

# **Desenvolvimento de Protótipo de Sistema de Suporte ao Diagnóstico de Patologias da Voz**

**Patrícia Manuela Andrade Rodrigues**

Dissertação apresentada à

**Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Bragança**

para a obtenção do grau de Mestre em

**Tecnologia Biomédica**

**Instrumentação e sinais Médicos**

**Orientadores: Professor Doutor João Paulo Teixeira e Professor Doutor Pedro João Rodrigues**

**2022**

## **AGRADECIMENTOS**

A realização desta dissertação de mestrado contou com importantes apoios e incentivos que, de alguma forma, permitiram que esta tese se concretizasse e aos quais estarei eternamente grata.

Em primeiro lugar quero agradecer aos meus orientadores, Prof. Doutor João Paulo Teixeira, e Prof. Doutor Pedro João Rodrigues, do Instituto politécnico de Bragança, por todo o apoio prestado, pelo trato simples, e correto, pela disponibilidade e paciência, e por ter acreditado em mim e nas minhas capacidades sem nunca ter permitido que o desalento se instalasse, mesmo quando as coisas não corriam bem.

À colega Joana Filipa Teixeira Fernandes, pelo apoio, disponibilidade e colaboração no solucionar de dúvidas e problemas que foram surgindo ao longo da realização deste trabalho.

Agradeço, de coração aos meus pais, irmãos, tios e primas, e a todos os meus amigos pela eterna compreensão da minha ausência nos vários momentos em que a realização desta tese assim o exigiu, pelo animo, apoio e incentivo dado ao longo do meu percurso académico, e que de tal maneira resultou em um incentivo maior para a realização do mesmo, e ser quem eu sou hoje.

A todos meu muito obrigada!

## RESUMO

A voz é uma ferramenta de comunicação primordial nas relações inter-humanas, por meio de inflexões, pausas, variações de ritmo e de intensidade. É considerada a integridade da nossa identidade, pois através dela somos reconhecidos e a sua qualidade permite-nos expressar eficazmente.

As patologias vocais encontram-se presentes na nossa sociedade, com profundo impacto na qualidade de vida das pessoas. A origem deve-se a várias causas e apresentam diferentes graus de gravidade. A patologia pode progredir de forma benigna ou maligna, por isso é de extrema importância ter atenção aos sinais de alteração. Um diagnóstico precoce é muito relevante para o tratamento. Porém, as formas de avaliação existentes nesta área são invasivas e desagradáveis, sendo incomodativas para o paciente. Estes aspetos motivaram o desenvolvimento de métodos não invasivos, que possam fazer uma avaliação exata e possam ser utilizados como um método de ajuda ao diagnóstico eficaz.

Neste trabalho desenvolveu-se um sistema de suporte à decisão médica no diagnóstico de patologias vocais. Para o desenvolvimento deste sistema foi necessário o estudo de um conjunto de parâmetros acústicos, bem como de classificadores, como rede neuronal artificial (RNA), com o objetivo de fazer a classificação final do paciente entre saudável e patológico. Os parâmetros utilizados neste trabalho são: *Jitter* Absoluto (*Jitta*), *Jitter* Relativo (*Jitter*), *Shimmer* Absoluto (*ShdB*), *Shimmer* Relativo (*Shim*), *Harmonic to Noise Ratio* (HNR), e a Autocorrelação. E como classificador o modelo da rede neuronal *Multi Layer Perceptron* (MLP).

O sistema interface gráfica desenvolvido neste trabalho servirá como um método complementar no pré-diagnóstico de patologias da voz. O modelo MLP utilizada obteve uma taxa de exatidão de 98.86% que se encontra entre os melhores valores tendo em conta estado a arte, no entanto a possibilidade da inserção deste sistema em clínicas e hospitais contribuirá para o seu aperfeiçoamento por meio de familiarização com profissionais de saúde.

**Palavras-chave:** Patologias da Voz; Sinal da Fala; Sistema; Parâmetros Acústicos; Redes Neurais.

## ABSTRACT

Voice is a powerful tool that make possible the communication between humans being, by way of inflections, pauses, variations in rhythm and intensity. As the integrity of our identity, it is considered to allow us to express ourselves effectively to be recognized.

Vocal pathology become a reality in modern society, having a deep impact by affecting people life quality. it is extremely important to be aware of signs of alterations that can be derived from different causes, with different degrees of severity, which evolve in a benign or malignant. Early diagnostic intervention revealed to be extremely crucial to treatment. However, the already existed examination procedures in this field show to be invasive and unpleasant, which can be uncomfortable to the patient. Therefore, the need to develop a non-invasive method emerged in order to delivery an accurate and effective diagnosis.

In this work, a medical decision support system in the diagnosis of vocal pathologies was developed. For the development of this system, it was necessary to study a set of acoustic parameters, as well as classifiers, of artificial neural networks (ANN), with the objective of making the final classification of the patient between healthy and pathological. The parameters used in this work are Absolute *Jitter (Jitta)*, Relative *Jitter (Jitt)*, Absolute *Shimmer (ShdB)*, Relative *Shimmer (Shim)*, *Harmonic to Noise Ratio (HNR)*, and Autocorrelation. And as a classifier, the *Multi-Layer Perception (MLP)* neural network model.

The graphic interface system developed in this work will serve as a complementary method in the pre-diagnosis of voice pathologies. The MLP model used obtained a success rate of 98.86%, which is among the best values taking into account the state of the art, however the possibility of inserting this system in clinics and hospitals will contribute to its improvement through familiarization with health professionals.

