura 2.6 - Fotografia clínica dos diferentes graus do edema de Reinke.....	27
Figura 2.7 - Superfície ótima para separação entre duas classes.....	34
Figura 2.8 - (a) Hiperplanos capazes de resolver o problema linearmente separável, com padrões de treinamento. (b) Problema resolvido por um hiperplano muito próximo a uma das classes.	35
Figura 2.9 - Interpretação geométrica de um padrão x ao hiperplano ótimo.....	38
Figura 2.10 - Problema hipotético binário não linearmente separável	45
Figura 3.1 - Página de download da base de dados SDV	52
Figura 3.2 – Visão geral do sistema de extração de atributos e classificação SVM	53
Figura 4.1 - Voz Saudável (a) no domínio do tempo e (b) no domínio da frequência.....	59
Figura 4.2 - Voz com Disfonia (a) no domínio do tempo e (b) no domínio da frequência.....	59
Figura 4.3 - Voz com Edema de Reinke (a) no domínio do tempo e (b) no domínio da frequência	60

LISTA DE TABELAS

Tabela 2.1 - Principais <i>Kernels</i>	47
Tabela 3.1 - Estrutura da matriz de confusão	56
Tabela 4.1 - Matriz de Confusão: FEMININA.....	61
Tabela 4.2 - Avaliação de desempenho	61
Tabela 4.3 - Matriz de Confusão: MASCULINA	62
Tabela 4.4 - Avaliação de desempenho	62
Tabela 4.5 - Acurácia para classificação de vozes Saudáveis x Patológicas.....	62
Tabela 4.6 - Sensibilidade para classificação de vozes Saudáveis x Patológicas	62
Tabela 4.7 - Especificidade para classificação de vozes Saudáveis x Patológicas.....	62
Tabela 4.8 - Matriz de Confusão: FEMININA.....	63
Tabela 4.9 - Avaliação de desempenho	63
Tabela 4.10 - Matriz de Confusão: MASCULINA	63
Tabela 4.11 - Avaliação de desempenho	63
Tabela 4.12 - Acurácia para classificação de vozes com Edema de Reinke x Disfonia	64
Tabela 4.13 - Sensibilidade para classificação de vozes com Edema de Reinke x Disfonia ...	64
Tabela 4.14 - Especificidade para classificação de vozes com Edema de Reinke x Disfonia .	64

LISTA DE ACRÔNIMOS

CAL	Cricoaritenóideo Lateral
CAP	Cricoaritenóideo Posterior
CAPE-V	<i>Consensus Auditory Perceptual Evaluation Of Voice</i>
CPPS	<i>Cepstral Peak Prominence Smoothed</i>
CV	<i>Cross Validation</i>
DFT	<i>Discrete Fourier Transform</i>
FIR	Resposta Finita Ao Impulso
FN	Falso Negativo
FP	Falso Positivo
Hz	Hertz
LST	<i>Department Of Language Science And Technology</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
ML	<i>Machine Learning</i>
RBF	Função De Base Radial
<i>MATLAB</i>	<i>MATrix LABoratory</i>
SVD	<i>Saarbrücken Voice Database</i>
SVM	<i>Support Vector Machine</i>
TA	Tireoaritenóideo
<i>UdS</i>	<i>Universität Des Saarlandes</i>
VC	<i>Vapnik-Chervonenkis</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

SUMÁRIO

1 INTRODUÇÃO	13
1.1 MOTIVAÇÃO.....	14
1.2 OBJETIVO GERAL.....	16
1.2.1 Objetivos específicos.....	16
1.3 ORGANIZAÇÃO DA DISSERTAÇÃO	16
2 FUNDAMENTAÇÃO TEÓRICA.....	18
2.1 LARINGE E PREGAS VOCALIS.....	18
2.2 VOZ E TRATO VOCAL	20
2.3 PATOLOGIAS LARÍNGEAS	25
2.4 ANÁLISE ACÚSTICA DOS SINAIS DE VOZES SAUDÁVEIS E PATOLÓGICAS	28
2.4.1 Série e Transformada de Fourier.....	28
2.4.2 Medidas acústicas do sinal de voz	29
2.5 <i>MACHINE LEARNING (ML)</i>	31
2.5.1 <i>Support Vector Machine (SVM)</i>	32
2.6 ESTADO DA ARTE.....	48
3 MATERIAIS E MÉTODOS.....	50
3.1. MATERIAIS	50
3.2. METODOLOGIA	51
3.2.1. Diagrama de blocos do sistema de extração de atributos e classificação	52
3.2.2. Extração de atributos.....	54
3.2.3. Classificador.....	54
4 RESULTADOS E DISCUSSÕES	58
4.1. EXTRAÇÃO DE ATRIBUTOS	58
4.2. CLASSIFICAÇÃO	61
4.2.1. Comentários e discussões.....	64
5 CONCLUSÃO E TRABALHOS FUTUROS.....	66
5.1. SUGESTÕES PARA TRABALHOS FUTUROS.....	68
REFERÊNCIAS BIBLIOGRÁFICAS	69

Capítulo 1

INTRODUÇÃO

A voz é formada através de tecidos musculares chamados de pregas vocais. Localizadas no interior da laringe, é um dos principais meios de comunicação entre os seres humanos. O som promovido por elas se dá através do ato de expulsar o ar dos pulmões, fazendo com que as vibrações produzam o som. Este som é um sinal acústico que contém informações importantes sobre o indivíduo que o gerou, a qualidade dessa voz depende diretamente da vibração e da abertura e fechamento da glote. A presença de patologias na laringe causa mudanças significativas nos padrões vibratórios, afetando a qualidade da produção vocal. Esses efeitos podem ser percebidos por um profissional especialista e então, através de exames, pode ser dado um diagnóstico (ROSA, 1998).

Dentre os tipos de disfunções provocadas por efeitos patológicos, as originadas de patologias laríngeas alteram o padrão da vibração das pregas vocais provocando modificações no timbre, na frequência, na amplitude, na composição dos harmônicos presentes na voz e também causando desordens no funcionamento laríngeo, respiratório e/ou trato vocal. Consequentemente, estas mudanças afetam a qualidade vocal do indivíduo, comprometendo a inteligibilidade e a efetividade da comunicação oral, assim como sua qualidade de vida devido aos desconfortos causados pela doença (ALMEIDA, 2010).

Há uma grande variedade de doenças relacionadas à laringe que causam modificações na voz, podendo ser de origem orgânica (organofuncionais), ou de origem neurológica (DIAS, 2020). As lesões organofuncionais, relacionadas ao trato vocal são nódulos, cistos, pólipos, edemas, consideradas enfermidades benignas decorrentes principalmente do comportamento vocal inadequado e a hábitos sociais não saudáveis como o tabagismo e o alcoolismo (CIELO et al., 2011). As patologias laríngeas de origem neurodegenerativas, como a paralisia das pregas vocais, decorrem de lesões do nervo vago ou de seus ramos que interligam

o tronco cerebral à laringe, podendo levar a alterações das funções que requerem o fechamento glótico (STEFFEN et al., 2004). No caso de paralisia, as principais causas incluem traumas cirúrgicos ou acidentais, doença cardiovascular e doenças neurológicas (STEMPLE, KLABEN, 2010).

As ações dos diferentes grupos musculares, presentes em todo o trato vocal, é comprometida pelo surgimento de alterações fisiológicas ocasionadas por patologias vocais capazes de impedir, comprometer ou dificultar a produção da voz (MENDOZA et al., 2003).

Diversas técnicas são utilizadas por profissionais da área médica para avaliar a qualidade vocal do paciente. O diagnóstico é feito a partir de procedimentos de rotina para examinar a laringe com propósitos clínicos ou de investigação, feitos por médicos otorrinolaringologistas ou fonoaudiólogos. Esses procedimentos incluem videolaringoscopia (realizado por meio de um aparelho endoscópico acoplado em um sistema de vídeo), videoestroboscopia (iluminação estroboscópica da laringe, útil para visualização dos movimentos), eletromiografia (observação indireta do estado funcional da laringe) e videofluoroscopia (técnica radiográfica na qual o paciente ingere uma substância rádio-opaca para avaliar a deglutição). Todos esses métodos são invasivos e causam desconforto ao paciente (SOBRÉ, 2016).

A desordem vocal é perceptível quando o sinal acústico é analisado. Logo, doenças que provoquem alguma alteração vocal podem ser diagnosticadas, acompanhadas e até tratadas por técnicas baseadas em análise acústica do sinal de voz (ROSA, 1998).

1.1 MOTIVAÇÃO

A possibilidade de avaliar a qualidade vocal e as patologias que a modificam a partir da análise das características do sinal acústico vocal através de métodos não-invasivos é a principal motivação para o desenvolvimento desta pesquisa.

Os métodos invasivos usados para o diagnóstico citados acima causam desconforto ao paciente além de serem procedimentos que exigem instrumentos sofisticados e caros (ALMEIDA, 2010). As técnicas de processamento digital de sinais têm sido utilizadas, para análise acústica, como uma eficiente ferramenta para o auxílio da detecção das alterações dos padrões dos sons provocados por patologias da laringe e a classificação dessas doenças da voz auxiliam no diagnóstico precoce e no desenvolvimento do processo terapêutico (COSTA et al, 2007).

Para a classificação de uma voz patológica, por meio de análise acústica, é fundamental que o processo de extração das características acústicas da patologia seja bem elaborado. Para tanto, deve-se buscar uma modelagem acústica para a patologia, utilizando características acústicas que diferenciem uma voz patológica de uma voz normal ou que seja capaz de fazer a distinção entre as patologias.

Apesar de existir um desenvolvimento significativo na classificação de vozes patológicas, usando ferramentas de processamento da fala, o reconhecimento automático no padrão dessas patologias ainda está numa fase inicial (CORDEIRO, 2016). A literatura ainda não é conclusiva com relação às quais características acústicas são mais adequadas para modelagem de uma patologia em particular. A maioria das pesquisas baseia-se na discriminação entre vozes normais e patológicas, sem especificar a patologia. Alguns estudos focam em uma determinada patologia sem, no entanto, apresentar um modelo acústico correspondente. A pesquisa para uma análise acústica mais representativa e detalhada de sinais de vozes patológicas é ainda um campo promissor (TORRES, 2018). Por esse motivo justifica-se a escolha do tema deste trabalho.

Para que o algoritmo classificador consiga detectar a presença patológica, faz-se necessário que características obtidas representem significativamente as desordens provocadas por uma determinada patologia. A escolha de características confiáveis que representem cada tipo de patologia pode se tornar uma tarefa difícil, uma vez que dependerá de vários fatores como o tipo e o grau da lesão, a severidade dos efeitos causados pela patologia na produção vocal e a quantidade de ruído presente no sinal de voz analisado (COSTA, 2012).

Para o algoritmo de classificação utilizado nesse trabalho, optou-se por uma técnica de *Machine Learning* (ML), área que atualmente, há grande disseminação do uso de técnicas de ML, em diversos tipos de aplicação do conhecimento. Tem como objetivo principal desenvolver ferramentas computacionais eficientes capazes de fornecer soluções satisfatórias, para problemas com alta complexidade de classificação de padrões, mesmo que sendo linearmente não-separáveis e com dimensionalidade elevada (ROCHA-NETO, 2011).

Com a parte que diz respeito à implementação e simulação dos algoritmos de extração de características e de classificação de vozes saudáveis e patologias, o desenvolvimento da ferramenta será feito utilizando o *software MATLAB (MATrix LABoratory)* da *MathWork*¹. O *software* apresenta grande popularidade dentro do meio acadêmico, pela facilidade da criação de algoritmos, é um ambiente de programação, visualização e análise de

¹ Disponível em <<https://www.mathworks.com/products/matlab.html>>

dados de alto nível com ampla utilização em pesquisa básica e aplicada, ensino e nos meios profissionais das mais diversas áreas. Ele possui linguagem própria, mas também interpreta outras linguagens como, por exemplo, C. Por tal fato, inúmeros trabalhos científicos são realizados utilizando o *MATLAB* (SALVIANO, 2009).

1.2 OBJETIVO GERAL

Este trabalho tem como objetivo geral o desenvolvimento de um procedimento de extração de características dos sinais acústicos da fala para que um algoritmo de reconhecimento de padrão diferencie vozes saudáveis e patológicas e o reconhecimento de duas patologias Edema de Reinke e Disfonia.

1.2.1 Objetivos específicos

No sentido mais específico, esta dissertação tem os seguintes objetivos:

- Levantamento da teoria sobre SVM para o processo de classificação de padrões;
- Extrair características significativas do sinal de voz (frequência fundamental, formantes, energia e entropia);
- Desenvolver o algoritmo de extração de características e de reconhecimento e classificação dos padrões das vozes saudáveis e patológicas e entre as patologias Edema de Reinke e Disfonia;
- Analisar os resultados obtidos;

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação é ordenada em cinco capítulos.

O Capítulo 2 aborda a fundamentação teórica para o estudo da voz: anatomia e fisiologia da laringe, pregas vocais, produção da fala e trato vocal e as patologias; análise acústica e extração das características do sinal de voz, a apresentação da teoria da SVM e a revisão da literatura no estado da arte.

Os detalhes do desenvolvimento dos algoritmos, os materiais utilizados para evolução da pesquisa e a metodologia proposta para atingir os objetivos, são descritos no Capítulo 3.

O Capítulo 4 trata dos resultados obtidos nas simulações e a discussão sobre eles.

No quinto e último capítulo, é apresentada a conclusão do trabalho de acordo com os dados apresentados no capítulo anterior, validando a proposta de pesquisa. Além de conter as contribuições desta dissertação e sugestões para trabalhos futuros, de modo a dar continuidade a pesquisa.

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

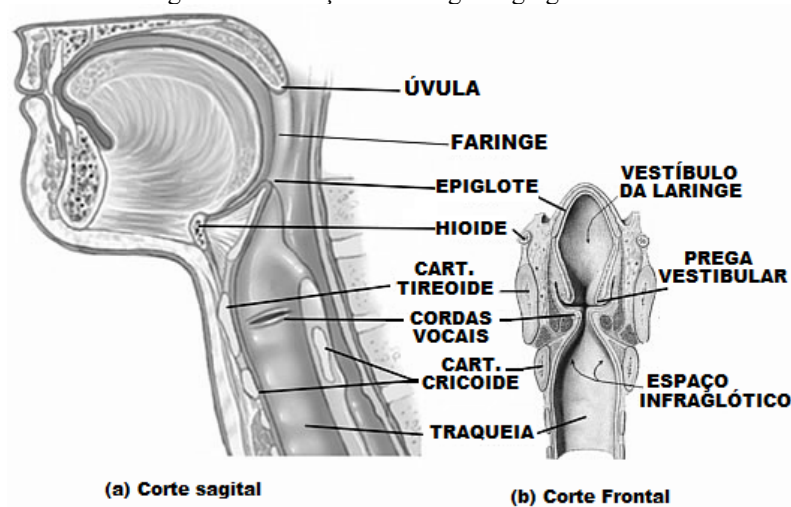
Neste capítulo são descritos os fundamentos para o estudo da voz: anatomia e fisiologia, conceitos estruturais, acústicos, características e ferramentas usadas para a discriminação e a classificação de vozes saudáveis e patológicas. São apresentadas as bases teóricas da anatomia funcional, fisiologia da laringe, seu funcionamento e duas das patologias mais comuns que atingem as pregas vocais, sendo elas a disfonia como patologia de lesão primária e Edema de Reinke como patologia de lesão secundária.

2.1 LARINGE E PREGAS VOCAIS

A produção da voz não é a função principal da laringe. A função primária desse órgão é conduzir o ar e proteger os pulmões da entrada de substâncias indesejadas (função principalmente realizada pela epiglote) e auxiliar na deglutição através de um movimento vertical de toda a estrutura laríngea, exercido pelos músculos extrínsecos (músculos que a conectam a outras partes do corpo). A ação de produzir o sinal glotal é uma função secundária da laringe. Isso ocorre por causa da morfologia específica de alguns de seus músculos, particularmente das pregas vocais. Resumidamente pode-se dizer que a laringe é uma válvula de controle de vazão de ar, um sistema complexo de músculos responsável pela geração de energia para o sistema fonador (ROSA, 2002).