**Keywords:** Voice Pathologies; Speech Sign; System; Acoustic Parameters; Neural Networks

**“Porque cada um, independente das habilitações que tenha, ao menos uma vez na vida fez ou disse coisas muito acima da sua natureza e condição, e se a essas pessoas pudéssemos retirar do cotidiano pardo em que vão perdendo os contornos, ou elas a si próprias se retirassem de malhas e prisões, quantas mais maravilhas seriam capazes de obrar, que pedaços de conhecimento profundo poderiam comunicar, porque cada um de nós sabe infinitamente mais do que julga e cada um dos outros infinitamente mais do que neles aceitamos reconhecer.” José Saramago.**

# ÍNDICE

AGRADECIMENTOS .....	i
RESUMO .....	ii
ABSTRACT .....	iii
ÍNDICE.....	v
NOMENCLATURAS, ABREVIATURAS E SÍMBOLOS.....	vii
ÍNDICE DE FIGURAS .....	ix
ÍNDICE DE TABELAS .....	x
1. INTRODUÇÃO.....	1
1.1 Aparelho fonador.....	1
1.2 Laringe.....	2
1.3 Análise Acústica.....	6
1.4 Objetivos do trabalho .....	6
1.5 Estrutura de tese.....	7
2. ESTADO DE ARTE.....	10
3. PATOLOGIA DA VOZ .....	19
3.1 Paralisia de Cordas Vocais .....	19
3.2 Laringite Crónica.....	20
3.3 Cancro da Laringe .....	20
3.4 Pólipo das Cordas vocais.....	21
3.5 Quistos.....	21
3.6 Edema das Cordas Vocais (Reinke) .....	22
3.7 Granuloma .....	22
3.8 Granulomas de Intubação .....	23
3.9 Fibroma Laríngeo .....	23
3.10 Tumor da Hipofaringe .....	23
3.11 Disfonia .....	24
3.11.1 Disfonia Funcional .....	24
3.11.2 Disfonia Hiperfuncional.....	25
3.11.3 Disfonia Hipofuncional.....	25
3.11.4 Disfonia Hipotónica .....	25

3.11.5	Disfonia Espasmódica.....	26
3.11.6	Disfonia Psicogénica.....	26
3.12	Efeitos das patologias da voz no sinal de fala.....	26
4.	METODOLOGIA.....	29
4.1	Interface gráfica do utilizador.....	29
4.1.1	Benefícios do uso de sistemas na área da saúde.....	29
4.1.2	Métricas para a classificação de uma interface gráfica.....	30
4.1.3	Linguagem de programação.....	31
4.1.4	Leitura do código <i>Matlab</i> para <i>Python</i> .....	31
4.2	Parâmetros extraídos do sinal acústico.....	32
4.2.1	Jitter - perturbação da frequência a curto prazo.....	33
4.2.1.1	<i>Jitter</i> Absoluto ( <i>Jitta</i> ).....	34
4.2.1.2	<i>Jitter</i> Relativo ( <i>Jitter</i> ).....	34
4.2.1.3	<i>Jitter</i> ( <i>RAP</i> ).....	34
4.2.1.4	<i>Jitter</i> ( <i>PPQ5</i> ).....	35
4.2.2	Shimmer - perturbação da frequência a curto prazo.....	35
4.2.2.1	<i>Shimmer</i> Relativo ( <i>Shim</i> ).....	35
4.2.2.2	<i>Shimmer</i> Absoluto ( <i>ShdB</i> ).....	35
4.2.2.3	<i>Shimmer</i> ( <i>APQ3</i> ).....	36
4.2.2.4	<i>Shimmer</i> ( <i>APQ5</i> ).....	36
4.2.3	Parâmetros Harmónicos.....	37
4.2.3.1	<i>Harmonic to Noise Ratio</i> ( <i>HNR</i> ).....	37
4.2.3.2	Autocorrelação.....	37
4.2.3.3	<i>Noise to Harmonic Ratio</i> ( <i>NHR</i> ).....	38
4.3	Modelo de classificação MLP.....	38
5.	RESULTADOS.....	41
5.1	Janela <i>RecordSound</i> .....	42
5.2	Janela <i>Analysis</i> .....	44
5.3	Janela <i>Diagnosis</i> .....	46
5.4	Modelo de classificação.....	47
6.	DISCUSSÃO DOS RESULTADOS.....	50
7.	CONCLUSÃO E TRABALHOS FUTUROS.....	52
8.	REFERENCIAS BIBLIOGRAFICAS.....	55

## NOMENCLATURAS, ABREVIATURAS E SÍMBOLOS

F0	Frequência Fundamental
FFT	Fast Fourier transform
WPT	Transformada <i>Wavelet Packet</i>
AVA	Analise Acústica Vocal
Jitta	Jitter Relativo
Jitter	Jitter Absoluto
RAP	<i>Relative Average Perturbation Jitter</i>
PPQ5	<i>Five-point Period Perturbation Quotients Jitter</i>
Shim	<i>Shimmer</i> Absoluto
ShdB	<i>Shimmer</i> Relativo
APQ3	<i>Three-point Amplitude perturbation Quotients Shimmer</i>
APPQ5	<i>Five-point Amplitude perturbation Quotients Shimmer</i>
HNR	<i>Harmonic to Noise Ratio</i>
NHR	<i>Noise to Harmonic Ratio</i>
LSF	Frequência Espetral Linear
LPC	Coefficientes de predição Linear
MFCC	Coefficientes Ceptrais de Frequência de Mel
RNA	Redes Neurais Artificiais
MLP	<i>Multi Layer Perceptron</i>
IA	Inteligência Artificial
GAN	Redes Neurais Adversárias
SVM	Máquina Vetor Suporte
HMM	Modelo Oculto de <i>Marcov</i>
VAE	Autoencoder Varacional

CNN	Rede Neuronal Convolutacional
SLL	Aprendizado Semissupervisionado
GMM	Modelo de Mistura Gaussiana
PCA	Análise de Componente Principais
kPCA	Análise de Componentes Principais do <i>Kernel</i>
NLPCA	Análise de Componentes Principais não Lineares
SVD	<i>Saarbrucken Voice Database</i>
MEEI	<i>Massachusetts Eye and Ear Infirmary</i>
AVPD	Base de Dados da Patologia da Voz Árabe
SPDD	Base de Dados de Parkinson em espanhol
MAE	<i>Mean Absolute Error</i>
MSE	<i>Mean Squared Error</i>
SGD	<i>Stochastic Gradient Descent Optimizer</i>
GUI	<i>Grafical User Interface</i>

## ÍNDICE DE FIGURAS

Figura 1- Estrutura do aparelho fonador (Cordeiro, 2016).....	2
Figura 2- Vista anterior da laringe. (Stojanovic & Belic, 2013). .....	3
Figura 3-Anatomia das cordas vocais (Michael P Underbrink, 2019). .....	4
Figura 4- Processo de interação utilizador-interface (Prates & Barbosa, 2003). .....	29
Figura 5-Representação dos parâmetros <i>jitter</i> e <i>shimmer</i> para um sinal de fala 0(Bäckström, 2022). .....	33
Figura 6- Representação do <i>Front-end</i> do sistema. ....	41
Figura 7- <i>Front-end</i> do sistema quando pressionado o botão <i>Help</i> .....	42
Figura 8- Janela <i>RecordSound</i> . ....	43
Figura 9- Janela <i>RecordSound</i> , ao efetuar uma gravação.....	43
Figura 10- Janela <i>RecordSound</i> ao selecionar o botão <i>Diagnosis</i> .....	44
Figura 11- Janela <i>Analysis</i> . ....	45
Figura 12- Janela <i>Analysis</i> ao clicar no botão com seta <i>Play</i> . ....	45
Figura 13- Janela <i>Históric</i> . ....	46
Figura 14- Janela <i>Diagnosis</i> . ....	46
Figura 15-Exatidão do modelo MLP. ....	48

## ÍNDICE DE TABELAS

Tabela 1: Valores dos parâmetros obtidos através <i>software Praat</i> e através do algoritmo correspondentes à vogal “a”. .....	47
--	----



# 1. INTRODUÇÃO

## 1.1 Aparelho fonador

A variação melódica da fala caracteriza-se pelo facto de ocorrerem vibrações das cordas vocais, que resultam numa forma de onda acústica periódica produzida pelo fluxo de ar da fonação. Essa forma de onda vibra com uma frequência fundamental ( $F_0$ ) e uma série de harmónicos que definem o timbre do som. Essas características podem ser encontradas a partir da análise espectral dos sons. Embora haja uma relação entre a produção de determinados segmentos e as características entoacionais, tal facto não tem sido considerado muito relevante para as análises linguísticas (Fry, 1976).

Na produção da voz intervêm vários sistemas e órgãos, representados na Figura 1. O sistema respiratório é a fonte de energia, responsável pelo ar expelido pelos pulmões através das cordas vocais. O sistema fonatório é a fonte de vibrações, representado pelas cordas vocais. O sistema de ressonância inclui a cavidade oral e nasal. Do sistema articulatorio faz parte a língua, lábios, céu da boca, palato e dentes. E, por fim, o sistema nervoso central e periférico, como o córtex permite a coordenação (Almeida, 2010) .

Pode dizer-se que a voz é produzida pela pressão de ar vinda dos pulmões, que faz vibrar as cordas vocais aduzidas, provocando um som que é articulado e modificado na boca e amplificado nas cavidades de ressonância. Para que a voz se ouça, o som deve ser impelido por uma onda sonora regularmente mantida e ao nível da laringe deve-se formar uma vibração periódica. Essa vibração surge pela possibilidade que o homem tem de unir as cordas vocais durante a expiração, produzindo uma série de aberturas/fechos que geram variações de pressão no interior do fluxo do ar. Esta vibração das cordas vocais depende de um fator fisiológico que lhe é associado – a mobilidade – e de uma componente psicológica – o sistema nervoso autónomo (Vale, 2011).

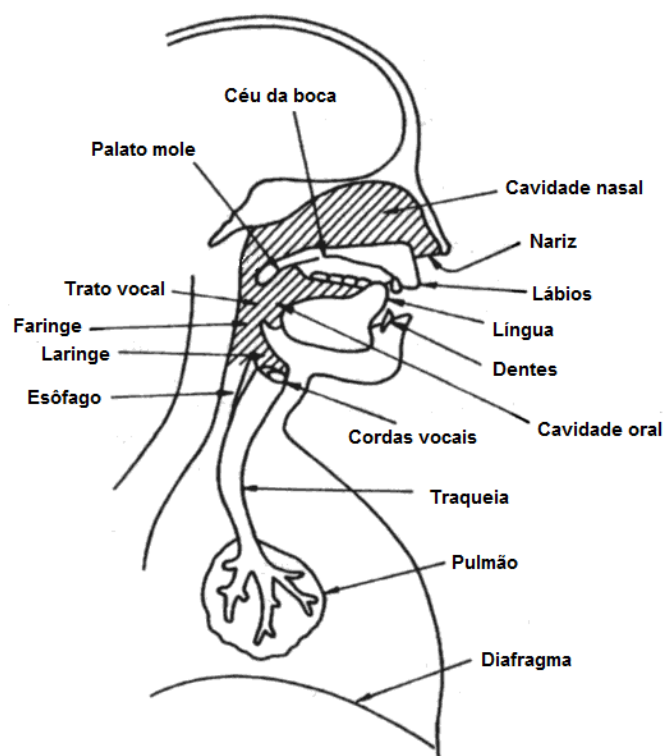


Figura 1- Estrutura do aparelho fonador (Cordeiro, 2016).