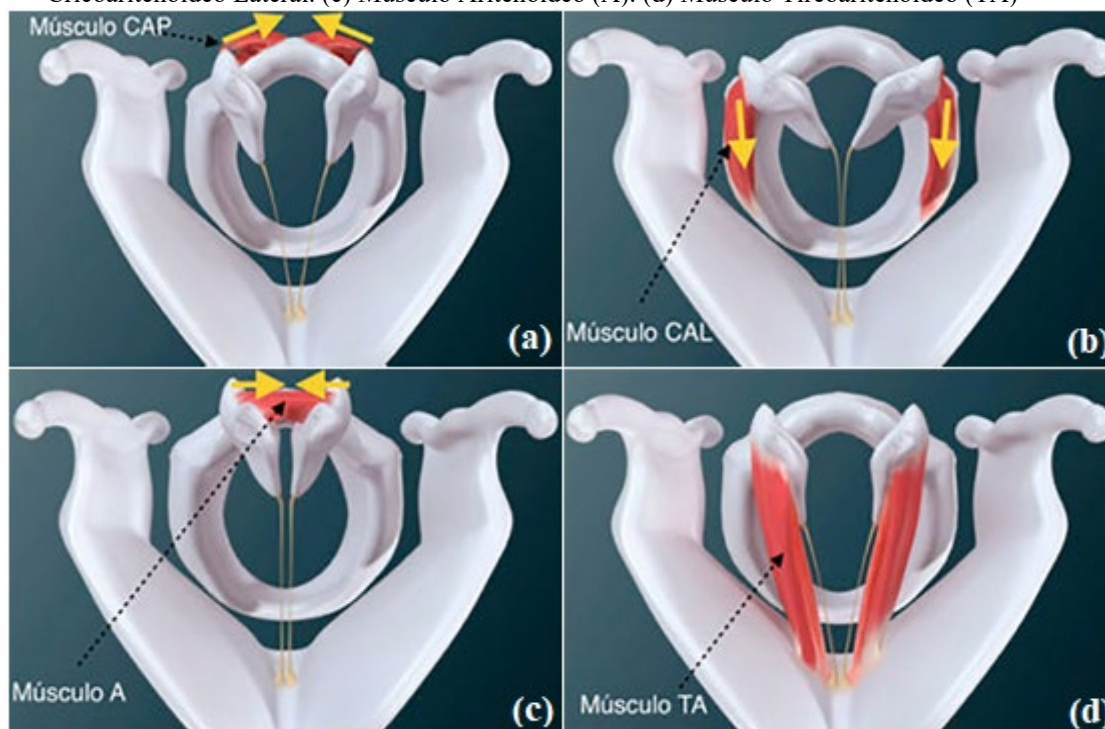
A laringe, como mostrada na Figura 2.1, é composta por um grupo de músculos intrínsecos e extrínsecos que movimentam ou sustentam as cartilagens na garganta. Os músculos extrínsecos movimentam e sustentam toda a laringe, já os músculos intrínsecos, movimentam a parte de dentro da laringe.

Figura 2.1 - Posição da laringe na garganta



FONTE: Adaptado GUERINO, 2015

Figura 2.2 - Músculos intrínsecos (a) Músculo Cricoaritenóideo Posterior (CAP). (b) Músculo Cricoaritenóideo Lateral. (c) Músculo Aritenóideo (A). (d) Músculo Tireoaritenóideo (TA)



FONTE: Adaptado Yamasaki e Tsuji, 2017

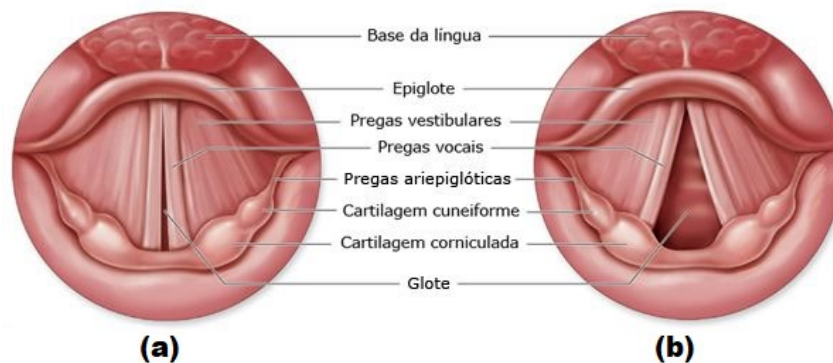
A fonação depende da ação desses músculos intrínsecos. Eles são responsáveis pela abertura (abdução) e fechamento (adução) do espaço glotal, atividades que efetivamente controlam a vibração das pregas vocais.

O músculo que realiza esse movimento é o cricoaritenóideo posterior (CAP). Esse par de músculos localiza-se na região posterior da laringe, mostrado na Figura 2.2 (a), em seguida a Figura 2.2 (b) apresenta o músculo cricoaritenóideo lateral (CAL), responsável pelo

fechamento da prega vocal, assim como o músculo aritenóideo (A), mostrado na Figura 2.2 (c), responsável pelo fechamento da glote posterior e enfim o músculo tireoaritenóideo (TA), destacado na Figura 2.2 (d), que encurta a prega vocal e torna a voz mais grave (YAMASAKI e TSUJI, 2017).

Entre a parte interna da base das aritenóides e da tireóide, estão as pregas vocais. Chamada de cordas, pregas ou dobras vocais refere-se a dois pares de lábios, simetricamente formados por um músculo e um tecido elástico, situadas no interior da laringe. As extremidades das pregas vocais estão fixadas sobre as aritenóides. Ao espaço triangular compreendido entre as pregas vocais, dá-se o nome de glote mostrado na Figura 2.3. (PARRAGA, 2002).

Figura 2.3 - Pregas vocais em: (a) adução e (b) abdução - visão endoscópica



FONTE: Adaptado de Central da Fonoaudiologia²

As pregas vocais têm papel importante na fonação, elas são estruturas multilaminadas e cada camada apresenta propriedades mecânicas diferentes (ZITTA, 2005).

Para facilitar o estudo do sistema de produção de fala, ele é normalmente reduzido à três grupos distintos: respiração, vocalização e ressonância apresentados nas próximas seções.

2.2 VOZ E TRATO VOCAL

Durante a fonação, as mudanças de tensão e longitude das pregas vocais, ampliação da abertura glótica e a intensidade do esforço respiratório provocam variações no tom da voz, tom que resulta da vibração das pregas vocais, modificado pelos movimentos da faringe, língua e lábios para a fala (ZITTA, 2005).

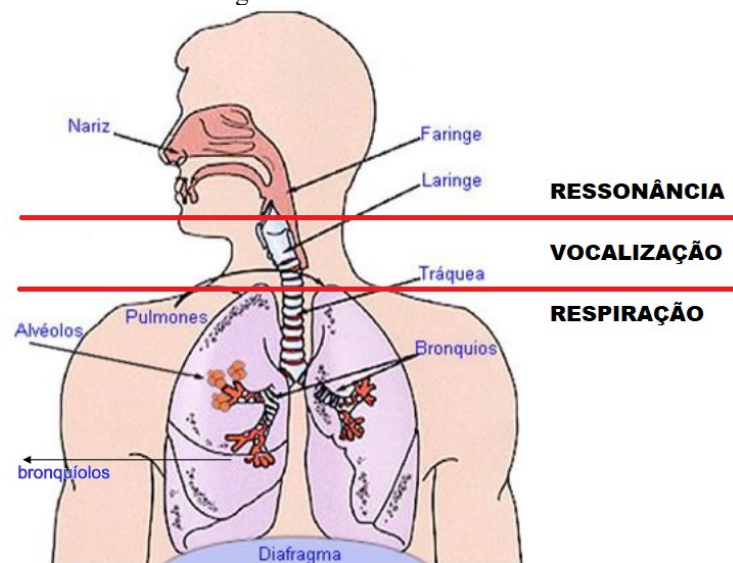
A produção do som se origina na laringe como um tom fundamental, que é modificado por várias câmaras de ressonância acima e abaixo desta. O som é convertido em

² Disponível em: <http://www.centraldafonoaudiologia.com.br/tratamentos/fonoaudiologia-nodulos-de-pregas-vocais>. Acesso em: 17 set. 2019

fala por ação da faringe, língua, palato, lábios e estruturas relacionadas. Durante a fonação, a corda vocal é aduzida para próximo da linha mediana pelo músculo CAP, que funciona como um tensor das pregas vocais e os ajustes isométricos finos são feitos pelo músculo TA. O movimento medial da prega resulta da tensão na corda vocal, diminuição na pressão de ar em cada abertura vibratória da glote e do efeito de sucção do ar que escapa (efeito Bernoulli), o período entre duas aberturas (ou fechamentos) sucessivos determina a frequência fundamental, o *pitch*. O resultado deste ciclo de abertura e fechamento da glote é a liberação de pequenos deslocamentos da coluna de ar subglótica, que formam as ondas sonoras.

O som é uma onda mecânica que necessita de um meio físico de propagação, se movimenta em sentido transversal à direção de propagação. As sucessivas compressões e descompressões do ar se propagam até o ouvido (ou transdutor, podendo ser um microfone, por exemplo), onde é adequadamente descodificado, traduzido em pulsos elétricos e interpretado (ROSA, 1998).

Figura 2.4 - Trato Vocal



FONTE: Adaptado de GUERINO, 2015

O estudo anátomo-fisiológico do trato vocal é dividido em três grupos estruturais, como ilustrado na Figura 2.4, onde a onda sonora que é produzida ou modificada pelos órgãos. O primeiro grupo corresponde a produção de um fluxo de ar, iniciando nos pulmões e encerrando no final da traqueia, recebe o nome de subgrupo de respiração. Situando-se na laringe encontram-se os órgãos do segundo subgrupo, responsáveis pela produção de uma onda quase-periódica mediante vibrações das pregas vocais em um movimento por variação de pressão (segundo a lei de Bernoulli). Com a responsabilidade de produzir os fonemas que

constituem o vocabulário acústico de uma língua encontra-se o sistema ressonador, constituindo assim o último grupo de órgãos do trato vocal (ROSA, 1998).

2.2.1 Respiração

O primeiro grupo de órgãos está ligado à geração de fluxo de ar variável que excitará o sistema produtor de fala, a laringe e os órgãos posteriores.

A respiração tem função primária de transporte de gases para a renovação de oxigênio, por isso existem limitações sobre a quantidade de som, mais precisamente, fonemas, que se pode sustentar por uma única expiração. Um importante fator a ser considerado nos experimentos de análise da voz normal e/ou patológica.

Anatomicamente, o sistema respiratório do ser humano compreende um par de pulmões, que funciona como reservatório de ar, os sacos aéreos e os bronquíolos, responsáveis pelo transporte de ar a partir dos pulmões, onde as trocas de gasosas são realizadas, entregando a partir destes a massa de ar para o par de brônquios que unem-se a traqueia, servindo de duto para o ar circular em ambas as direções. Além destas estruturas, há o diafragma que é composto de musculatura, tendão e membrana, importante na atividade respiratória, logo, diretamente ligado ao processo fonatório.

A função respiratória é necessária à produção de voz para que as ondas sonoras sejam “levadas” por este fluxo de ar. Como esta capacidade respiratória é limitada pelo volume de ar armazenado nos pulmões, a capacidade da fonação fica limitada tanto em altura como em duração.

Para que ocorra o processo de respiração, uma ação conjunta de músculos intercostais e do diafragma é necessária, para que se permita a entrada e a saída do ar. O ato de respirar dividido em dois movimentos a inspiração e expiração (SANTOS, 2010).

Inspiração é o conjunto de movimentos que permite a entrada de ar nos pulmões. Nesse processo ocorre a contração do diafragma, ocasionando seu abaixamento. Os músculos intercostais se contraem e as costelas se levantam, fazendo com que o tórax aumente de tamanho e a pressão interna dos pulmões torne-se menor que a externa. Com a redução da pressão intrapulmonar, o ar entra pelas vias respiratórias e chega até os alvéolos para que ocorram as trocas gasosas. Já na expiração o conjunto de movimentos resulta na retirada do ar do interior das vias respiratórias. Nesse caso, ocorre o relaxamento dos músculos intercostais e do diafragma. Diferentemente do que ocorre na inspiração, há uma redução do volume da caixa torácica, que volta ao seu tamanho de repouso, retraindo os pulmões, o que faz com que a

pressão interna fique maior que a externa. O aumento da pressão intrapulmonar faz com que o ar seja expelido para fora, onde a fonação é produzida.

A capacidade vital de um indivíduo está relacionada com o volume total de ar que pode ser expirada dos pulmões a partir de uma inspiração máxima. Estas medidas possibilitam avaliação da capacidade pulmonar e respiratória de um ser humano e servem para determinar a diferença entre o tempo de expiração e inspiração, verificando-se que o tempo de expiração é levemente superior ao da inspiração, principalmente devido as necessidades de fonação e controle do tempo de expulsão de ar para os órgãos glotais e supraglóticos (ROSA, 1998).

2.2.2 Vocalização

Mediante o controle coordenado da capacidade valvular da laringe e pregas das vocais pode-se produzir um som de baixa intensidade, quase-periódico, que quando corretamente modulado pelo trato supraglotal torna-se inteligível. A laringe é composta por um conjunto de cinco cartilagens principais que compõem o sistema vibratório, sendo elas: cricoidea, tireoidea, um par de aritenoideas e a epiglote. As pregas vocais que produzem este sinal quase-periódico através da vibração da mucosa composta pelo epitélio e lâmina superficial, chamado espaço de Reinke, junto com a camada de transição (ligamento vocal), deslizam sobre o músculo TA, por quase toda a extensão do espaço glotal (ROSA, 1998).

Segundo Rosa (1998), para produzir o sinal característico os músculos intrínsecos adutores CAL e TA contraem-se rapidamente (na ordem de até 15ms), de modo a unir as pregas vocais e obstruir a passagem de ar no espaço glotal, aproximando as pregas vocais antes do início da vocalização. As pregas vocais não se fecham totalmente, uma fenda de até 3mm possibilita a vibração da mucosa. Ao alterar o nível de pressão e velocidade do fluxo de ar através desta fenda, as pregas vocais fecham-se e abrem-se, criando um ciclo de atividade do ar, e efeito sugador de Bernoulli as reaproxima novamente. Este ciclo vibratório da mucosa repete-se aproximadamente 125 vezes por segundo nos homens e 215 vezes por segundo em mulheres, ambos adultos.

Características com o comprimento e a espessura das pregas vocais, são importantes para a formação da frequência fundamental do sinal produzido pela abertura glotal. Pregas curtas e espessas tendem a ser frouxas, resultando em uma baixa velocidade na realização do ciclo de vibração, já pregas mais finas e tensas levam a produção de sons mais agudos.

O aumento e a redução no comprimento das pregas vocais produzem sons mais altos ou mais baixos, uma vez que aumentam ou reduzem a densidade do tecido das pregas

vocais. O aumento da tensão sobre as pregas vocais, exige um maior fluxo de ar para produzir a vibração na mucosa, devido a alteração massa-volume desta. A intensidade da voz, o volume, está diretamente relacionada com alterações na queda de pressão do ar na glote. A quantidade de ar e a velocidade do fluxo provida pelos pulmões, influenciam na velocidade do ciclo de abdução/adução das pregas vocais, sendo comum em pessoas não treinadas elevarem tanto a frequência fundamental quando a intensidade o sinal glotal.

2.2.3 Ressonância

O sinal de baixa intensidade proveniente das pregas vocais necessita ser amplificado e suas componentes harmônicas sofrem ênfase para que os fonemas sejam caracterizados (ROSA, 1998).

O trato supraglotal inicia-se a partir das hipofaringe (cavidade diretamente acima do esôfago próximo a epiglote), seguido pelos constritores médios da faringe, região denominada orofaringe, acompanhado pela valécula e o início da língua, do trato bucal e nasal. O acoplamento do trato nasal é feito através do véu palatino, ou palato mole, finalizando a cavidade nasal. A cavidade bucal apresenta a língua sustentada pela mandíbula, finalizando com os dentes e lábios.

A produção dos fonemas é regida por um fenômeno chamado ressonância, onde há ênfase na atenuação de regiões do espectro, os diversos músculos e estruturar supraglotalis agem de forma a alterar a frequência natural que será amplificada. Assim, as componentes harmônicas e não-harmônica são modificadas adequadamente de acordo com cada fonema a ser produzido.

O formato anatômico destas estruturas também influencia na produção da fala, o acoplamento da cavidade nasal. Grau de tensão ou relaxamento nos constritores faríngeos, nível de abertura mandibular e dos lábios influem no fonema a ser produzido. Os elementos ressonadores fornecem a percepção do dialeto regional e auxiliam na identificação pessoal.

É importante citar que a frequência fundamental proveniente da cavidade glotal sofre modificações ao longo do trato vocal, por modificações na intensidade das componentes harmônicas. A escolha da vogal para avaliação patológica é de suma importância para uma correta análise acústica.

2.3 PATOLOGIAS LARÍNGEAS

O ato de falar exige o bom funcionamento e condições das estruturas do trato vocal. Uma voz é considerada normal e saudável quando não há esforço na sua produção. Qualquer interferência nessa dinâmica implica na geração de um sinal acústico diferente do emitido pelo falante, geralmente exigindo algum esforço adicional, fatores ambientais e físicos podem interferir no padrão adequado da qualidade vocal (COSTA, 2008).

Alterações morfológicas nas estruturas do trato vocal ou de seu mau funcionamento resultam na desordem vocal. Enquanto a audição é essencialmente uma função neurossensorial, a voz depende essencialmente da atividade de todos os músculos que participam de sua produção, além da integridade de todos os tecidos do aparelho fonador. Quando essa harmonia é mantida, o som é emitido sem dificuldade pelo falante e apresenta uma boa qualidade para os ouvintes. Quando a voz está fora dos padrões de normalidade, diz-se que está perturbada ou disfônica. (ROSA, 2002).

2.3.1 Disfonia

A voz é uma função que depende principalmente da atividade muscular. Quando a harmonia entre os músculos responsáveis pela produção da fala e a integridade dos tecidos do aparelho fonador é mantida, um som de boa qualidade é emitido sem dificuldade para o falante. Em oposição, quando certas condições não são atingidas, se tem uma disfonia (ZITTA, 2005).

Uma disfonia representa qualquer dificuldade que impeça a produção natural da voz, alguns dos sintomas para a identificação de uma disfonia são: dificuldade em manter a voz, rouquidão, cansaço ao falar, variações na frequência habitual, esforço na emissão, falta de intensidade e pouca resistência (BEHLAU, 2001).