## 1.2 Laringe

A laringe representada na Figura 2, é o órgão onde se localizam as cordas vocais, responsáveis pela produção da voz. Localizada na linha mediana do pescoço anterior ela situa-se na encruzilhada aerodigestiva e é o início do sistema respiratório inferior. É um tubo fibroelástico entre o osso hióide e a traqueia, cuja camada externa é composta por cartilagem e músculos e a camada interna é mucosa. A laringe está ligada por ligamentos e músculos ao osso hióide e, portanto, acompanha seus movimentos. Ela estende-se da terceira à quarta vértebras cervicais. A partir de sua abertura na parede anterior da faringe inferior, a laringe desce pelo colo anterior e continua seu caminho pela traqueia. A borda superior da laringe é apresentada por uma borda livre de epiglote e plicas ariepiglóticas. A borda inferior da cartilagem cricóide forma a borda inferior da laringe (Stojanovic & Belic, 2013).

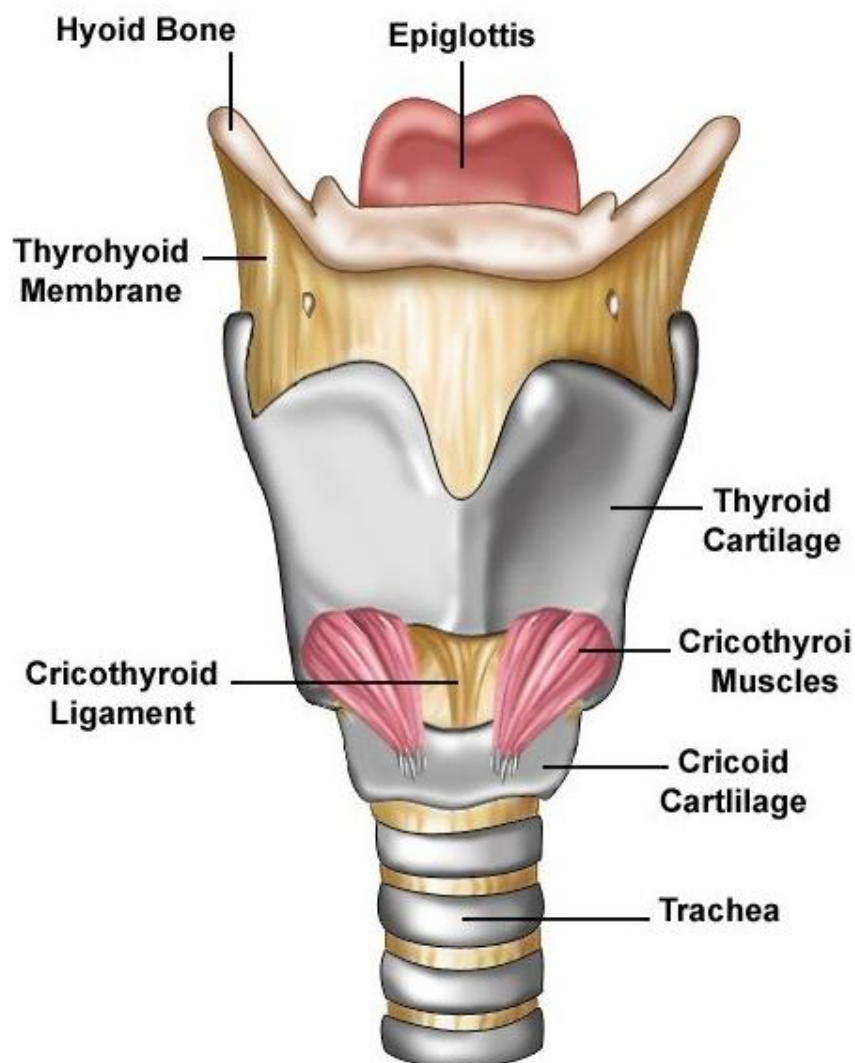


Figura 2- Vista anterior da laringe. (Stojanovic & Belic, 2013).

É um órgão móvel, pois movimenta-se para cima durante a inspiração, deglutição, pronúncia da vogal “e” e para baixo na expiração, após a deglutição e na pronúncia da vogal “u”. O alargamento ou estreitamento do espaço supraglótico decorrente dos movimentos de descida e de subida da laringe e tem implicação direta na ressonância dos sons, desta forma quanto mais largo o canal, maior a ressonância. No interior da laringe encontram-se as cordas vocais (Figura 2), duas estruturas altamente flexíveis, semelhantes a dois grandes lábios, constituídas por um músculo e um ligamento cada, que se encontra na laringe bem acima da traqueia. Em estado de repouso as cordas vocais ficam separadas e aparece um espaço entre elas. Esse espaço recebe o nome de glote, e quando falamos, no entanto, elas se tornam bastante próximas e vibram, para produzir o som (Braga & Oliveira, 2014).

Através da passagem do ar faz-se vibrar as duas cordas de fibras elásticas e pode-se modificar a sua forma e elasticidade através dos músculos da laringe. Assim sendo, é possível produzir diferentes sons, bem como, alterar o tom de voz ou cantar. O cérebro controla estes músculos enviando mensagens através dos nervos para controlar a aproximação e a tensão das cordas vocais, de modo, que estas 2 vibrem quando o diafragma e os músculos empurram o ar para fora dos pulmões (Cordeiro, 2016).

O processo de abertura e fecho das cordas vocais durante o fluxo de ar vindo dos pulmões funciona como gerador de som (Teixeira, 1995).