Segundo Behlau (2001), o processo de desenvolvimento de uma disfonia leva a diferentes propostas de classificação. As disfonias são classificadas em orgânicas, funcionais e organofuncionais. Disfonias funcionais são desordens do comportamento vocal, o uso incorreto da voz é considerado uma disfonia funcional primária. Disfonias funcionais secundárias resultam das inaptações vocais e são discretas alterações que comprometem apenas a produção vocal, causadas por inaptação respiratória, ressonatal, fônica ou desequilíbrios entre o tamanho da laringe e as caixas de ressonância. As disfonias psicogênicas são resultantes do estado emocional. As disfonias orgânicas são aquelas que não são causadas pelo mau uso da

voz, e sim por uma série de processos, tais como: tumores benignos e malignos, doenças neurológicas e lesões por uso de armas brancas ou de fogo.

As disfonias organofuncionais aquelas relacionadas ao funcionamento da voz com lesões secundárias; uma evolução de uma disfonia funcional. Geralmente caracterizada pelo seu diagnóstico tardio ou pelo não reconhecimento da possibilidade de se desenvolver uma lesão secundária. Essas lesões, diagnosticadas por meio de um exame laringoscópico, abrangem nódulos, pólipos e edemas de Reinke (BEHLAU, 2001).

O campo de interesse desta pesquisa estão nas disfonias organofuncionais representadas pelo Edema de Reinke, que serão estudados a seguir.

2.3.2 Edema de Reinke

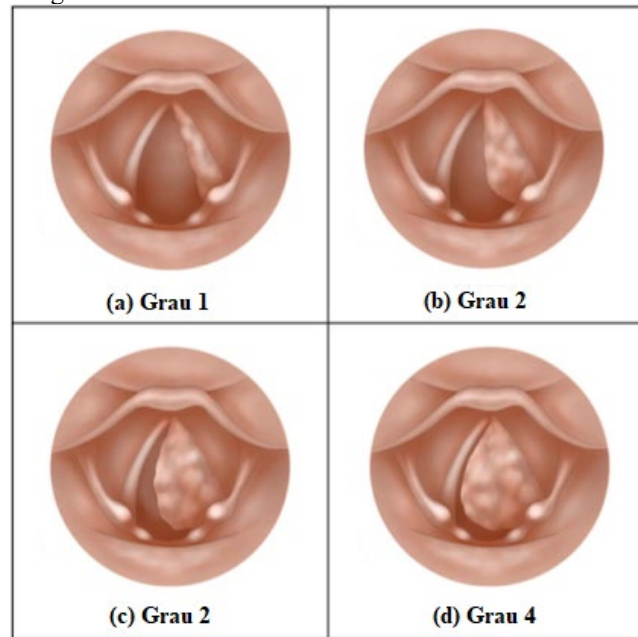
Edema de Reinke é uma lesão bilateral alargada, de coloração rosada, com fenda irregular e pouca abertura respiratória. Localiza-se por toda prega vocal e é causada por fonotrauma e/ou tabagismo (NICOLSI, 2003).

O edema de Reinke mostrado nas Figura 2.5 e 2.6, caracteriza-se pelo aumento e inchaço das pregas vocais e pelo acúmulo de material líquido, gelatinoso ou semissólido na camada superficial da lâmina própria, chamada de espaço de Reinke, das pregas vocais. Classifica-se em 4 graus, sendo eles:

- Grau 1: Degeneração polipoide mínima da corda vocal ocupando até 25% da fenda glótica;
- Grau 2: Lesão polipoide expandida ocupando entre 25% e 50% da fenda glótica;
- Grau 3: Lesão polipoide expandida ocupando entre 50% e 75% da fenda glótica;
- Grau 4: Lesão obstrutiva independente do lado ocupando mais de 75% da fenda glótica.

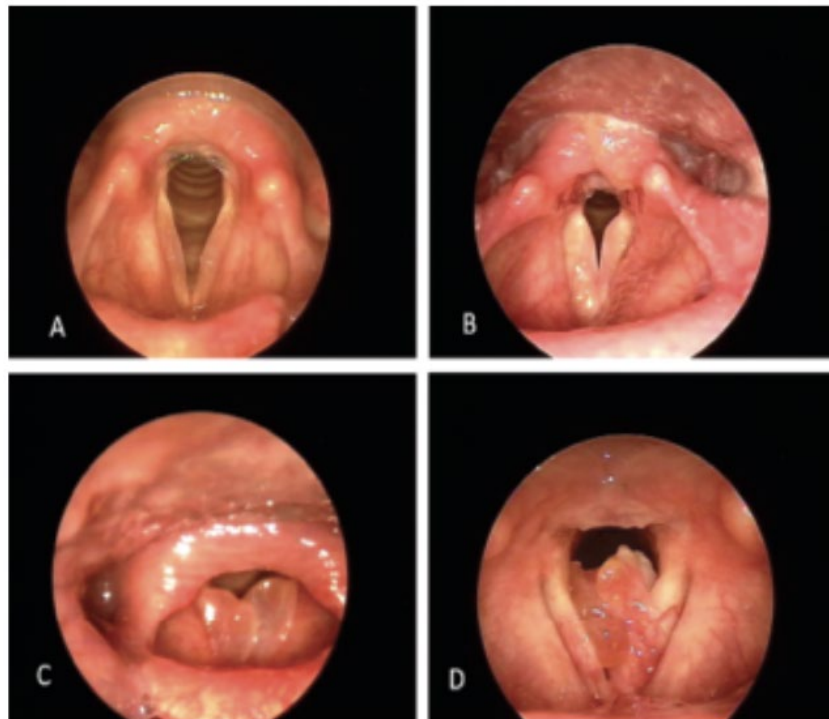
Segundo Barbosa (2000), a incidência é maior em mulheres que atingiram a menopausa. Entretanto, encontra-se em ambos os sexos, e é uma patologia da idade adulta. O sintoma clássico do edema de Reinke é a disfonia, diferentemente de qualquer tumor benigno de pregas vocais, esta disfonia é de um timbre mais baixo, como produzido por pregas vocais mais flácidas. Em mulheres portadoras desta patologia, o timbre vocal se assemelha a uma voz masculina.

Figura 2.5 - Diferentes Graus do Edema de Reinke



FONTE: Adaptado de *Anesthesiology*, 2018.

Figura 2.6 - Fotografia clínica dos diferentes graus do edema de Reinke



FONTE: *The American Laryngological, Rhinological and Otological Society, Inc.*

Seu diagnóstico pode ser feito através da laringoscopia indireta, por meio da videolaringoscopia ou pode estar imediatamente evidente nos casos moderados ou avançados. As pregas vocais apresentam um edema submucoso ao longo de toda a prega vocal, bilateralmente, cheia de fluido, flácidas e pálidas.

2.4 ANÁLISE ACÚSTICA DOS SINAIS DE VOZES SAUDÁVEIS E PATOLÓGICAS

O sinal de acústico da voz pode ser representador por uma sequência de vetores de características obtidas através da utilização de ferramentas matemáticas como a Transformada de Fourier. Nesta seção, são apresentadas algumas técnicas baseadas na análise acústica de extração de características do sinal de voz que serão adicionadas ao vetor de características para posterior classificação, são elas frequência fundamental (*pitch*), formantes, energia e entropia.

2.4.1 Série e Transformada de Fourier

A Série de Fourier mostra que um sinal periódico $x(t)$ pode ser composto por uma soma de funções seno e cosseno com frequências múltiplas de uma frequência fundamental f_0 , a Equação 2.1 mostra a série de Fourier trigonométrica (SPIEGEL, 1976, p.249):

$$x(t) = a_0 + \sum_{n=1}^{\infty} [a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)], \quad (2.1)$$

Onde os coeficientes são:

$$a_0 = \frac{1}{T} \int_0^T x(t) dt, \quad (2.2)$$

$$a_n = \frac{2}{T} \int_0^T x(t) \cos(2\pi n f_0 t) dt, \quad (2.3)$$

$$b_n = \frac{2}{T} \int_0^T x(t) \sin(2\pi n f_0 t) dt. \quad (2.4)$$

Para sinais pares, ou seja, quando $x(t) = x(-t)$, pode-se reescrever a Equação 2.1 da seguinte maneira:

$$x(t) = a_0 + \sum_{n=1}^{\infty} [a_n \cos(2\pi n f_0 t)]. \quad (2.5)$$

Já quando o sinal é ímpar, $x(t) = -x(-t)$, podemos reescrever a Equação 2.1 da seguinte maneira:

$$x(t) = a_0 + \sum_{n=1}^{\infty} [b_n \sin(2\pi n f_0 t)]. \quad (2.6)$$

Assim com a Série de Fourier é usada na análise de sinais periódicos, a Transformada de Fourier é utilizada na análise de sinais aperiódicos, definida na Equação 2.7 pelo par de transformadas direta e na Equação 2.8 a inversa:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt, \quad (2.7)$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df. \quad (2.8)$$

Com o surgimento do processamento de sinais discreto, tornou-se possível calcular a Transformada Discreta de Fourier (DFT – *Discrete Fourier Transform*) de um sinal através do par definido pelas Equações 2.9 e 2.10:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)nk}, \quad (2.9)$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j(2\pi/N)nk}, \quad (2.10)$$

basta que um grupo tenha $x(n)$ com N amostras do sinal $x(t)$. Para isso, respeitando o teorema da amostragem, estabelece-se que a frequência de amostragem f_s deve ter valor maior ou igual ao dobro do valor da frequência do sinal analisado. N é o número de amostras, $x(n)$ são as amostras de $x(t)$ e $X(k)$ são as componentes espectrais. O espaçamento entre as componentes espectrais é dado por f_s/N .

2.4.2 Medidas acústicas do sinal de voz

A seguir, são descritas as medidas acústicas frequência fundamental (f_0), formantes, energia e entropia do sinal de voz

2.4.2.1 Frequência fundamental (f_0) – *Pitch*

A Frequência Fundamental (f_0) é dada em Hertz (Hz) e corresponde ao número de ciclos glóticos por segundo que se aproxima de 125 Hz para homens, 215 Hz para mulheres e acima de 250 Hz para crianças. Essa frequência recebe o nome de primeiro harmônico e é

definida por uma interação complexa entre o comprimento, massa e tensão das pregas vocais, sendo esses parâmetros controlados pelos músculos intrínsecos e extrínsecos da laringe (ROSA, 1998).

Neste estudo, as vozes saudáveis femininas estão entre 170 e 270 Hz e as masculinas entre 80 e 160 Hz.

A presença de patologias pode afetar a frequência fundamental, fazendo com que mulheres e homens produzam frequências fundamentais baixas ou elevadas, respectivamente, ou seja, fora dos valores médios usuais para o sexo (COSTA, 2008).

2.4.2.2 Formantes

As formantes são representadas pelas frequências naturais de ressonância do trato vocal, especificamente na posição articulatória da vogal falada. As vogais são identificadas por suas formantes (BEHLAU, 2001). As formantes determinam a qualidade das vogais e contribuem muito para o timbre pessoal do indivíduo (CORDEIRO, PINHO E CAMARGO, 2007).

Segundo Costa (2008), os valores das formantes, para o caso patológico, em geral, são mais baixos, apresentando alguns valores fora da faixa para voz normal.

A ressonância do tubo ressonador depende de seu comprimento e de seu diâmetro transversal (KENT e READ, 2002). As formantes são as frequências das ressonâncias do trato vocal, onde se concentram as maiores energias acústicas. Através dos formantes podemos reconhecer as características segmentais da fala. Por exemplo, os três primeiros formantes são os principais determinantes da qualidade fonética de uma vogal. A fonte de energia e o ressonador são independentes, a frequência de vibração das pregas vocais não afeta as propriedades do ressonador (CAMARGO, 2000; KENT e READ, 2002).

2.4.2.3 Energia

Em processamento digital de sinais de voz, a medição da intensidade sonora pode ser obtida por meio do cálculo da energia do sinal. Para isso, são usadas técnicas no domínio do tempo (análise temporal) e/ou no domínio da frequência (análise espectral) (ALMEIDA, 2010).

A vibração quase periódica das pregas vocais produz a fonte de energia conhecida como vocalização. Assim, esta fonte de energia gera um espectro harmônico, onde a energia é

calculada pela interação da energia contida em cada frequência que compõem esse sinal em um dado instante de tempo. Esse é um método que fornece, de forma precisa a variação da energia da voz ao longo do tempo, por considerar as contribuições de todas as frequências do sinal, isto é, tanto da energia dos harmônicos quanto a energia do componente de ruído, O componente de ruído é provocado pela turbulência do ar ao passar pela glote e o componente harmônico é produzido pela vibração das pregas vocais e alterado de acordo com cada fonema (ZITTA, 2005).

2.4.2.4 Entropia

A entropia aparece como uma medida do grau de ordem ou desordem do sinal, fornecendo informações úteis sobre o processo dinâmico associado ao sinal. A entropia de Shannon (SHANNON, 1948) é um critério útil para analisar e comparar a distribuição de probabilidade, já que fornece uma medida da informação para qualquer distribuição de probabilidade.

A definição de quantidade de informação ou incerteza probabilística, proposta por Shannon, utiliza um tratamento probabilístico e faz uso de uma função logarítmica, tendo sua forma funcional sendo representada por:

$$H(p_1, \dots, p_n) = -\sum_{i=1} p_i \log(p_i), \quad (2.11)$$

onde p_i é a probabilidade de ocorrência de cada evento e $\sum_{i=1} p_i = 1$.

2.5 MACHINE LEARNING (ML)

As técnicas de ML e reconhecimento de padrões têm um grande leque de aplicações em inúmeras áreas científicas e tecnológicas, principalmente no âmbito de desenvolvimento de sistemas inteligentes. Chamado de padrão de teste ou simplesmente padrão é definido como uma entidade numérica que recebe um rótulo, uma classificação. As demandas em sistemas de reconhecimento de padrões automáticos estão aumentando bastante devido à disponibilidade de bases de dados grandes e de exigências de desempenho restritas (velocidade, exatidão e custo) (ROCHA-NETO, 2006).

O projeto de um sistema de reconhecimento de padrões envolve essencialmente os seguintes aspectos:

1. aquisição e pré-processamento de dados;
2. representação de dados;
3. treinamento;
4. tomada de decisão (a qual classe um padrão em teste pertence).

O papel do módulo de pré-processamento é remover ruídos, normalizar padrões, ou qualquer outra operação que contribuir na definição das amostras de uma forma compacta. A representação de um padrão ou vetor $\mathbf{x} = (x_1, x_2, x_3, \dots, x_d)^T$ é dado por um conjunto de d características ou atributos, e pode ser visto como um ponto (ou vetor) no espaço d -dimensional. O objetivo da área de reconhecimento de padrões é a escolha das características, que pertencem aos vetores, que possam diferenciar os padrões em regiões compactas e separadas por classe em um espaço característico d -dimensional, então considerar o vetor $\mathbf{x}$ pertencente a uma classe W (em $W_1, W_2, \dots, W_d$) baseado nos valores de suas d características. A eficácia do conjunto de características é determinada por quão bem os padrões de diferentes classes possam ser separados. A extração do conhecimento de um conjunto de padrões é uma propriedade importante e desejada para os sistemas de reconhecimento de padrões, que a partir desse processo, podem realizar a tomada de decisão.

No modo de treinamento, o módulo de extração/seleção de características busca os atributos apropriadas para representar os padrões de entrada, já o classificador treinado é usado para dividir o espaço de características, para que no modo de classificação este seja capaz de determinar a qual classe um dado vetor ou padrão de entrada pertence. Classificadores de padrões podem ser classificados quanto a forma de aprendizado, em supervisionados, em que os padrões no treinamento apresentam-se com seus rótulos que indicam a qual classe os padrões pertencem, ou não-supervisionados em que no treinamento não se usa ou não se dispõe dos rótulos para o aprendizado e o número de classes muitas vezes deve ser também determinado (WATANABE, 1985).

2.5.1 *Support Vector Machine (SVM)*

O desenvolvimento desta teoria aplica-se não só à problemas de classificação de padrões, chamado de *Support Vector Classification* (SMOLA; SCHÖLHOPF, 1998), mas também em problemas de aproximação de funções denominado *Support Vector Regression* (BURGES; CHRISTOPHER, 1998).

O objetivo de aprendizagem dos classificadores SVM não está ligado apenas a diminuir o risco empírico, associado à minimização do erro relacionado aos padrões de treinamento, mas também a buscar a reduzir do risco estrutural que está associado à minimização do erro relacionado aos padrões de teste (aqueles que não foram vistos no processo de aprendizagem). Dessa forma, o processo de aprendizagem busca aumentar a capacidade de generalização diretamente na fase de treinamento.

O procedimento de aprendizagem é supervisionado e realizado com base nos diversos pares de entrada e saída pertencentes à base de dados. Os classificadores SVM possuem a característica comum que objetiva a resolução de problemas binários. Mesmo havendo formulação do SVM para problemas multiclases (CRAMMER; SINGER, 2001), tornando o processo de aprendizagem consideravelmente mais complexo, a partir da combinação das saídas apresentadas por classificadores binários têm sido amplamente utilizadas. As abordagens um-contra-um (DUAN; KEERTHI, 2005), um-contra-todos (DUAN; KEERTHI, 2005), DAGSVM (PLATT et al., 2000) e por códigos corretores de erros (DIETTERICH, 1995) são as mais utilizadas.