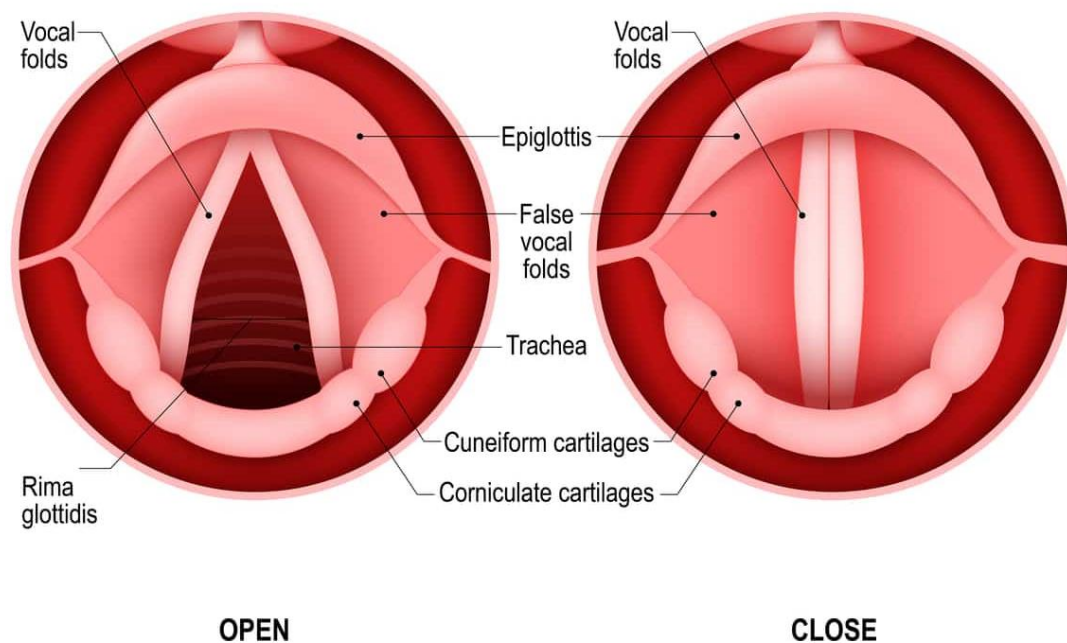


Figura 3-Anatomia das cordas vocais (Michael P Underbrink, 2019).

Qualquer perturbação na voz trará implicações profundas na vida social e profissional de uma pessoa (Almeida, 2010) (Lopes, 2008). A dificuldade de detetar a patologia da voz está intimamente associada à gravidade da patologia. Um paciente com patologia menos grave pode pronunciar uma vogal sem muita dificuldade, no entanto, um paciente com patologia vocal grave, impede o fecho e a abertura adequada das cordas vocais ao pronunciar um som de vogal (Muhammad et al., 2016)

### **1.3 Análise Acústica**

Como a função vocal apresenta uma multiplicidade de fatores envolvidos na sua produção, nenhuma medida única da voz pode, por si só, estimar um conjunto total de informações sobre a fonação. A análise acústica, usando parâmetros múltiplos, tem-se revelado o método mais apropriado de se descrever vozes patológicas. Por outro lado, como há muitos parâmetros acústicos, é importante selecioná-los com acurada reflexão das características vocais da função laríngea (Gama, 1997).

A análise acústica permite medir propriedades do sinal acústico de uma voz gravada, quer se trate de uma fala contínua ou uma vogal sustentada (Alves, 2016). Esta análise é capaz de fornecer o formato da onda sonora permitindo avaliar determinadas características como a frequência fundamental (F0), definida como o número de vibrações por segundo produzidas pelas cordas vocais; as medidas de perturbação da frequência, como o *jitter* e o *shimmer*, são considerados os principais parâmetros acústicos utilizados na detecção de patologias (Teixeira et al. 2011).

Embora existam vários sistemas de detecção de fonoaudiologia, esses sistemas não foram convertidos em dispositivos assistidos para o pessoal clínico. Isso levou ao desenvolvimento de um dispositivo para apoiar a tomada de decisões. Os exames utilizados nesta área, são invasivos e tornam-se desconfortáveis para os pacientes. O desenvolvimento de métodos de avaliação que possibilita o reconhecimento de patologias vocais com maior facilidade e rapidez, de forma eficiente e não invasivo, pode ser benéfico numa primeira análise de rastreio, como método complementar no diagnóstico de patologias vocais, assim sendo permite a disponibilidade de um sistema de pré-diagnóstico principalmente para regiões com acesso limitado a profissionais clínicos. (Fernandes et al. 2021).

Para que um sistema de auxílio ao diagnóstico possa ser implementado com eficiência, torna-se primordial o estudo de técnicas que ofereçam um excelente desempenho, para que o erro de diagnóstico seja mínimo (Costa et al., 2013).

### **1.4 Objetivos do trabalho**

As alterações da voz podem ocorrer em qualquer fase da vida, muitas vezes causando transtornos que afetam diretamente as relações no trabalho ou na escola e na vida familiar e social.

O processamento de fala encontra-se em constante progressão em áreas como o reconhecimento de fala, o reconhecimento de orador, a síntese de fala a partir de texto, a diarização e a codificação. Estas áreas são exemplos em que parâmetros extraídos do sinal de fala são usados com o intuito de reconhecer palavras, reconhecer pessoas, produzir palavras ou transmitir sinais, construindo aplicações de inegável utilidade.

O reconhecimento de vozes patológicas pode ser realizado com recurso a medidas que caracterizam o movimento das cordas vocais, tais como o *Jitter* e o *Shimmer*. No entanto, apesar das patologias afetarem tipicamente as cordas vocais, estas medidas não produzem resultados conclusivos no reconhecimento de patologias da voz.

Este trabalho tem como finalidade desenvolver uma interface gráfica para um sistema de apoio à decisão médica, com base no estudo e incorporação de ferramentas desenvolvidas em projetos anteriores, tais como, algoritmos que efetuam a medição dos parâmetros *Jitter*, *Shimmer*, Autocorrelação, HNR, bem como outras ferramentas de classificação de vozes patológicas e de identificação de patologias.

A determinação dos parâmetros acústicos utilizados neste trabalho teve por base um algoritmo desenvolvido anteriormente em linguagem MATLAB, por Gonçalves, (2015) e modificado por Fernandes et al., (2021) . Neste trabalho, esse algoritmo foi transcrito para linguagem *Python*.

No algoritmo constam-se os principais parâmetros acústicos usados na deteção de patologias, sendo estes, o *jitt*, *jitta*, *Shim*, *ShdB*, HNR, NHR, Autocorrelação, *Entropia de Shannon*, e a *Entropia logarítmica*, *F0*.

Para a classificação final do paciente como sendo patológico ou saudável, pretende-se utilizar uma rede neuronal *Multi Layer Perceptron* (MLP) implementada em linguagem *Python*.

## **1.5 Estrutura de tese**

O presente trabalho encontra -se dividido em 7 capítulos.

**capítulo 1** é feita uma introdução ao tema e alguns conceitos relativos ao funcionamento e anatomia do aparelho fonador, uma pequena descrição da análise do sistema acústico na deteção de patologias da fala e são enunciados os objetivos deste trabalho.

**Capítulo 2** é feita uma revisão da literatura dos trabalhos realizados anteriormente, relacionados ao tema do trabalho.

**Capítulo 3** aborda-se a descrição de algumas das patologias da fala.

**Capítulo 4** descreve a interface gráfica, os parâmetros acústicos que constam neste trabalho e a forma de os determinar. Por fim, é descrito o sistema de classificação utilizado neste trabalho.

**Capítulo 5** são apresentados os resultados obtidos relativamente à interface desenvolvida, bem como a explicação do funcionamento da mesma.

**Capítulo 6** é feita a discussão e a análise dos resultados.

**Capítulo 7** estão presentes as conclusões e algumas sugestões de trabalhos futuros.



## 2. ESTADO DE ARTE

Neste capítulo é feita uma abordagem teórica de alguns trabalhos realizados na área de diagnóstico da patologia da voz, usando parâmetros acústicos e redes neuronais artificiais.

Em Boersma, (1993) foi desenvolvido um algoritmo para a detecção da periodicidade baseado no método de Autocorrelação. Neste algoritmo são efetuados testes para sinais periódicos, com adição de ruído e variação de *Jitter*, portanto, mais preciso do que outros métodos utilizados habitualmente para análise de fala. O autor considera que o método é capaz de medir o parâmetro HNR no domínio do tempo com uma precisão e fiabilidade superior aos métodos que usam o domínio da frequência.

No trabalho de Martínez & Rufiner, (2000) implementaram um sistema para detecção automática de patologias da laringe usando análise acústica da fala no domínio da frequência. Esta análise utiliza alguns parâmetros que caracterizam os sinais: cepstrum, mel-cepstrum, delta cepstrum e delta mel-cepstrum e *fast Fourier transform* (FFT). Para a análise dos padrões de voz utilizaram 2 tipos de redes neuronais, uma para distinguir entre vozes normais e patológicas, seja qual for a patologia, e um sistema treinado para classificar a voz normal, bicíclica e áspera. Foram obtidas altas percentagens de reconhecimento, sendo que a análise cepstral foi a técnica que atingiu os valores mais altos. Com isso concluiu-se que esse tipo de análise fornece uma caracterização da voz em condição patológica de maneira direta e não invasiva. Os resultados obtidos tornam promissora a aplicação dessa alternativa como ferramenta de apoio ao diagnóstico de patologias do sistema vocal.

Em Dibazar et al., (2002) desenvolveram um sistema robusto, rápido e preciso para a detecção automática de fala normal e patológica. Os *Coefficientes Cepstrais de Frequência Mel* (MFCCs) e as medidas da dinâmica do *pitch* foram modelados por misturas gaussianas num classificador de modelo oculto de *Markov* (HMM). Para a avaliação desse método foi utilizada a vogal /a/ de mais de 700 indivíduos normais e de diferentes patologias do base de dados do *Massachusetts Eye and Ear Infirmary* (MEEI). Obtiveram uma taxa de acerto de 99,44% na classificação correta para discriminar entre fala normal e fala patológica.

Em Murphy & Akande, (2005) destaca-se o processo de estimativa da linha de base do ruído baseado no cepstrum, mostra-se análoga a ação de um filtro de média móvel aplicada ao espectro de potência da fala. A linha de base do ruído é influenciada pelo comprimento da janela e pela forma do espectro da fonte glotal. Foram testados sistematicamente usando fluxo glotal gerado sinteticamente e sinais de fala vozeada com conhecimento à priori do HNR, a duas técnicas de estimativa existentes. A influência da fonte foi removida utilizando uma nova técnica de pré-ênfase harmônico. Os resultados indicaram uma estimativa precisa do HNR usando a presente abordagem. Uma investigação preliminar do método com um conjunto de dados normais/patológicos é investigada.

Em Almeida, (2010) foi desenvolvido um sistema de classificação de vozes para auxiliar no pré-diagnóstico de patologias na laringe, bem como, no acompanhamento de tratamentos farmacológicos e pós-cirúrgicos. Como características relevantes do sinal de voz utiliza *coeficientes de Predição Linear (LPC)*, *Coeficientes Cepstrais de Freqüência Mel (MFCC)* e os coeficientes obtidos através da *Transformada Wavelet Packet (WPT)*. Para a classificação utilizou Máquina de Vetor de Suporte (SVM), pois, o objetivo era construir hiperplanos ótimos que maximizassem a margem de separação entre as classes envolvidas. O hiperplano gerado foi determinado pelos vetores de suporte, que são subconjuntos de pontos dessas classes. Os resultados apresentam um bom desempenho, com uma taxa de acerto de 98,46% na classificação de vozes normais e patológicas, e 98,75% na classificação de patologias entre si: edemas e nódulos.