O objetivo de um classificador SVM é estimar uma função $f: \mathbb{R}^N \rightarrow \{\pm 1\}$ usando um conjunto:

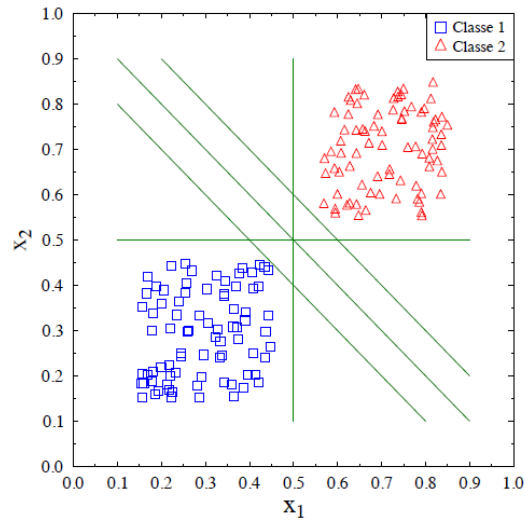
$$(x_1, d_1), \dots, (x_n, d_n) \in \mathbb{R}^N \times \{\pm 1\}, \quad (2.12)$$

em que x_i , representa o vetor de características e d_i , a classe da i -ésima amostra; e a função f classifica de forma correta outros padrões (x_i, d_i) não utilizados para a estimação da função, isto é, $f(x_i) = d_i$. O conjunto de dados $(x_1, d_1), \dots, (x_n, d_n) \in \mathbb{R}^N \times \{\pm 1\}$ usado para estimar f é denominado conjunto de treinamento. Enquanto o conjunto de dados $(\bar{x}_1, \bar{d}_1), \dots, (\bar{x}_n, \bar{d}_n) \in \mathbb{R}^N \times \{\pm 1\}$ com os exemplos não utilizados na estimação de f é denominado conjunto de teste. No contexto de classificação binária, ideia principal do SVM é estimar um hiperplano como superfície de separação ótima entre os exemplos (HAYKIN, 2001).

2.5.1.1 Hiperplano ótimo

A Figura 2.7, mostra os 5 hiperplanos encontrados por uma SVM para um problema linearmente separável, com duas classes hipotéticas com padrões descritos por quadrados para a classe positiva e triângulos para a classe negativa.

Figura 2.7 - Superfície ótima para separação entre duas classes.



FONTE: Rocha-Neto, 2011.

Considerando o problema linearmente separável mostrado na Figura 2.7, matematicamente, as soluções para este tipo de problema podem ser apresentadas na forma da equação de um hiperplano, ou seja:

$$w^T x + b = 0, \quad (2.13)$$

em que o vetor w é um vetor perpendicular ao hiperplano, x é o vetor de entrada e b é uma variável que permite que a margem do hiperplano seja maximizada. Nesse sentido, os hiperplanos que separam as classes devem satisfazer as seguintes restrições:

$$\begin{aligned} w^T x_i + b &\geq a \quad \text{para } d_i = +1, \\ w^T x_i + b &\leq a \quad \text{para } d_i = -1, \end{aligned} \quad (2.14)$$

em que $a > 0$ e $i = 1 \dots n$. Logo, as restrições acima devem ser satisfeitas para todos os padrões do conjunto de treinamento $(x_1, d_1), \dots, (x_n, d_n)$.

A Figura 2.8 ilustra o problema relacionado à escolha do melhor hiperplano de separação, a escolha se dá a partir da necessidade de não apenas minimizar o risco empírico como também minimizar o risco estrutural. Uma das possibilidades de escolha ilustrado em (b) resolve o problema para os padrões de treinamento, porém quando introduzido os padrões de teste, mostrado em (c), o hiperplano já não é a solução ideal. Uma solução ideal para este problema é o hiperplano apresentado em (d), visto que se posiciona equidistante das classes e,

desta maneira, os padrões de teste são classificados corretamente após o processo de treinamento.

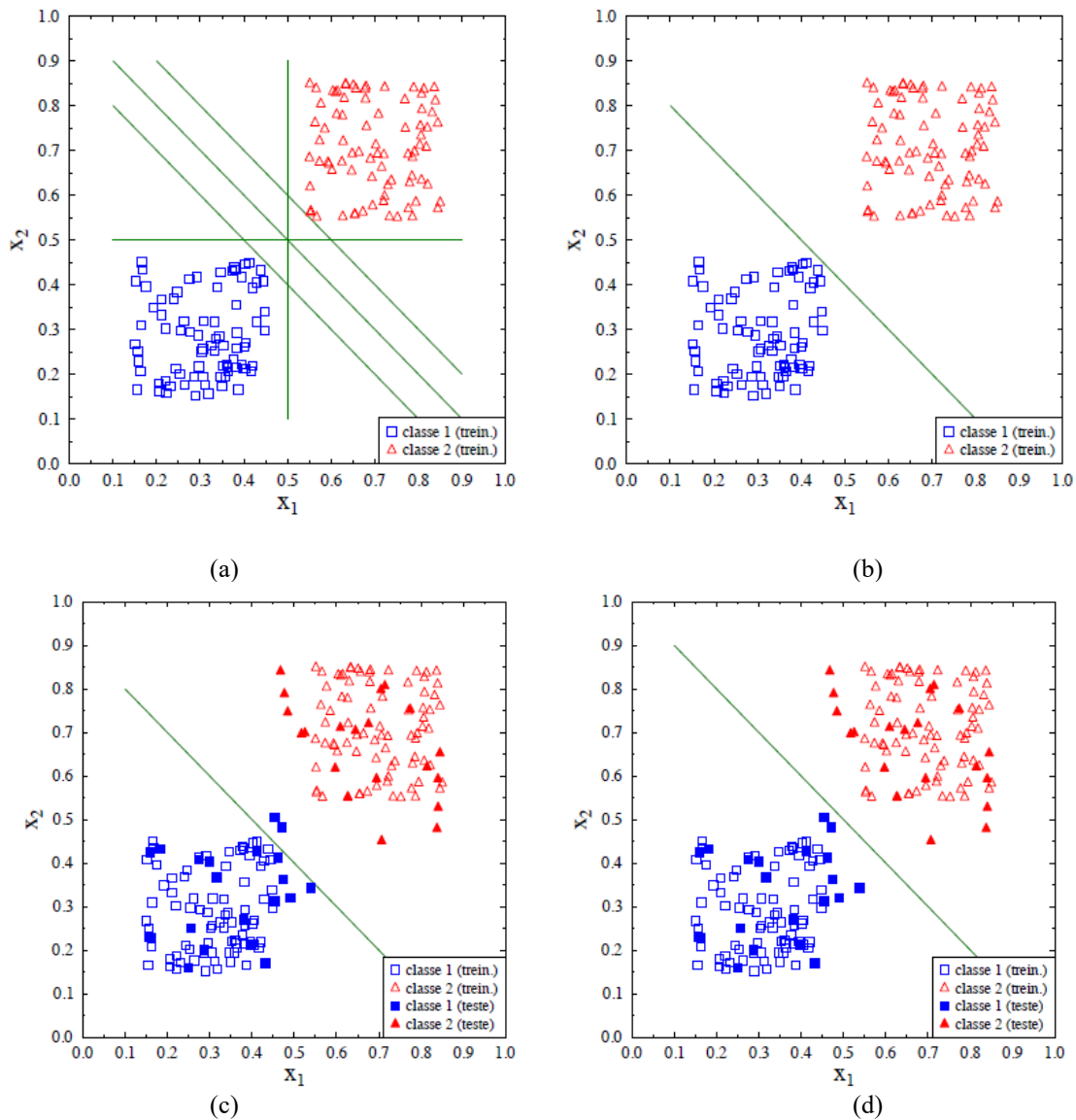
Desta forma, para os hiperplanos possíveis que solucionam o problema, a escolha deve ser um que tenha a máxima distância em relação aos padrões mais próximos do conjunto de treinamento.

Figura 2.7 - (a) Hiperplanos capazes de resolver o problema linearmente separável, com padrões de treinamento.

(b) Problema resolvido por um hiperplano muito próximo a uma das classes.

(c) O mesmo problema descrito em (a) com padrões de treinamento e teste.

(d) Problema resolvido por um hiperplano mais adequado equidistante das duas classes.



FONTE: Rocha-Neto, 2011.

Chamando-o de hiperplano ótimo, é representado por:

$$w_0^T x + b_0 = 0. \quad (2.15)$$

A menor distância entre o hiperplano ótimo e o padrão de treinamento mais próximo é chamado de margem de separação. Sendo o objetivo de uma máquina de vetor suporte, encontrar o hiperplano particular para o qual a margem de separação ρ , seja máxima através da função discriminante:

$$f(x) = w_0^T x + b_0, \quad (2.16)$$

a qual fornece uma medida algébrica da distância de x ao hiperplano ótimo (DUDA & HART, 1973). Para problemas de classificação de padrões, classificadores SVM utilizam a função sinal, isto é:

$$f(x) = \text{sign}(w_0^T x + b_0), \quad (2.17)$$

tal que:

$$f(x) = \begin{cases} -1, & \text{se } w_0^T x + b_0 < 0, \\ +1, & \text{se } w_0^T x + b_0 \geq 0. \end{cases} \quad (2.18)$$

Para uma solução ótima, o vetor de pesos ótimo w_0 e o viés ótimo b_0 devem ser encontrados a partir do conjunto de treinamento $\{x_i, d_i\}_{i=1}^n$. Para este fim, o problema apresentado na Equação 2.14 pode-se reescrever como:

$$\begin{aligned} w_0^T x + b_0 &\geq +1 & \text{para } d_i &= +1, \\ w_0^T x + b_0 &\leq -1 & \text{para } d_i &= -1. \end{aligned} \quad (2.19)$$

Os padrões particulares $\{(x^{(s)}, d^{(s)}) | s \in \{1 \dots N\}\}$ pertencem ao conjunto de treinamento $\{x_i, d_i\}_{i=1}^n$ e que satisfazem a Equação 2.19 com sinal de igualdade são denominados *vetores suporte*, ou seja]:

$$f(x^{(s)}) = \begin{cases} -1, & w_0^T x^{(s)} + b_0 = -1, \\ +1, & w_0^T x^{(s)} + b_0 = +1. \end{cases} \quad (2.20)$$

Considerando x_p como a projeção de x sobre o hiperplano ótimo, e r a distância do ponto x até o hiperplano. Portanto, um padrão x pode ser descrito como sendo

$$x = x_p + r \frac{w_0}{\|w_0\|}. \quad (2.21)$$

A distância r do ponto até o hiperplano pode ser obtida com base nas equações 2.16 e 2.21 e no conhecimento do valor de f no ponto x_p , $f(x_p) = 0$, da seguinte maneira:

$$f(x) = w_0^T x + b_0, \quad (2.22)$$

$$f(x) = w_0^T \left[x_p + r \frac{w_0}{\|w_0\|} \right] + b_0, \quad (2.23)$$

$$f(x) = w_0^T x + b_0 + w_0^T r \frac{w_0}{\|w_0\|}, \quad (2.24)$$

$$f(x) = 0 + r \frac{w_0^T w_0}{\|w_0\|}, \quad (2.25)$$

$$f(x) = r \|w_0\|, \quad (2.26)$$

$$r = \frac{f(x)}{\|w_0\|}. \quad (2.27)$$

A distância $r^{(s)}$ dos vetores suporte ao hiperplano ótimo pode ser obtida a partir da Equação 2.22 e da Equação 2.19 (a distância deve assumir valores positivos, desta forma, é desconsiderado o valor negativo), logo:

$$r^{(s)} = \frac{f(x^{(s)})}{\|w_0\|}, \quad (2.28)$$

$$r^{(s)} = \frac{1}{\|w_0\|}.$$

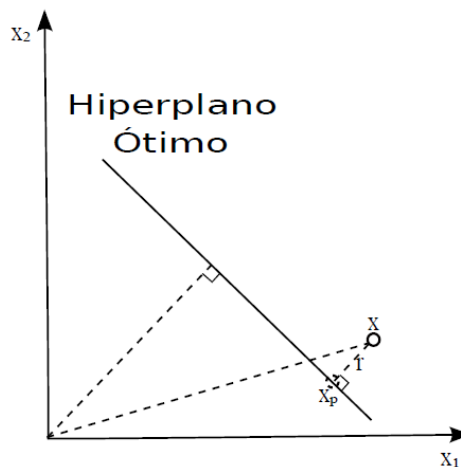
A distância geométrica de um dado padrão x ao hiperplano ótimo pode ser visualizada na Figura 2.9. Quando o padrão é um vetor suporte $x^{(s)}$, tem-se a distância $r^{(s)}$. Nesse sentido, a margem de separação ótima é:

$$\rho = 2r^{(s)}, \quad (2.29)$$

pode ser obtida substituindo a Equação 2.28, na Equação 2.29, tal que

$$\rho = \frac{2}{\|w_0\|} \quad (2.30)$$

Figura 2.8 - Interpretação geométrica de um padrão x ao hiperplano ótimo



FONTE: Rocha-Neto, 2011.

Baseado em Vapnik (1982) pode-se compreender que a maximização da margem de separação ρ implica, simultaneamente, na minimização da dimensão *Vapnik-Chervonenkis* (VC), associando à complexidade da função discriminante deve-se, portanto, adequar ao conjunto de treinamento. Analisando a Equação (2.30), verifica-se que ao se minimizar a norma $\|w\|$ do vetor de pesos, obtém-se a minimização da dimensão VC.

A partir deste raciocínio a formulação do problema de obtenção do hiperplano ótimo, pode começar a ser realizada a partir da Equação (2.30), com a maximização da margem de separação ρ :

$$\max \rho = \max \frac{2}{\|w\|} \quad (2.31)$$

Considerando a minimização da dimensão VC, também podendo ser representada como:

$$\min \|w\| \Leftrightarrow \min \sqrt{w^T w} \Leftrightarrow \min \frac{1}{2} w^T w, \quad (2.32)$$

ou seja, maximizar a margem de separação $\rho = \frac{2}{\|w\|}$ é o mesmo que minimizar a norma do vetor de pesos $\|w\|$. Portanto o hiperplano que separa os dados de entrada, pode ser descrito como um que minimize a função $\tau(w)$ a ser considerada a partir de agora:

$$\tau(w) = \frac{1}{2}w^T w, \quad (2.33)$$

desde que, satisfaça a restrição:

$$d_i(w^T x_i + b) \geq +1, \quad (2.34)$$

a qual é obtida pela combinação das duas linhas da Equação 2.19.

Então, é apresentado o problema clássico de obtenção dos parâmetros ótimos do classificador SVM como o seguinte problema de otimização:

$$\begin{aligned} \min \tau(w) &= \min \frac{1}{2}\|w\|^2 = \min \frac{1}{2}w^T w, \\ \text{s. a. } d_i[(w^T x_i) + b] &\geq 1, \quad i = 1, \dots, n \end{aligned} \quad (2.35)$$

em que $\tau(w)$ representa descreve a função-custo a ser minimizada.

O problema mostrado na Equação 2.35 baseia-se na linearidade das classes, ou seja, assume-se que as duas classes sejam totalmente separáveis por um único hiperplano, denominado problema de otimização primal (ROCHA-NETO, 2011).

2.5.1.2 Classificador SVM com margem rígida (linearmente separável)

Para a formulação de restrição apresentada na Equação 2.30, resulta no classificador denominado SVM com margem rígida, ou seja, ideal para problemas que sejam linearmente separáveis. Sendo a função $\tau(w)$, uma função convexa em w , enquanto as restrições são lineares em w , pode-se usar o método dos multiplicadores de Lagrange construindo a função lagrangeana:

$$L(w, b, \alpha) = \frac{1}{2}w^T w - \sum_{i=1}^n \alpha_i [d_i(w^T x_i + b) - 1], \quad (2.36)$$

em que os multiplicadores de Lagrange $\{\alpha_i\}_{i=1}^n$ são grandezas não-negativas (i.e. $\{\alpha_i \geq 0\}_{i=1}^n$).

A forma expandida da função lagrangeana apresenta-se como:

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n [\alpha_i d_i w^T x_i + \alpha_i d_i b - \alpha_i], \quad (2.37)$$

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i d_i w^T x_i - \sum_{i=1}^n \alpha_i d_i b + \sum_{i=1}^n \alpha_i, \quad (2.38)$$

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i d_i w^T x_i - b \sum_{i=1}^n \alpha_i d_i + \sum_{i=1}^n \alpha_i. \quad (2.39)$$

A solução é determinada pelo ponto de sela da função $L(w, b, \alpha)$, que deve ser minimizada em relação a w e b e maximizada em relação a α . Assim, diferenciado em relação a w e b e igualando a zero, obtém-se as seguintes condições de otimização:

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0, \quad (2.40)$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0, \quad (2.41)$$

resultando em:

$$w = \sum_{i=1}^n \alpha_i d_i x_i \quad (2.42)$$

e

$$\sum_{i=1}^n \alpha_i d_i = 0, \quad (2.43)$$

respectivamente.

Desta maneira, nota-se que o terceiro termo $(-b \sum_{i=1}^n \alpha_i d_i)$ da forma expandida da função apresentada na Equação 2.32 é zero, devido a condição obtida na Equação 2.43. Reescrevendo a Equação 2.39:

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i d_i w^T x_i + \sum_{i=1}^n \alpha_i. \quad (2.44)$$

Ao substituir os valores de w , obtidos na Equação 2.42, na Equação 2.44, tem-se que:

$$L(w, b, \alpha) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j x_i^T x_j - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j x_i^T x_j + \sum_{i=1}^n \alpha_i, \quad (2.45)$$

$$L(\alpha) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j x_i^T x_j + \sum_{i=1}^n \alpha_i. \quad (2.46)$$

Portanto, o problema de otimização dual é formulado como:

$$\begin{aligned} \max L(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j x_i^T x_j, \\ \text{s. a. } &\sum_{i=1}^n \alpha_i d_i = 0, \\ \text{s. a. } &\alpha_i \geq 0 \text{ para } i = 1, \dots, n. \end{aligned} \quad (2.7)$$

No problema de resolução do problema dual são obtidos os multiplicadores de Lagrange ótimos $\{\alpha_i^0\}_{i=1}^n$, w_0 e b_0 , podem ser escritos como:

$$w_0 = \sum_{i=1}^n \alpha_i^0 d_i x_i \quad (2.48)$$

e

$$b_0 = 1 - w_0^T x, \quad (2.49)$$

A função discriminante definida no problema primal para o hiperplano ótimo descrita pela Equação 2.16, pode-se reescrita com base no problema dual, aplicando-se a Equação 2.48, resultando em:

$$f(x) = \sum_{i=1}^n \alpha_i^0 d_i x_i^T x + b_0. \quad (2.50)$$

2.5.1.3 Classificador SVM com Margem Flexível (não linearmente separável)

Em problemas reais, nem sempre é possível obter um hiperplano que separe, sem erros, todos os padrões pertencentes à duas classes distintas, principalmente quando ocorre sobreposição entre dados que compõem as classes, fazendo necessário uma formulação para o problema do classificador SVM que considere tal dificuldade, permitindo que alguns padrões sejam incorretamente classificados. Uma das justificativas desta situação é impedir que a função discriminante se torne muito complexa. Para diminuir a complexidade da função permite-se alguns erros com a intenção de obter melhor desempenho. Com problemas em que existem padrões fora de seus valores típicos, chamados de padrões discrepantes, facilmente percebe-se que essa situação se aplica. Outra justificativa para a flexibilidade da margem é evitar o sobreajuste da superfície de decisão aos dados (*overfitting*).

A Equação 2.51 permite um relaxamento das restrições dos classificadores SVM com margens rígidas com base em um limiar, ou seja

$$d_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, \dots, n, \quad (2.51)$$

em que $\xi_i \geq 0$, $i = 1, \dots, n$. Os limites ξ_i são chamados de variáveis de folga.

No processo de aprendizagem, as variáveis de folga assumem valores não-negativos, $\xi_i \geq 0$. Os vetores de treinamento que não são vetores suporte, após o processo de aprendizagem, assumem o valor igual a zero. Os que ultrapassam a margem de separação, mas ainda permanecem do lado correto em relação ao hiperplano de separação ótimo, assumem valores no intervalo, $0 < \xi_i \leq 1$. Os padrões que estão localizados opostos à sua classe e classificados incorretamente, devido à sua posição em relação ao hiperplano ótimo possuem valores $\xi_i > 1$.

O problema de otimização (primal) para classificadores SVM com margem flexível é formulado como:

$$\begin{aligned} \min \tau(w, \xi) &= \min \left\{ \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \right\} \\ \text{s. a. } d_i[(w^T x_i) + b] &\geq 1 - \xi_i, \quad i = 1, \dots, n \\ \text{s. a. } \xi_i &\geq 0, \quad i = 1, \dots, n \end{aligned} \quad (2.52)$$

onde a constante C é um parâmetro que faz uma regularização entre o primeiro $\left(\frac{1}{2} w^T w\right)$ e segundo termo $(C \sum_{i=1}^n \xi_i)$ da função-custo. Objetivando minimizar os valores assumidos pelas variáveis de folga, consequentemente minimizando os erros permitidos.

A solução é determinada pelo ponto de sela da função lagrangeana $L(w, b, \xi, \alpha, \beta)$, a qual deve ser minimizada em relação a w, b e ξ e maximizada em relação a α e β , semelhante ao realizado para o classificador SVM com margem rígida, sendo:

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (d_i (x_i^T w + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \xi_i, \quad (2.53)$$

em que os termos $\alpha = \{\alpha_i\}_{i=1}^n$, $\beta = \{\beta_i\}_{i=1}^n$ e $\xi = \{\xi_i\}_{i=1}^n$ possuem valores não negativos. Expandindo a função lagrangeana da Equação 2.53, tem-se:

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i d_i w^T x_i - b \sum_{i=1}^n \alpha_i d_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n \beta_i \xi_i. \quad (2.54)$$

Para a solução do problema, a função-custo deve ser diferenciada em relação a w, b e ξ e igualada a zero, para se obter as condições de otimização a seguir:

$$\frac{\partial L(w, b, \xi, \alpha)}{\partial w} = 0, \quad (2.55)$$

$$\frac{\partial L(w, b, \xi, \alpha)}{\partial b} = 0, \quad (2.56)$$

$$\frac{\partial L(w, b, \xi, \alpha)}{\partial \xi} = 0. \quad (2.57)$$

A primeira condição da Equação 2.55, ao ser resolvida, resulta em:

$$w = \sum_{i=1}^n \alpha_i d_i x_i. \quad (2.58)$$

Enquanto a condição da Equação 2.56, resulta em:

$$\sum_{i=1}^n \alpha_i d_i = 0. \quad (2.59)$$

Por fim, ao resolver a Equação 2.57, tem-se:

$$C = \alpha_i + \beta_i . \quad (2.60)$$

Substituindo a Equação 2.60 na Equação 2.53 e rearranjando seus termos, tem-se:

$$L(w, b, \xi, a, \beta) = \frac{1}{2} w^T w + \sum_{i=1}^n \xi_i (\alpha_i + \beta_i) - \sum_{i=1}^n \alpha_i d_i w^T x_i - b \sum_{i=1}^n \alpha_i d_i + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \xi_i (\alpha_i + \beta_i) . \quad (2.61)$$

Verifica-se que o segundo e o último termo podem ser eliminados, resultando na Equação 2.62

$$L(w, b, a) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i d_i w^T x_i - b \sum_{i=1}^n \alpha_i d_i + \sum_{i=1}^n \alpha_i . \quad (2.62)$$

Sabendo que o terceiro termo é zero, segundo a Equação 2.59, a Equação 2.62 pode ser reescrita da seguinte maneira:

$$L(w, b, a) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i d_i w^T x_i + \sum_{i=1}^n \alpha_i . \quad (2.63)$$

Aplicando o resultado obtido na Equação 2.58 no segundo termo da Equação 2.63, tem-se $-w^T w$. Logo:

$$L(w, b, a) = \frac{1}{2} w^T w - w^T w + \sum_{i=1}^n \alpha_i , \quad (2.64)$$

$$L(w, b, a) = -\frac{1}{2} w^T w + \sum_{i=1}^n \alpha_i . \quad (2.65)$$

Para que a função não dependa de w , substitui-se o resultado da Equação 2.58 na Equação 2.65, obtém-se então:

$$L(\alpha) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j x_i^T x_j + \sum_{i=1}^n \alpha_i , \quad (2.66)$$

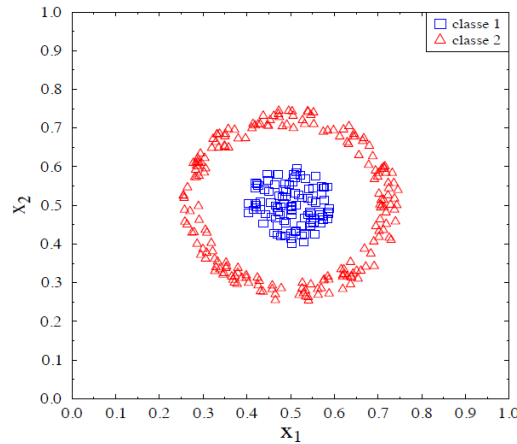
Então, o problema de otimização dual para o classificador SVM com margem flexível é dado por:

$$\begin{aligned} \max L(\alpha) = \max & \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j x_i^T x_j \right\}, \\ \text{s. a. } & \sum_{i=1}^n \alpha_i d_i = 0, \\ \text{s. a. } & 0 \leq \alpha_i \leq C \text{ para } i = 1, \dots, n. \end{aligned} \quad (2.67)$$

Analisando a Equação 2.67 percebe-se apenas uma diferença em relação à Equação 2.47 que é a limitação superior dos multiplicadores de Lagrange $0 \leq \alpha_i \leq C$. Os outros termos se mantêm inalterados. Isto acontece em decorrência da relação $\alpha_i + \beta_i = C$ apresentada na Equação 2.60. Como $\beta_i \geq 0$, isto é, o menor valor assumido por β é zero quando $\alpha_i = C$.

O uso da margem flexível pelo classificador SVM pode haver situações em que o problema não é resolvido de forma satisfatória. Um exemplo do tipo de problema é ilustrado na Figura 2.10, que mesmo tolerando alguns erros de classificação, traz um problema de classificação hipotética, de natureza não linearmente separável.

Figura 2.9 - Problema hipotético binário não linearmente separável



FONTE: Rocha-Neto, 2011.

2.5.1.4 Função Kernel

O conceito chave para se manipular problemas não linearmente separáveis é usar um mapeamento não linear a fim de se obter um problema linear a partir de um não linear, o conjunto de dados deve ser transformado para um espaço de características de elevada dimensão. O espaço de entrada de um padrão $x_i \in \mathbb{R}^N$ é mapeado para um espaço de

características $\phi(x) \in \mathbb{R}^M$, tal que $M > N$. A Figura 2.10 exemplifica um problema binário, não linearmente separável.

Para qualquer função contínua e simétrica $K(x, y)$ que satisfaça o teorema de Mercer (MERCER, 1909), há um espaço de Hilbert H , um mapeamento $\phi: \mathbb{R}^N \rightarrow H$ e um número $\beta_i > 0$, sendo:

$$K(x, y) = \sum_{i=1}^M \beta_i \bar{\phi}_i(x) \bar{\phi}_i(y), \quad (2.68)$$

em que $x, y \in \mathbb{R}^N$ e M é a dimensionalidade do espaço de Hilbert. O teorema de Mercer exige que para qualquer função quadrada integrável $g(\cdot)$, tal que $g(\cdot) \neq 0$,

$$\int K(x, y) g(x) g(y) dx dy \geq 0. \quad (2.69)$$

Definindo $\phi_i(\cdot) = \bar{\phi}_i(\cdot) \sqrt{\beta_i}$, pode-se escrever a Equação 2.68 como:

$$K(x, y) = \sum_{i=1}^M \bar{\phi}_i(x) \sqrt{\beta_i} \bar{\phi}_i(y) \sqrt{\beta_i}, \quad (2.70)$$

$$K(x, y) = \sum_{i=1}^M \phi_i(x) \phi_i(y), \quad (2.70)$$

de modo que a função *kernel* $K(x, y)$ pode ser representada pelo produto interno dos vetores no espaço de características, sendo

$$K(x, y) = \sum_{i=1}^M \phi_i(x) \phi_i(y) \quad (2.71)$$

$$K(x, y) = (\phi_i(x), \phi_i(y)) \quad (2.72)$$

$$K(x, y) = \phi_i^T(x) \phi_i(y). \quad (2.73)$$

A Equação 2.73 é conhecida como truque de *kernel* (*Kernel Trick*), permite que se evite o uso explícito de características de elevada dimensão. O truque de *kernel* permite que não seja necessário realizar o mapeamento para se obter o vetor $(\phi_i(x))$, desde que se conheça a função que descreva o produto interno $(\phi_i(x), \phi_i(y))$.

Pode-se, então, redefinir as Equações 2.47 e 2.67 referentes ao classificador SVM com margem rígida e ao SVM com margem flexível como:

$$\begin{aligned} \max L(\alpha) &= \max \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j K(x_i, x_j) \right\}, \quad (2.74) \\ \text{s. a. } \sum_{i=1}^n \alpha_i d_i &= 0, \\ \text{s. a. } \alpha_i &\geq 0 \text{ para } i = 1, \dots, n, \end{aligned}$$

$$\begin{aligned} \max L(\alpha) &= \max \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j K(x_i, x_j) \right\}, \quad (2.75) \\ \text{s. a. } \sum_{i=1}^n \alpha_i d_i &= 0, \\ \text{s. a. } 0 \leq \alpha_i &\leq C \text{ para } i = 1, \dots, n. \end{aligned}$$

Assim, como a Equação 2.50, que descreve a função discriminante no espaço de características, pode ser reescrita como:

$$f(x) = \sum_{i=1}^n \alpha_i^0 d_i K(x_i, x) + b_0. \quad (2.76)$$

Desta maneira, a classe para um dado vetor de teste x pode ser determinada a partir da função sinal, ou seja

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i^0 d_i K(x_i, x) + b_0 \right), \quad (2.77)$$

Os *kernels* mais comumente utilizados são o linear, polinomial e gaussiano (RBF), mostrados na Tabela 2.1. Neste trabalho, utiliza-se o kernel gaussiano.

Tabela 2.1 Principais *Kernels*

<i>Kernel</i>	Descrição
Linear	$K(x_i, x) = x_i^T \cdot x$
Polinomial	$K(x_i, x) = (x_i^T \cdot x + 1)^d$ em que d é o grau do polinômio
Gaussiano (RBF)	$K(x_i, x) = \exp\{-\ x - x_i\ ^2 / \sigma^2\}$ em que σ é uma constante

FONTE: Adaptado Rocha-Neto, 2011.

2.6 ESTADO DA ARTE

É possível encontrar na literatura diversas técnicas utilizadas para detectar e classificar disfunções vocais. O interesse na análise qualitativa da voz utilizando de processamento de sinais motivado pelo bem estar do paciente, pelo fato de ser um método de diagnóstico não invasivo, tem crescido nos últimos anos. A maioria dos sistemas de detecção e classificação das patologias laringeas usam a vogal sustentada /a/, por ela não se afeta por outros fenômenos não vocais, já que em um contexto fonético é uma configuração em que a laringe, ou seja, ou músculos supralaríngeos permanecem relativamente estáticos, com o mínimo de variação durante a produção da onda sonora (Maryn & Roy 2012, Martins et al. 2015).

Com um trabalho de grande relevância, na dissertação de Rosa (1998), intitulada “Análise acústica da voz para pré-diagnóstico de patologias da Laringe”, é descrito um conjunto de medidas acústicas baseada no sinal de voz, e a partir da avaliação estatística, empregando redes neurais artificiais, definiu um método automático para identificação probabilística das patologias e obteve taxa de acerto máxima de 81%. Foi amplamente usada como referência para trabalhos posteriores que usaram das mesmas medidas acústicas para classificação de patologias com diferentes ferramentas de classificação, inclusive usada como referência para sua tese de doutorado em 2002.

Rosa ainda publicou diversos artigos sobre o mesmo tema entre eles no trabalho com Sodré, e Dassi-Leite (2017). Neste trabalho, os autores relacionaram 19 medidas acústicas a 23 condições da laringe por redes neurais artificiais avaliando o uso de redes neurais e medidas acústicas para identificar patologias laringeas, tendo como resultados 90,3% de sensibilidade e especificidade alcançados com apenas 6 medições extraídas

Frazão et al (2013) trata da avaliação de desempenho de características temporais de sinais de vozes no pré-diagnóstico de patologias laringeas (edema de Reinke, paralisia e nódulos vocais). As medidas temporais usadas foram frequência fundamental, *jitter*, *shimmer*, energia segmental, autocorrelação e as ferramentas de classificação matemática usadas foram a análise discriminante com as funções linear (LDA) e quadrática (QDA). A medida temporal que apresentou os melhores resultados na classificação foi *shimmer* percentual, chegando a uma precisão de 100% entre sinais saudáveis ou com paralisia, como também para sinais com edema de Reinke ou com nódulo vocal.

Fontes et al (2014) investigou o uso de uma ferramenta matemática baseada na teoria da informação, chamada função de autocorrentropia, na análise de sinais de voz. Foi

apresentada uma arquitetura, baseado em distância Euclidiana, para a classificação entre voz normal e patológica e também entre patologias (edema e nódulo). A função de autocorrentropia, se mostrou eficiente na extração de características de um sinal de voz.

Al-Nasheri et al (2018) utilizaram métodos baseados em autocorrelação e entropia para extrair características de um sinal de voz, no intuito de detectar e classificar patologias vocais. A estratégia busca investigar o sinal de voz em diferentes regiões de frequência. O sinal de voz passa por oito filtros passa-faixa do tipo FIR (Resposta Finita ao Impulso), com frequências centrais espaçadas em escala oitavada. Sobre o sinal de saída de cada filtro calcula-se a autocorrelação e a entropia. A partir dos valores obtidos, formam-se dois vetores de características classificados individualmente por uma SVM.