Em Fezari et al., (2014) são usados como parâmetros os MFCC's em conjunto com *Jitter* e *Shimmer*, para a detecção de uma patologia chamada disfonia espasmódica. A base de dados usada é a *Saarbrücken Voice Database (SVD)* e o método de classificação é o *Gaussian Mixture Model (GMM)*. Este trabalho é feito com base na frase "Bom dia, como está" dita em alemão. O objetivo em usar a frase, por um lado é obter mais dados para treino, onde o GMM precisa de uma quantidade importante de dados especialmente quando se usa um número elevado de mistura (Gaussiana), por outro, a diversidade de dados que pode aumentar a precisão de um sistema. No pré-processamento é feita a remoção de silêncios e selecionadas apenas as vogais "a" e "u". Um aspeto importante é o uso de várias *frames* por pessoa. Ficando a classificação dependente de um valor de *threshold*. "Se mais de 70% das *frames* de um sinal forem atribuídos a determinada classe

então assume-se que todo sinal pertence aquela classe.” A melhor precisão obtida foi de 82,31%.

Em Teixeira & Fernandes, (2015) foi elaborada uma análise acústica da disfonia vocal, onde utilizaram estatísticas de um conjunto de parâmetros da voz como *Jitter*, *Shimmer* e HNR. Esta análise foi feita com indivíduos de controlo e de quatro patologias distintas: disfonia, disfonia funcional, disfonia hiperfuncional e disfonia psicogénica. Fizeram uma análise estatística dos três parâmetros utilizando 3 vogais a/, /i/ e /u/ e cada vogal em 3 tons, alto, baixo e normal, num total de 9 segmentos de fala para os 4 grupos patológicos e grupo de controlo utilizando a base de dados SVD. O grupo de controlo é composto por 34 pacientes do sexo feminino e 7 do sexo masculino. Para a disfonia funcional foram utilizados 22 sujeitos do sexo feminino e 7 do masculino. Na disfonia hiperfuncional utilizaram 23 pacientes do sexo feminino e 6 do masculino. Para a disfonia psicogénica usaram 21 mulheres e 8 indivíduos do sexo masculino e para a disfonia utilizaram 15 mulheres e 11 homens. Segundo os autores não foram encontradas diferenças de significância estatística entre os grupos patológicos, mas verifica-se que o parâmetro HNR não mostra uma forte capacidade discriminante, enquanto, *Jitter* e o *Shimmer* são relevantes para serem utilizados num sistema de diagnóstico de patologias da disfonia.

No trabalho de Gonçalves, (2015) foi desenvolvido um algoritmo capaz de medir corretamente os parâmetros da voz *Jitter*, *Shimmer* e HNR, em vozes patológicas, e prosseguir com a identificação destas patologias, recorrendo a sistemas inteligentes. Previamente foi feita uma análise de cada um destes parâmetros para vozes masculinas e femininas, ou entre vozes de controlo e vozes patológicas. Este algoritmo permite medir o *Jitter* de quatro formas distintas: *Jitter*, *Jitta*, *RAP* e o *PPQ5*; e a do *Shimmer* em: *Shim*, *ShsB*, *APQ3* e o *APQ5*. Foi feita uma comparação entre os valores obtidos pelo programa *Praat* com os resultados obtidos neste algoritmo. Para testar o algoritmo utilizou sinais sintetizados, com valores bem conhecidos para os parâmetros, sinais de voz normal (controlo) e sinais de voz patológicos retirados da base de dados SVD. Na comparação realizada utilizando o sinal sintetizado, o algoritmo produziu um erro inferior a 5 microssegundos ( $\mu s$ ) para o parâmetro *Jitta* e inferior a 0,1% para o *Shim*. As diferenças de valores entre o *Praat* e o algoritmo foram pouco significativas, quando comparado com vozes reais (sinais de vozes de controlo e patológicas). Além disso, foi feita também uma comparação estatística do comportamento dos parâmetros *Jitter* e *Shimmer* em sinais de voz de controlo e sinais de pacientes com laringite, disfonia hiperfuncional, disfonia

espasmódica, pólipos das cordas vocais e envelhecimento das cordas vocais. Contudo, apenas as três últimas patologias apresentam distinção estatisticamente significativa dos parâmetros em relação ao grupo de controlo.

No trabalho de Panek et al., (2015) foi utilizado um vetor de 28 parâmetros acústicos, que foram avaliados utilizando a análise de componentes principais (PCA), a análise da componente principal de *Kernel* (kPCA) e uma rede neuronal auto associativa (NLPCA). A partir destas avaliações, utilizando as vogais /a/, /i/ e /u/ em três tons, alto, baixo e normal., foi possível a determinação de quatro tipos de patologias (disfonia hipertónica, disfonia funcional, laringite e paralisia das cordas vocais). A classificação entre saudável e patológico foi feita separadamente para cada doença e género e teve-se em conta o número de pessoas de controlo ser igual ao número de pacientes para cada patologia, porém, não é referido o número de pessoas que foram utilizadas. Constatou-se como resultado níveis de eficiência de cerca de 100% e que os métodos kPCA e NLPCA podem ser considerados para a deteção da patologia das cordas vocais.

Nos estudos de Alves, (2016) foi utilizado dois conjuntos de parâmetros, um constituído por HNR e quatro medidas de *Jitter* e *Shimmer*, onde examinou a capacidade de predição deste conjunto, quando usados com apenas uma vogal e um tom e quando usado com várias vogais e vários tons. O outro conjunto de parâmetros é constituído por 12 coeficientes cepstrais, frequências e larguras de banda dos três primeiros formates, frequência fundamental, energia, potência, momentos espectrais de ordem zero, um, dois, três e curtose. Estes parâmetros serviram para aferir utilidade de outro tipo de parâmetros na deteção de patologias da laringe, no entanto, foram aplicadas técnicas de seleção de variáveis e redução da dimensão como regressão linear passo a passo e PCA. Na classificação entre patológico ou saudável, foram treinados dois tipos de sistemas, as RNA e SVM. Foi utilizada a base de dados SVD e como grupos patológicos foram usadas a disfonia e paralisia das cordas vocais, separando o género feminino do masculino, onde constam um total de 334 pessoas do sexo feminino e 196 do sexo masculino. Dos 334 pacientes do sexo feminino, 126 tinham paralisia das cordas vocais e 41 disfonia, e no sexo masculino 69 tinham paralisia das cordas vocais e 29 disfonia. Usando tanto a disfonia feminino como masculino como grupo patológico, foi possível obter uma precisão de 100% através da análise do primeiro conjunto de parâmetros (HNR, quatro medidas de *Jitter* e *Shimmer* para três vogais e três tons diferentes) e 78,9% usando a

paralisia das cordas vocais feminina como grupo patológico; 81,8% usando a paralisia das cordas vocais masculina como grupo patológico.