Teixeira (2019) utilizou três grupo de parâmetros, o grupo I (a), contém *jitter* relativo, *shimmer* relativo e *Harmonic to Noise Ratio*, determinados em segmentos de fala estacionária, onde se atingiu 80,7% de exatidão para distinguir saudáveis e patológicos com SVM. O grupo I (b), contém os parâmetros do grupo I(a), *Noise to Harmonic Ratio* e Autocorrelação determinados em segmentos de fala estacionária, onde se atingiu 79,2% de exatidão para distinguir saudáveis e patológicos com SVM. O grupo II é baseado em *Mel Frequency Cepstral Coefficients* (MFCC's), determinados nos segmentos de fala estacionários, onde se atingiu 83,3% de exatidão para distinguir saudáveis e laringite com SVM. O grupo III é formado por coeficientes MFCC's extraídos de fala contínua onde se atingiu 71% de exatidão para distinguir saudáveis e patológicos com Redes Neurais.

Em dezembro de 2020, na Escola Superior de Saúde - Politécnico do Porto, Faria (2020) propôs uma avaliação complementar da voz através de medidas acústicas de longo termo em vozes disfônicas. Realizou um estudo analítico, observacional, quantitativo transversal, em que doze indivíduos, com e sem disfonia, foram sujeitos a avaliação por otorrinolaringologista, gravação da voz (vogal sustentada e leitura em diferentes intensidades) e avaliação áudio-percetiva (por três peritos, usando a escala CAPE-V [*Consensus Auditory Perceptual Evaluation of Voice*] adaptada), ficando distribuídos por três graus: sem alterações, ligeiro grau e moderado grau de disfonia. Foram calculadas medidas de curto e longo termo e da análise estatística (correlação de Pearson) e concluiu que tarefas de leitura em diferentes intensidades são importantes na avaliação acústica, verificando-se correlação inversa entre a dinâmica de intensidades no harmônicos que reflete na contribuição tanto na região fonte glótica como do trato vocal, *Cepstral Peak Prominence Smoothed* (CPPSfala e CPPSvogal) que diz respeito à diferença entre o pico cepstral mais proeminente e o grau de disfonia.

Capítulo 3

MATERIAIS E MÉTODOS

Neste capítulo, são apresentados os processos utilizados para a caracterização acústica e classificação de vozes afetadas por patologias da laringe, mais especificamente Disfonia e Edema de Reinke. Será mostrado como os algoritmos retiram características acústicas das vozes, criando um banco de dados a partir das medidas acústicas, para posterior classificação do algoritmo de reconhecimento de padrões, resultando, assim, numa hipótese de diagnóstico.

3.1. MATERIAIS

Todas as simulações utilizadas foram implementadas utilizando o ambiente *Matlab R2013a*, com a licença nº 28878-48087-65497-12872-42979-61998-52336-35817-09598 do laboratório de Processamento de Sinais do IFCE - Campus Fortaleza e sua respectiva linguagem de programação nativa.

A base de dados usada neste trabalho foi desenvolvida pelo *Department of Language Science and Technology (LST)* do *Institut für Phonetik* da *Universität des Saarlandes (UdS)*, localizada em *Saarbrücken* na Alemanha. Chamada *Saarbruecken Voice Database (SVD)*³, inclui gravações de vozes de mais de 2000 pessoas, entre saudáveis e com diferentes tipos de patologia, cada sessão de gravação contém as vogais sustentadas /a/, /i/, /u/ produzidas

³ Disponível em: < <http://stimmdb.coli.uni-saarland.de/> >

em entonação neutra, alto, baixo e em frequências crescentes, também possui gravação da frase “*Guten Morgen, wie geht es Ihnen?*” (“Bom dia, como vai você?”) (Al-Nasheri et al., 2018).

Inclui amostras de pacientes com uma larga variedade de desordens vocais por causas orgânicas, neurológicas, traumáticas, psicogênicas entre outras. Todas as amostras foram coletadas em um ambiente controlado com as seguintes características: baixo nível de ruído, distância constante do microfone, taxa de amostragem de 50kHz com resolução de amplitude de 16 bits (WOKUREK e PUTZER, 2017).

Para este trabalho, foram consideradas as vozes com a vogal sustentada /a/ com entonação neutra nos seguintes casos:

- Vozes patológicas – 172 amostras:
 - Edema de Reinke: 75 vozes (59 femininas e 16 masculinas);
 - Disfonia: 97 vozes (52 amostras femininas e 45 masculinas);
- Vozes saudáveis – 614 amostras:
 - Feminina: 384 vozes
 - Masculina: 230 vozes

3.2. METODOLOGIA

O processo de discriminação das vozes para classificação das patologias inicia-se com o download dos arquivos dos sinais de das vozes com a extensão .wav, na base de dados do *site* ilustrado na Figura 3.1. Nele foi possível escolher quais as patologias e características físicas de cada paciente como sexo, idade, saudável e patologias.

Para este trabalho foram escolhidos os pacientes femininos e masculinos de idade variada, saudáveis e com as patologias Edema de Reinke ou Disfonia. É importante pontuar algumas informações relevantes. As vozes dos pacientes escolhidos tinham as patologias Edema de Reinke OU Disfonia. Essa observação é importante visto que existem pacientes que apresentam essas duas e outras patologias associadas com agravamento da fonação.

Figura 3.1 - Página de download da base de dados SDV

version 2.0 **Saarbruecken Voice Database** 

[Help](#) [Logout](#) [Accept](#) [Reset](#) [Export](#)

Database request

Speaker:	☐ male ☐ female Age: No.:
Recording session:	☐ healthy ☐ pathological No.:
Pathologies and Diagnosis:	Obligatory pathologies: ← → Excluded pathologies: ← → Selection: - Amyotrophe Lateralsklerose - Aryluxation - Balbuties - Bulbärparalyse - Carcinoma in situ - Chondrom - Chordektomie - Cyste - Diplophonie Remarks w.r.t. diagnosis: Extension of request
Sorting criteria:	Speaker number Date of recording Sex of the speaker ☐ descending order ☐ descending order ☐ descending order

[Help](#) [Logout](#) [Accept](#) [Reset](#) [Export](#)

[Manfred Pützer](#)  [William J. Barry](#)  Responsible administrator 
[Institut für Phonetik](#)
[Universität des Saarlandes](#)
[Impressum](#) Supported by [Deutsche Forschungsgemeinschaft, Bonn](#)
(BA 737/9-6)

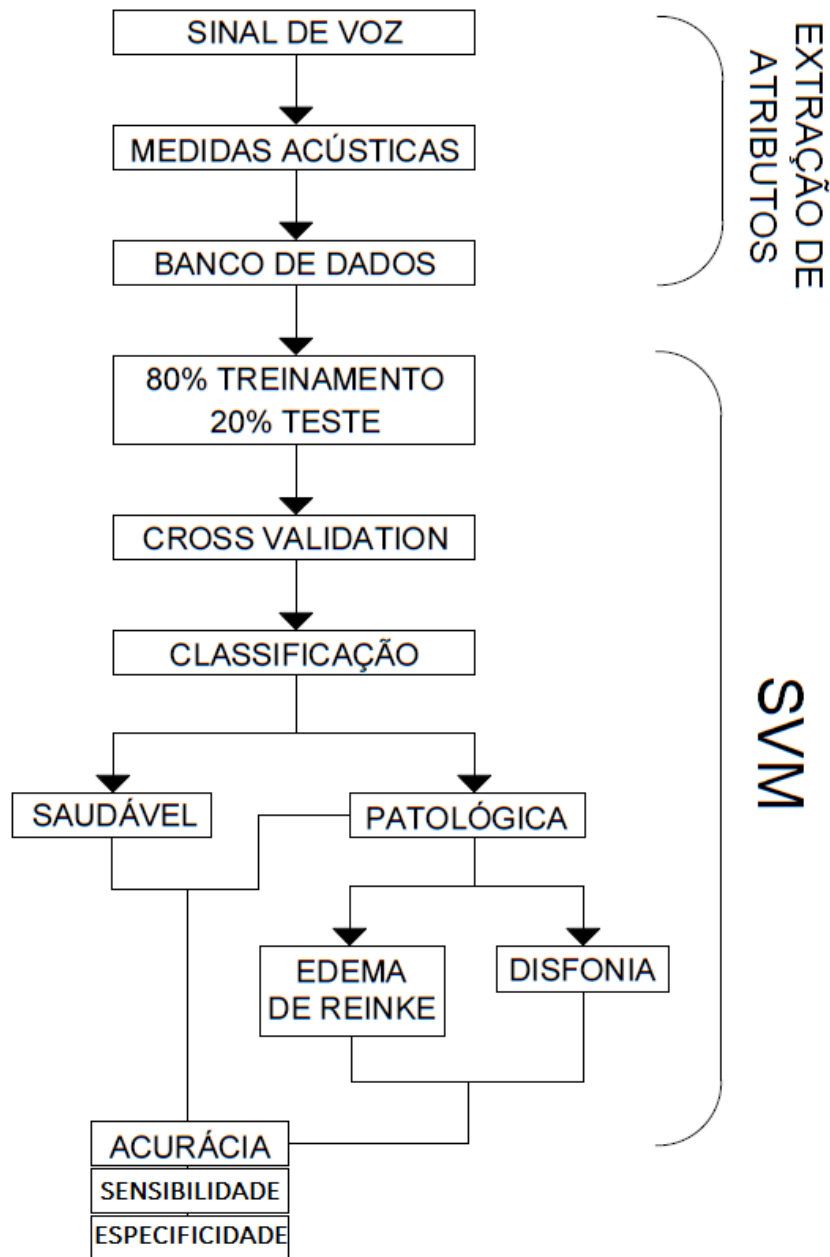
FONTE: *Saarbruecken Voice Database*

Outro ponto importante é que as vozes femininas e masculinas foram separadas em diretórios diferentes pelo motivo de que o algoritmo utilizado trabalha com as frequências e a caracterização das vozes femininas e masculinas possuem suas particularidades, fazendo com que, ao serem “misturadas” o algoritmo não as classificava corretamente, este ponto será mais bem detalhado no item 3.2.3.

3.2.1. Diagrama de blocos do sistema de extração de atributos e classificação

Na Figura 3.2, é apresentado o diagrama de blocos do sistema de extração de atributos das vozes e do sistema de classificação.

Figura 3.2 – Visão geral do sistema de extração de atributos e classificação SVM



Fonte: Autora

A entrada para o sistema são os sinais de áudio das vozes no sistema de extração de atributos é feito o processamento deste sinal para retirar características através de ferramentas matemáticas para a criação de um banco de dados, este que por sua vez é parâmetro de entrada para o sistema de classificação SVM, tendo como resultado a acurácia, sensibilidade e especificidade. Assim como o autor Rocha-Neto (2011) utiliza em sua tese, para a etapa de treinamento, foram usadas 80% dos padrões do banco de dados e 20% para o teste, como ilustrado no diagrama de blocos da Figura 3.2.

3.2.2. Extração de atributos

O sistema de extração de atributos do sinal de voz, tem como objetivo uma nova representação, em outro espaço de menor dimensão para um menor custo computacional, gerando uma matriz com os resultados calculados de cada um dos atributos apresentados no Capítulo 2.

Como a gravação do sinal de áudio da base de dados é de alta resolução, isto é, alta qualidade, não foi necessário fazer uma pré-filtragem para retirar possíveis ruídos.

Inicialmente feita a leitura do sinal de áudio da voz em ‘.wav’ feminina ou masculina separadas em diretórios diferentes, a partir do diretório **D** (conjunto com todas as amostras) e seus valores de amplitude no domínio do tempo são armazenados em um vetor, neste vetor é aplicada a FFT para extração das características no domínio da frequência deste sinal e calculado seu valor em dB.

Um procedimento de limpeza nos dados foi feito para remoção de amostras discrepantes, aqueles que apresentaram falhas e/ou erros de cálculo para valores fora da faixa de frequência para a melhorar eficiência do algoritmo, em seguida foram separadas a primeira e segunda formantes da voz a partir deste vetor de frequências e armazenados seus valores f_0 , f_1 e f_2 no vetor de formantes.

O próximo passo foi calcular a energia total (e_0) deste sinal de voz, onde é feito o somatório das intensidades das frequências do vetor de frequências, em seguida foram calculadas as energias e_1 e e_2 e a relação entre elas (e_1/e_2), todos os valores armazenados no vetor energia e_0 , e_1 , e_2 , e_1/e_2 .

Por fim, é calculado o grau de desordem do sinal através da entropia que é o oitavo e último descritor escolhido para o padrão, repetido o mesmo processo para todas as amostras do diretório formando uma matriz de atributos.

O último passo foi criar um arquivo em “.txt” com os dados da matriz de atributos a fim de criar um banco de dados onde cada linha é a representação numérica de uma voz e cada coluna representa um atributo que descreve suas características.

3.2.3. Classificador

Este classificador é capaz de diferenciar as vozes saudáveis e as patologias Edema de Reinke e Disfonia de indivíduos do mesmo sexo. Essa separação das vozes femininas e

masculinas foi necessária pelo fato de que este trabalho trata com características de frequência, como a voz feminina tem suas características embasadas em frequências mais altas e a masculina com frequências mais baixas, ao mesclar ambas as vozes o algoritmo não é capaz de fazer a correta classificação, visto que ele busca características específicas naquela região de frequência.

Os atributos extraídos na etapa anterior serão utilizados como entrada para o SVM, que atua como método de reconhecimento de padrões. O processo de classificação das vozes patológicas está dividido em duas etapas: treinamento e teste.

O classificador é dividido em duas máquinas SVMs, onde a primeira máquina, chamada de SVM1, apresenta resultados referentes ao processo de classificação entre vozes saudáveis e patológica, na qual este tipo de classificação em um ambiente clínico é chamado de triagem. Já a segunda máquina, chamada SVM2, classifica entre as patologias Edema de Reinke e Disfonia. As características selecionadas dos dados de entrada são atribuídas às máquinas, elas são treinadas e validadas resultando numa hipótese do diagnóstico.

Para cada máquina as etapas de treinamento e teste funcionam da mesma maneira. Na primeira etapa, o treinamento da máquina de vetor suporte é realizado através do aprendizado supervisionado e sua utilização nesse trabalho justifica-se devido algumas características particulares, como a detecção automática dos padrões mais relevantes, chamados de vetores suporte e o mapeamento implícito dos padrões em dimensões elevadas, através da função *kernel*, como descrita sua teoria no Capítulo 2 deste trabalho.

Foi utilizado um processo de *Cross Validation* (CV) que permite dividir uma estrutura de mineração em seções cruzadas e iterativamente treinar e validar os padrões com cada seção cruzada, especificando um conjunto de dados para o treinamento e um conjunto para teste. São criadas k partições aleatórias de dados cujas partições possuem aproximadamente a mesma quantidade de dados. É utilizado apenas um subconjunto para teste o restante é para treino. Sendo assim o classificador é treinado e testado k vezes com diferentes partições de dados e a cada uma destas interações são obtidas as medidas de desempenho do classificador, ao final dos testes são calculadas a média e o desvio padrão dos resultados das métricas referentes a cada interação. Neste trabalho é empregada a técnica de validação cruzada *k-fold* com o valor de k igual a 5.

O *kernel* utilizado em todas as máquinas é o RBF (Função de Base Radial), pois quando comparado com o polinomial obteve-se melhor performance.

Para cada classificador são realizadas 50 rodadas de treinamento/teste. Os padrões de treinamento são selecionados aleatoriamente, sendo as restantes usadas para teste. São

determinadas a taxa média de acerto (acurácia) e o desvio-padrão das taxas de acerto observadas durante os testes.

A remoção dos padrões discrepantes foi feita a partir da faixa de frequência estipulada de 170 a 270 Hz para vozes femininas e 80 e 160 Hz para vozes masculinas.

Além das taxas das médias de acerto (acurácia) e seus desvios-padrões, os classificadores são também comparados entre si com base no número de falsos negativos e falsos positivos. Dessa forma, é possível calcular a sensibilidade e especificidade do classificador a partir da matriz de confusão relacionando esse estudo de falso negativo e falso positivo com a classificação dando uma hipótese de diagnóstico ao paciente.

A matriz de confusão é uma importante ferramenta que determina se o valor classificado corresponde ou não ao valor real, de tal forma que todos os casos em cada classe são aferidos e os mesmos são exibidos na matriz. A Tabela 3.1 mostra a estrutura de uma matriz de confusão que fornece as possíveis situações encontradas em um teste de classificação binária.

Tabela 3.1- Estrutura da matriz de confusão

	Saudável	Patológica
Saudável	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Patológica	Falso Negativo (FN)	Verdadeiro Negativo (VN)

FONTE: Autora

A acurácia baseada em matriz de confusão é definida na Equação 3.1. A acurácia é a taxa de acerto obtida durante a classificação na etapa de teste, desse modo mensurando a proporção de predições corretas (KOHAVI, PROVOST, 1998). Podendo ser calculada por:

$$Acurácia = \frac{VP+VN}{VP+VN+FP+FN} . \quad (3.1)$$

Já sensibilidade baseada em matriz de confusão é definida pela equação

$$Sensibilidade = \frac{VP}{VP+FN} \quad (3.2)$$

mensurando a capacidade que o classificador possui em excluir corretamente aqueles sinais que não são afetados por patologias (VIEIRA, 2014; COSTA, 2012).

A especificidade baseada em matriz de confusão é definida pela equação

$$Especificidade = \frac{VN}{VN+FP}, \quad (3.3)$$

a especificidade indica a capacidade do classificador identificar de modo correto as vozes que apresentam alguma patologia (VIEIRA, 2014; COSTA, 2012).

Para facilitar o entendimento do classificador, é apresentado um pseudocódigo de como o algoritmo foi estruturado para apresentação dos resultados apresentados no capítulo 4.

ENTRADA: conjunto de dados;

Definir: conjunto de atributos: $X = \{x_1, x_2, \dots, x_n\}$;

conjunto de classes: $Y = \{y_1, y_2, \dots, y_n\}$;

Separar: conjunto de treinamento: $X_{tr} = \{x_{tr1}, x_{tr2}, \dots, x_{trm}\}$ e $Y = \{y_{tr1}, y_{tr2}, \dots, y_{trm}\}$;

conjunto de teste: $X_t = \{x_{t1}, x_{t2}, \dots, x_{tn}\}$;

Aplicar *cross validation*;

Definir Função Kernel: $K(x_i, x) = \exp\{-\|x - x_i\|^2/\sigma^2\}$;

Encontrar o hiperplano ótimo utilizando o conjunto de treinamento

PARA CADA padrão de observação $x_{ti} \in X_t$, **FAÇA:**

Classificar calculando: $y_{ti}(x_{ti}) = \text{sign}\left(\sum_{i=1}^n \alpha_i^0 d_i K(x_{ti}, X_t) + b_0\right)$;

FIM

SAÍDA: conjunto das classes $Y_t = \{y_{t1}, y_{t2}, \dots, y_{tn}\}$;

Capítulo 4

RESULTADOS E DISCUSSÕES

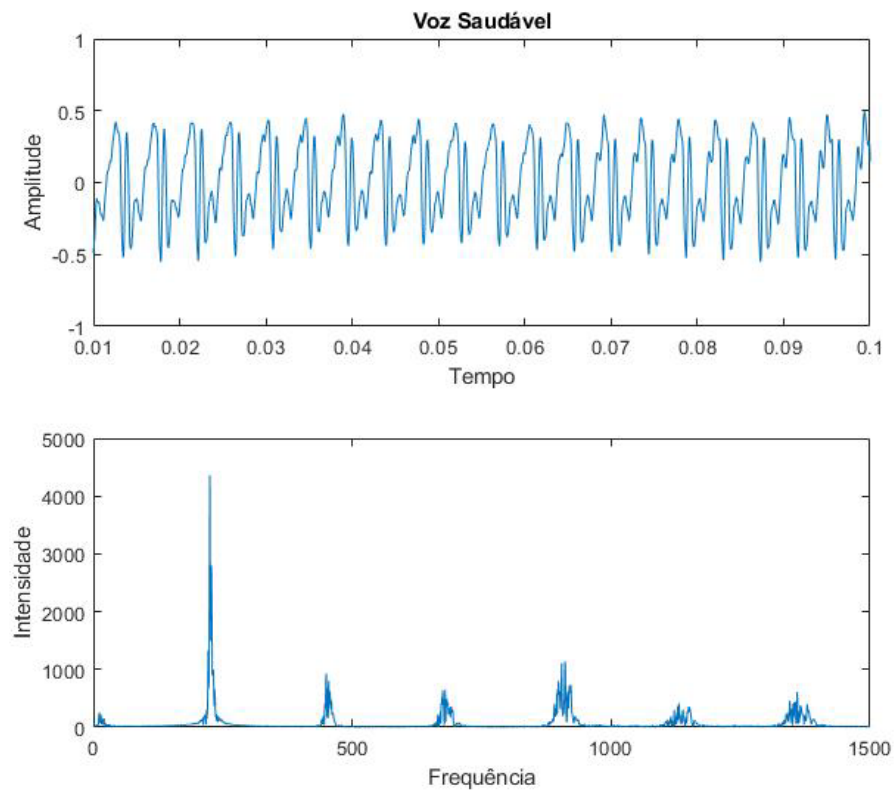
Nesta seção, são apresentados e detalhados os resultados obtidos na extração de características e classificação das patologias da laringe através da análise acústica do sinal de voz.

4.1. EXTRAÇÃO DE ATRIBUTOS

Ao analisar os sinais de voz no domínio do tempo e no domínio da frequência, pode-se observar diferenças significativas entre vozes normais e patológicas. O primeiro, representado pela Figura 4.1, é um típico sinal de voz saudável para a vogal sustentada /a/, o segundo, mostrado na Figura 4.2 é um sinal caracterizado pela presença de disfonia no paciente e o último, representado na Figura 4.3, é de um sinal de voz gerado por um paciente portador do Edema de Reinke.

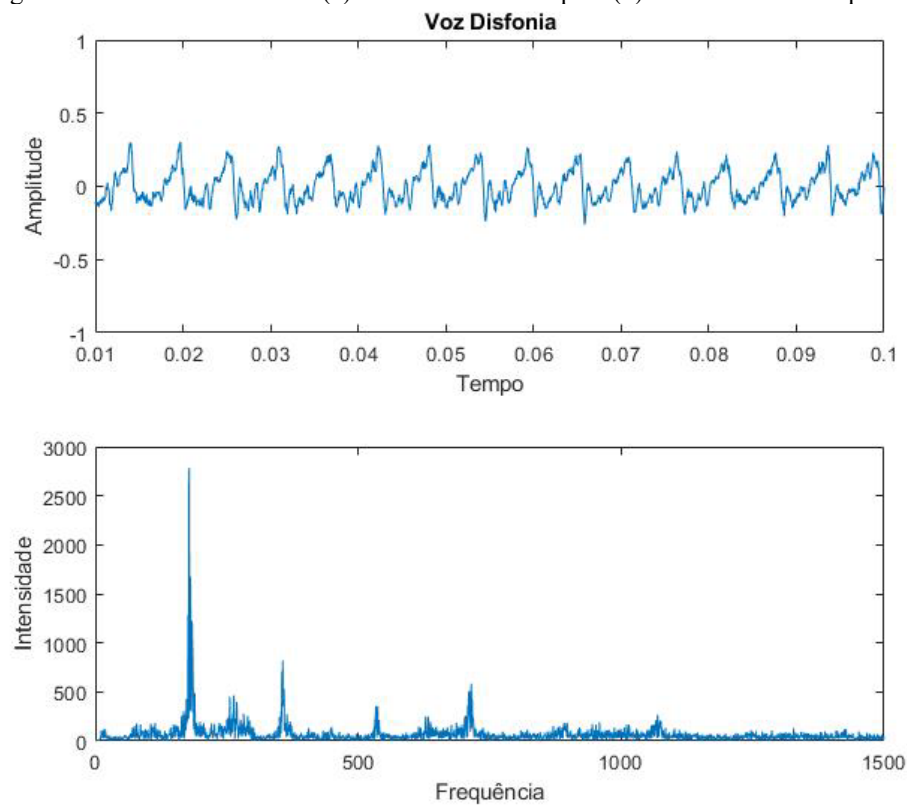
Para uma voz saudável, é importante observar a periodicidade do sinal, no domínio do tempo, como o sinal se comporta em um curto período de tempo de 100ms, como seu comportamento é constante e em toda sua extensão. Já no domínio da frequência observa-se a frequência fundamental bem definida, com sua intensidade muito acentuada e suas formantes com baixo grau de desordem, isto é, um baixo nível de ruído.

Figura 4.1 - Voz Saudável (a) no domínio do tempo e (b) no domínio da frequência



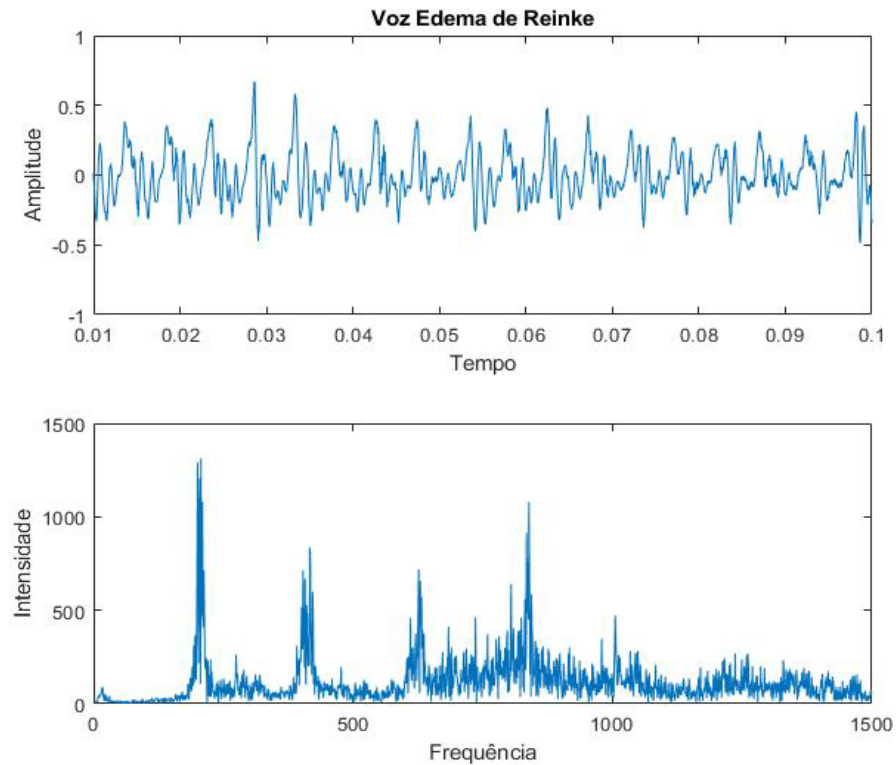
Fonte: Autora

Figura 4.2 - Voz com Disfonia (a) no domínio do tempo e (b) no domínio da frequência



Fonte: Autora

Figura 4.3 - Voz com Edema de Reinke (a) no domínio do tempo e (b) no domínio da frequência



Fonte: Autora

Nas vozes diagnosticadas com disfonia, observa-se uma diminuição da amplitude do sinal no domínio do tempo e um grau de distorção na forma de onda, sonoramente essa voz é mais baixa e “desafinada” quando comparada com a voz saudável sem presença de patologia. No domínio da frequência, pode-se observar a frequência fundamental ainda é bem definida, porém com intensidade mais baixa, já as formantes não são bem definidas e a presença de ruído entre as formantes.

Para as vozes diagnosticadas com Edema de Reinke, observa-se que o sinal no domínio do tempo não possui periodicidade e grande distorção na forma da onda, já no domínio da frequência, este sinal apresenta um alto grau de ruído característico do edema, dificultando a identificação das formantes.

Apenas observando as formas de ondas das Figuras 4.1, 4.2 e 4.3, já consegue-se identificar a presença de ruído para o caso das vozes patológicas, bem como redução do seu nível de amplitude e intensidade. De acordo com a figuras, é possível sugerir uma anormalidade ou desordem vocal, sendo necessário maior aprofundamento nos estudos das medidas acústicas que representem bem a patologia que está causando a desordem.

4.2. CLASSIFICAÇÃO

O classificador é dividido em duas máquinas SVMs, e seus resultados apresentados em 4 etapas distintas. Para o SVM1, tem-se os resultados obtidos na classificação entre as vozes saudáveis e patológicas femininas, a segunda etapa, idem com a classificação da primeira, ainda na máquina SVM1, aplicada às vozes masculinas.

As próximas etapas de classificação acontecem na segunda máquina SVM, chamada de SVM2, são para àquelas vozes que foram classificadas anteriormente como patológicas agora, na terceira e quarta etapas, haverá a separação dentre as patologias Edema de Reinke e Disfonia, na terceira etapa para vozes femininas e na quarta e última etapa para vozes masculinas.

Executou-se 50 repetições do algoritmo, com o objetivo de obter-se 50 resultados para avaliação das médias dos descritores escolhidos.

Na Tabela 4.1, é apresentada a matriz confusão da primeira repetição da máquina SVM1 para vozes femininas e calculadas sua acurácia, sensibilidade e especificidade, apresentados na Tabela 4.2.

Tabela 4.1 - Matriz de Confusão: FEMININA

	Saudável	Patológica
Saudável	70	7
Patológica	3	19

Fonte: Autora

Tabela 4.2 - Avaliação de desempenho

Acurácia	89,90
Sensibilidade	90,91
Especificidade	86,36

Fonte: Autora

Na Tabela 4.3, é apresentada a matriz confusão da primeira repetição da máquina SVM1 para vozes masculinas e calculadas sua acurácia, sensibilidade e especificidade, apresentados na Tabela 4.4.

Tabela 4.3 - Matriz de Confusão: MASCULINA

	Saudável	Patológica
Saudável	42	4
Patológica	2	11

Fonte: Autora

Tabela 4.4 - Avaliação de desempenho

Acurácia	89,83
Sensibilidade	91,30
Especificidade	84,62

Fonte: Autora

As Tabelas 4.5, 4.6 e 4.7, mostram a avaliação de desempenho da primeira máquina do algoritmo de classificação SVM1, onde são apresentados os valores mínimos e máximos, os valores médios, desvio padrão e variância alcançados pelos avaliadores.

Tabela 4.5 - Acurácia para classificação de vozes Saudáveis x Patológicas

	Feminina	Masculina
Mínimo (%)	86,49	85,43
Máximo (%)	93,78	92,79
Média (%)	90,15	89,03
Desvio Padrão	1,68	1,96
Variância	2,85	3,86

Fonte: Autora

Tabela 4.6 - Sensibilidade para classificação de vozes Saudáveis x Patológicas

	Feminina	Masculina
Mínimo (%)	88,08	87,37
Máximo (%)	92,04	91,30
Média (%)	89,82	89,64
Desvio Padrão	1,21	1,14
Variância	1,46	1,31

Fonte: Autora

Tabela 4.7 - Especificidade para classificação de vozes Saudáveis x Patológicas

	Feminina	Masculina
Mínimo (%)	84,23	84,67
Máximo (%)	90,67	89,01
Média (%)	85,84	86,30
Desvio Padrão	1,73	1,48
Variância	3,03	2,19

Fonte: Autora

Na Tabela 4.8, é apresentada a matriz confusão da primeira repetição da máquina SVM2 para vozes femininas e calculadas sua acurácia, sensibilidade e especificidade, apresentados na Tabela 4.9.

Tabela 4.8 - Matriz de Confusão: FEMININA

	Edema de Reinke	Disfonia
Edema de Reinke	11	1
Disfonia	2	8

Fonte: Autora

Tabela 4.9 - Avaliação de desempenho

Acurácia	89,36
Sensibilidade	91,67
Especificidade	80,00

Fonte: Autora

Na Tabela 4.10, é apresentada a matriz de confusão da primeira repetição da máquina SVM2 para vozes masculinas e calculadas sua acurácia, sensibilidade e especificidade, apresentados na Tabela 4.11.

Tabela 4.10 - Matriz de Confusão: MASCULINA

	Edema de Reinke	Disfonia
Edema de Reinke	3	1
Disfonia	1	8

Fonte: Autora

Tabela 4.11 - Avaliação de desempenho

Acurácia	84,62
Sensibilidade	75,00
Especificidade	88,89

Fonte: Autora

As Tabelas 4.12, 4.13 e 4.14, são apresentadas a avaliação de desempenho da primeira máquina do algoritmo de classificação SVM1, onde são apresentados os valores mínimos e máximos, os valores médios, desvio padrão e variância alcançados pelos avaliadores.

Tabela 4.12 - Acurácia para classificação de vozes com Edema de Reinke x Disfonia

	Feminina	Masculina
Mínimo (%)	77,26	69,23
Máximo (%)	90,91	92,31
Média (%)	88,64	88,23
Desvio Padrão	2,15	2,04
Variância	4,63	4,17

Fonte: Autora

Tabela 4.13 - Sensibilidade para classificação de vozes com Edema de Reinke x Disfonia

	Feminina	Masculina
Mínimo (%)	66,67	50,00
Máximo (%)	91,67	100
Média (%)	80,77	78,57
Desvio Padrão	2,30	2,13
Variância	5,29	4,53

Fonte: Autora

Tabela 4.14 - Especificidade para classificação de vozes com Edema de Reinke x Disfonia

	Feminina	Masculina
Mínimo (%)	70,00	66,67
Máximo (%)	90,00	88,89
Média (%)	82,15	82,73
Desvio Padrão	2,18	1,88
Variância	4,78	3,56

Fonte: Autora

4.2.1. Comentários e discussões

Ao analisar os resultados da máquina SVM1 entre vozes saudáveis e patológicas, observa-se na Tabela 4.5 que a melhor taxa de acerto foi na classificação de vozes femininas com 93,78% com variância de 2,85, porém não muito diferente para vozes masculinas com melhor taxa de acerto em 92,79% e variância de 3,86, demonstrando que o classificador se comporta da mesma maneira para ambos os sexos e é treinado de maneira eficiente para as diferentes faixas de frequência fundamental.

O que não acontece quando os sinais de ambos os sexos estão misturados, no espectro de frequência, vozes femininas patológicas e masculinas saudáveis se sobrepõem, ficando sobre a mesma faixa de frequência, isso faz com que o classificador tenha um baixo

desempenho por não reconhecer os padrões de maneira correta, já que a proposta deste trabalho é a análise acústica através das frequências fundamentais e suas formantes, ao misturar os dois padrões de frequência, o classificador não consegue identificar a diferença entre uma voz feminina com patologia ou masculina saudável, logo, pode-se concluir que, nessa modalidade, este algoritmo possui baixa capacidade de generalização.

Como ferramenta de validação do classificador, os valores de sensibilidade mostrados na Tabela 4.6, indicam que o classificador exclui corretamente vozes não-patológicas, ou seja, as vozes saudáveis, tendo a melhor taxa de acerto para vozes femininas com 92,04% e 91,30% para vozes masculinas. A Tabela 4.7 indica o quanto o SVM1 é capaz de classificar corretamente vozes patológicas, os valores apresentados são mais abaixo que os valores da sensibilidade, indica uma maior incerteza na exclusão dos sinais patológicos, porém ainda satisfatório com maior especificidade para vozes femininas em 90,67% e para vozes masculinas 89,01%.

Para a máquina SVM2, os valores médios da acurácia mostrados na Tabela 4.12 indicam que o ambas as máquinas se comportam da mesma maneira para a classificação dos sinais, porém a valor superior da variância para a máquina SVM2 (4,63% e 4,17% para as vozes femininas e masculinas respectivamente), indica que a máquina teve uma maior dificuldade de classificação, onde a cada uma das rodadas, apresentava valores distintos para a acurácia, isso se dá pelo fato de haver um menor número de padrões no treinamento disponibilizado pela base de dados, fazendo com que o modelo não seja tão bem treinado quanto a máquina SVM1, que há um número bem maior de padrões de treinamento.

Os valores de sensibilidade na Tabela 4.13 retratam a capacidade do SVM2 classificar as vozes com Edema de Reinke, tendo seu valor máximo em 100% para as vozes masculinas, justificado pelo menor número de padrões de teste, assim como, justificando o seu menor valor em 50%. Porém, o classificador ainda assim se comporta de maneira satisfatória com a média de acerto em 78,57%. Já a especificidade da Tabela 4.14 é a capacidade do SVM2 em excluir os sinais que não são classificados com Edema de Reinke, ou seja, aqueles que foram classificados como disfônico, apresentando seu melhor resultado em 90% para vozes femininas e 88,89% para vozes masculinas, com a média em 82% para ambos.

Capítulo 5

CONCLUSÃO E TRABALHOS FUTUROS

Buscando minimizar o desconforto de pacientes que necessitam submeter-se a exames invasivos, como laringoscopia, o presente trabalho visou apresentar uma abordagem de um método de reconhecimento de padrões e análise acústica como forma não invasiva para hipótese de diagnóstico de patologias, os resultados apresentados são bastante promissores e apontam que o uso dessas medidas e métodos para um sistema de classificação de vozes, podem ser uma ferramenta auxiliar de um médico especialista de pré-diagnósticos de patologias laringeas, porém o diagnóstico automático dessas patologias empregando métodos baseados unicamente na voz do paciente ainda demanda estudos mais extensos pois as similaridades entre algumas disfonias, juntamente com a complexidade do sistema produtor da fala ainda não produzem métodos 100% confiáveis.

A presença de patologias na laringe como o Edema de Reinke, provoca um mau fechamento da glote, devido ao aumento de massa nas pregas vocais, causando uma vibração irregular. Com essa irregularidade na vibração das pregas vocais é possível utilizar algumas ferramentas matemáticas para retirar características acústicas para servir como parâmetro de reconhecimento dos padrões dessas patologias.

Utilizou-se a base de dados alemã SVD, que está disponível *online* de forma gratuita pelo Instituto de Fonética da Universidade de *Saarland*, nesta base de dados em que é possível encontrar sinais de voz, entre saudáveis e patológicos de mais de 2000 sujeitos.

As medidas acústicas utilizadas foram a frequência fundamental, as formantes, a energia e entropia como vetor de característica de entrada para as máquinas de vetor suporte.

Como contribuição, este trabalho propôs diferentes descritores dos que foram encontrados no estado da arte, além da utilização das frequências e seus harmônicos, a escolha do descritor da energia do sinal, diferencia-se dos demais trabalhos, trazendo uma nova maneira de análise destes sinais, atrelado com o descritor entropia que vem auxiliar a caracterização do sinal com o cálculo do grau de desordem do mesmo.

Apresentou-se duas máquinas SVM de classificação, é importante apontar que estas máquinas não foram implementadas para serem capazes de separar os sinais vozes femininas e masculinas, então toda classificação e separação de patologias é feita para cada sexo individualmente. Diante disso, foi feita uma pré-classificação para que o vetor de entrada fosse apenas de vozes de sujeitos do mesmo sexo.

A primeira é a máquina, chamada SVM1, teve como objetivo separar as vozes saudáveis das patológicas, com isso ela foi usada de maneira binária na classificação entre as vozes saudáveis e vozes que apresentam algumas das patologias. O vetor de característica das medidas acústicas foi utilizado como parâmetro de entrada e após ser processado e analisado pelo SVM1, como parâmetro de saída, resultou em outros dois vetores de características, àquele com as vozes saudáveis e àquele com as vozes que apresentam alguma patologia.

Para a segunda máquina, chamada de SVM2, o parâmetro de entrada foi o vetor de sinais das vozes classificadas como patológicas na máquina SVM1, e então esta máquina classificou as patologias Edemas de Reinke e Disfonia, todos os resultados e discussões da metodologia empregada nas máquinas estão detalhadas no capítulo 4 deste trabalho.

De maneira geral, mostrou-se que o SVM1 possui um melhor desempenho, apesar de seus valores médios de acurácias serem muito próximos com os do SVM2, os valores médios de sensibilidade e especificidade são menores na classificação entre as patologias no SVM2, isto é, a segunda máquina apresentou uma maior incerteza na classificação dos dados, mas ainda suficientemente satisfatória para uma hipótese de diagnóstico.

Conclui-se que o método proposto permite uma maior precisão no diagnóstico, ou seja, o classificador permitirá ao médico especialista uma análise adicional proporcionando uma maior probabilidade de detectar patologias onde realmente existam. O médico poderá dispor de uma ferramenta complementar aos exames existentes auxiliando num diagnóstico definitivo.

5.2. SUGESTÕES PARA TRABALHOS FUTUROS

Diversas atividades podem dar continuidade a este trabalho como sugestão para trabalhos futuros, recomenda-se:

- análise com outros bancos de dados;
- análise de outras patologias utilizando a extração das mesmas características acústicas apresentadas neste trabalho;
- utilização de novas medidas acústicas, implementando novos atributos para melhorar o desempenho do sistema de classificação;
- utilização de outras estruturas de SVM, mais robustas;
- desenvolvimento de um sistema automático no processo de separação das vozes femininas e masculinas;
- inclui a opção de rejeição para melhorar o desempenho do algoritmo;
- desenvolvimento de uma interface gráfica; e
- comparar resultados obtidos com outras ferramentas de *machine learning*.

REFERÊNCIAS BIBLIOGRÁFICAS

AL-NASHERI, Ahmed et al. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. **Ieee Access**, v. 6, p. 6961-6974, 2017.

ALMEIDA, Náthalee Cavalvante de. **Sistema inteligente para diagnóstico de patologias na laringe utilizando máquinas de vetor de suporte**. Orientador: Prof. Dr. Sandro Gonçalves da Silva. 2010. 118 p. Dissertação (Mestrado em Engenharia Elétrica) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2010.

BARBOSA, Hildegard Paulino; FECHINE, Joseana Macêdo; RANGEL, José Eustáquio. Classificação de patologias da fala a partir do PPM. **Congresso Brasileiro de Inteligência Computacional**, Campina Grande, p. 1-6, março 2016.

BEHLAU, Mara; MADAZIO, Glaucya; FEIJO, Debora; PONTES, Paulo. **A Voz do Especialista V.1**. Rio de Janeiro: Revinter, 2001. Capítulo 3. p. 130-164.

BURGES, Christopher JC. A tutorial on support vector machines for pattern recognition. **Data mining and knowledge discovery**, v. 2, n. 2, p. 121-167, 1998.

CENTRAL DA SAÚDE. Central da Fonoaudiologia. **Nódulos de Pregas Vocais**. Brasil, 2018. Disponível em: <http://www.centraldafonoaudiologia.com.br/tratamentos/fonoaudiologia-nodulos-de-pregas-vocais>. Acesso em: 17 set. 2019.

CIELO, Carla Aparecida et al. Lesões organofuncionais do tipo nódulos, pólipos e edema de Reinke. **Revista CEFAC**, v. 13, p. 735-748, 2011.

CORDEIRO, Gislaine Ferro; PINHO, Sílvia M.R.; CAMARGO, Zuleica Antonia. Formante do cantor—um enfoque fisiológico. **PINHO, Silvia M R. Temas em voz profissional**. São Paulo: Revinter, p. 23-30, 2007.

CORDEIRO, Hugo Tito. **Reconhecimento de patologias da voz usando técnicas de processamento da fala**. Orientador: Prof. Dr. Carlos Meneses. 2016. 120 p. Dissertação (Doutorado em Engenharia Eletrônica e de Computadores) - Faculdade de Ciências e Tecnologia, Universidade de Lisboa, Lisboa, 2016.

COSTA, Silvana Luciana do Nascimento Cunha. **Análise acústica, baseada no modelo linear de produção de fala, para discriminação de vozes patológicas**. Orientador: Prof. Dr. Benedito Guimarães Aguiar Neto. 2008. 160 p. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Campina Grande, Campina Grande, 2008.

COSTA, Washington César de Almeida et al. **Análise dinâmica não linear de sinais de voz para detecção de patologias laríngeas**. 2012. 178 f. Tese (Doutorado) - Ciências no Domínio da Engenharia Elétrica, Programa de Pós-Graduação em Engenharia Elétrica (COPELE), Universidade Federal de Campina Grande.

CRAMMER, Koby; SINGER, Yoram. On the algorithmic implementation of multiclass kernel-based vector machines. **Journal of machine learning research**, v. 2, n. Dec, p. 265-292, 2001.

DIAS, Lucas Cardoso. **Detecção de patologias laríngeas por meio da análise de sinais de voz utilizando Deep Neural Networks**. 2020. 82 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia Elétrica, Programa de Pós-Graduação em Engenharia Elétrica (Ppgee), Instituto Federal de Educação da Paraíba, João Pessoa, 2020.

DIETTERICH, T. G. Ensemble learning. In: ARBIB, M. A. (Ed.). **The Handbook of Brain Theory and Neural Networks**. 2nd. ed., Cambridge/MA: Editora MIT Press, 2002.

DUAN, Kai-Bo; KEERTHI, S. Sathya. Which is the best multiclass SVM method? An empirical study. In: **International workshop on multiple classifier systems**. Springer, Berlin, Heidelberg, 2005. p. 278-285.

DUDA R.O., HART, P.E. **Pattern classification and scene analysis**. John Wiley Sons, New York: Editora The MIT Press, 1973.

FARIA, Joana Costa Frutuoso. **Avaliação complementar da voz através de medidas acústicas de longo termo em vozes disfônicas**. 2021. Tese de Doutorado.

FRAZÃO, Maria do Carmo de Luna Malheiros; COSTA, Silvana Luciene do Nascimento Cunha; CORREIA, Suzete Élide Nóbrega. Análise acústica de sinais de vozes patológicas por meio de características temporais. **Revista Principia**, João Pessoa, v. 1, n. 22, p. 28-34, 2013.

GUERINO, Marcelo. **Sistema Respiratório**. 2015. Disponível em: <https://slideplayer.com.br/slide/4118084/>. Acesso em: 06 nov. 2019.

HAYKIN, S. **Redes Neurais – Princípios de Práticas** (2ª ed). Porto Alegre/RS: Editora Bookman, 2001.

KOHAVI, Ron et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Ijcai**. 1995. p. 1137-1145.

MARTINS, Priscila Campos; COUTO, Thalita Evaristo; GAMA, Ana Cristina Côrtes. Auditory-perceptual evaluation of the degree of vocal deviation: correlation between the Visual Analogue Scale and Numerical Scale. In: **Codas**. Sociedade Brasileira de Fonoaudiologia, 2015. p. 279-284.

MARYN, Youri; ROY, Nelson. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. **Jornal da Sociedade Brasileira de Fonoaudiologia**, v. 24, p. 107-112, 2012.

MINISTÉRIO DA SAÚDE. Instituto Nacional do Câncer. **Câncer de laringe**. Brasil, 24 jul. 2019. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-laringe>. Acesso em: 10 set. 2019.

MUNOZ, J. et al. Acoustic and perceptual indicators of normal and pathological voice. **Folia Phoniatria et logopaedica**, v. 55, n. 2, p. 102-114, 2003.

NICOLOSI, R. **Edema de Reinke**, 2003. Disponível em: <http://www.reginanicolosi.com.br/patologialaringe.asp>. Acesso em: 15 de agosto de 2019

PARRAGA, Adriano. **Aplicação da transformada Walet Packet na análise e classificação de sinais de vozes patológicas**. Orientador: Prof. Dr. Adalberto Schuck Jr. 2002. 148 p. Dissertação (Mestrado em Engenharia Elétrica) - Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

PLATT, J. C.; CRISTIANINI, N.; SHAW-TAYLOR, J. **Large margin dags for multiclass classification**. In: **Advances in Neural Information Processing Systems**. Editora MIT Press, 2000. p. 547–553.

ROCHA-NETO, A. R. **SINPATCO - Sistema Inteligente para o Diagnóstico de Patologias da Coluna Vertebral**. Orientador: Prof. Dr. Paulo Cesar Cortês. Dissertação (Mestrado). Departamento de Engenharia de Teleinformática. Programa De Pós-Graduação em Engenharia de Teleinformática - Universidade Federal do Ceará, 2006.

ROCHA-NETO, A. R. **SINPATCO II: Novas estratégias de aprendizado de máquina para classificação de patologias da coluna vertebral**. Orientador: Prof. Dr. Guilherme de Alencar Barreto Tese (Doutorado). Departamento de Engenharia de Teleinformática. Programa De Pós-Graduação em Engenharia de Teleinformática - Universidade Federal do Ceará, 2011.

ROSA, Marcelo de Oliveira. **Análise acústica da voz para pré-diagnóstico de patologias da laringe**. Orientador: Prof. José Carlos Pereira. 1998. 218 p. Dissertação (Mestrado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 1998.

ROSA, Marcelo de Oliveira. **Laringe digital**. Orientador: Prof. Tit. José Carlos Pereira. 2002. 262 p. Tese (Doutorado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2002.

ROSA, Marcelo de Oliveira. Modelagem da laringe: da biologia ao computador. **Revista de Letras**, Brasil, v. 30, n. 4, p. 70-81, dez. 2011.

ROSA, Marcelo de Oliveira; PEREIRA, José Carlos; GRELLET, Marcos. Adaptive estimation of residue signal for voice pathology diagnosis. **IEEE Transactions on Biomedical Engineering**, v. 47, n. 1, p. 96-104, 2000.

ROSA, M. O. et al. Signal processing and statistical procedures to identify laryngeal pathologies. In: **ICECS'99. Proceedings of ICECS'99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No. 99EX357)**. IEEE, 1999. p. 423-426.

SALVIANO FILHO, Paulo de Tarso. **Ferramenta didática para o ensino de vibrações e de dinâmica em engenharia**. 2009. 111 f. Dissertação (Mestrado) - Curso de Departamento de Engenharia Mecânica, Universidade de Brasília, Brasília (Df), 2009.

SANTOS, Vanessa Sardinha dos. **"O que é inspiração e expiração?"**; *Brasil Escola*. Disponível em: <https://brasilecola.uol.com.br/o-que-e/biologia/o-que-e-inspiracao-expiracao.htm>. Acesso em 19 de agosto de 2019.

SHANNON, Claude Elwood. A mathematical theory of communication. **The Bell system technical journal**, v. 27, n. 3, p. 379-423, 1948.

SODRÉ, Bruno Ribeiro. **Reconhecimento de padrões aplicados à identificação de patologias de laringe**. Orientador: Marco José da Silva. 2016. 108 p. Dissertação (Mestrado em Engenharia de Automação e Sistemas) - Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná, Curitiba, 2016.

SODRÉ, Bruno; ROSA, Marcelo; DASSIE-LEITE, Ana. Evaluating the use of neural networks and acoustic measurements to identify laryngeal pathologies. In: **2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. IEEE, 2017. p. 4590-4593.

SMOLA, Alex J.; SCHÖLKOPF, Bernhard. A tutorial on support vector regression. **Statistics and computing**, v. 14, n. 3, p. 199-222, 2004.

STEFFEN, Luciane M. et al. Paralisia unilateral de prega vocal: associação e correlação entre tempos máximos de fonação, posição e ângulo de afastamento. **Revista Brasileira de Otorrinolaringologia**, v. 70, p. 450-455, 2004.

VAPNIK, Vladimir. **Estimation of dependences based on empirical data**. Springer Science & Business Media, 1982.

WATANABE, Satoshi. **Pattern recognition: human and mechanical**. John Wiley & Sons, Inc., 1985.

WOKUREK, Wolfgang. PUTZER, Manfred. Phonation Quality Detection on the Saarbrücken Voice Database using Harmonic Spectrum-based Parameters. In **Proceedings of the 10th International Workshop MAVEBA**, 2017, pages 15–18, Firenze, Italy

YAMASAKI, Rosiane; TSUJI, Domingos. **Produção da voz**. 2017. Disponível em: http://disfoniaespasmodica.com.br/producao_vocal.html. Acesso em: 01 nov 2019.

ZITTA, Simone Maganhotto. **Análise perceptivo-auditiva e acústica em mulher com nódulos vocais**. Orientador: Prof. Dr. Álvaro Luiz Stelle. 2005. 112 p. Dissertação (Mestrado em Engenharia Biomédica) - Centro Federal de Educação do Paraná- CEFET-PR, Curitiba, 2005.