Em Cordeiro, (2016) foi proposto soluções que permitam identificar patologias da voz através do processamento de fala. Nos métodos utilizados envolveu alguns classificadores inteligentes, geralmente usados em reconhecimento de fala, tais como, SVM e modelo de misturas Gaussianas. Pretendeu modelar o trato vocal, como por exemplo os *mel-frequency cepstral coeficientes*, os *line spectral frequencies* e *mel-line spectral frequencies* a partir dos parâmetros utilizados nos classificadores. Foram utilizadas duas bases de dados: a base de dados da Universidade de São Paulo, constituída por 47 pessoas, divididas em 3 grupos: 16 pessoas saudáveis, 16 pessoas diagnosticadas com edema de *Reinke* e 15 diagnosticadas com nódulos. E a base de dados da MEEI, constituídos por 53 sujeitos saudáveis e 724 com patologia da voz. Também foi proposto o uso de fala contínua como sinal para a identificação de patologias. A Partir desta abordagem realizou testes onde usou três grupos: sujeitos saudáveis, sujeitos com patologia laríngeas fisiológicas (edemas e nódulos); e sujeitos com patologias laríngeas neuromusculares (paralisia unilateral das cordas vocais). Como resultado obteve uma taxa de acerto de 84% para os três grupos e, ainda, foi tido em conta outra abordagem, tendo por base a análise dos formates e a relação harmónica-ruído. Desta maneira, efetuou a implementação de um algoritmo simples baseado em árvores de decisão que permitiu uma taxa de reconhecimento de 95%.

Em Forero et al., (2016) utilizaram parâmetros do sinal glotal, para auxiliar na identificação de dois tipos de distúrbios vocais relacionados às patologias das cordas vocais: nódulo e paralisia unilateral. Os parâmetros do sinal glótico são obtidos através de um conhecido método de filtragem inversa e são usados como entradas para uma Rede Neural Artificial, uma Máquina de Vetores de Suporte e também para um Modelo de Markov Oculto, para obter a classificação e comparar os resultados dos sinais vocais em três grupos distintos: falantes com nódulo nas cordas vocais; falantes com paralisia unilateral das cordas vocais; e falantes com vozes normais, ou seja, sem nódulo ou paralisia unilateral presente nas pregas vocais. Foi utilizado uma base de dados composta por 248 gravações de voz (sinais de produção de vogais) que contém as amostras correspondentes aos três grupos, em que cada paciente tem 8 gravações, e existem 12 pessoas com nódulos, 8 com paralisia das cordas vocais e 11 saudáveis. Para o método de classificação utilizando redes neuronais artificiais obteve uma taxa de acerto de 95,8%,

máquinas de vetor de suporte 82% e cadeias de *Markov* escondidas 96.2%, onde se obtiveram taxas de acerto de 95,8%, 82% e 96,2% respetivamente.

Em Teixeira et al., (2018) desenvolveram um sistema classificador baseado em redes neuronais artificiais e/ou máquina de vetor de suporte para classificar, com grande precisão, sinais de fala entre as diferentes classes de laringite crónica e controlo. Nesta análise encontra-se uma análise estatística de um conjunto de parâmetros sobre os grupos envolvidos (grupo de controlo e grupo com laringite crónica). A análise foi realizada com as vozes dos dois géneros (masculino e feminino). Os parâmetros utilizados foram o *jitter*, *shimmer*, HNR, NHR e a autocorrelação extraídos do som das vogais sustentadas /a/, /i/ e /u/ nos tons baixo, alto e normal. Utilizaram a base de dados SVD. Foram comparados os parâmetros por género para ambos os grupos e posteriormente foram comparados o grupo patológico em cada parâmetro. Verificou-se que na primeira fase só há diferenças de voz no *jitter* absoluto entre género masculino e feminino no grupo de controlo. A comparação entre o grupo patológico e de controlo mostram conclusões semelhantes para os restantes 6 parâmetros. Estes parâmetros poderão ser importantes para usar como ferramenta de decisão inteligente para classificar entre laringite crónica e saudável.

No trabalho de Fernandes, (2018) foi desenvolvido um algoritmo com o objetivo de extrair os parâmetros HNR, NHR e autocorrelação, posteriormente utilizados como entrada de um sistema inteligente para diagnóstico de patologias da fala. Primeiramente foi realizada uma análise onde comparou os valores do algoritmo com os do *software Praat* de modo a perceber qual a melhor janela e o seu comprimento, em número de períodos glotais. Desta análise decidiu usar janela de *hanning* com um comprimento correspondente a 6 períodos glotais. Esta comparação feita levou a conclusão que o algoritmo permite extrair os parâmetros HNR, NHR e autocorrelação com valores suficientemente próximos dos valores de referência. Desenvolveu, também, outro algoritmo que elimina as zonas de silêncio no início e no final do sinal, selecionando assim, apenas as zonas onde ocorre fala, para posterior extração dos MFCCs, os *Linear Prediction Coefficients* (LPC) e os *Line Spectral Frequency* (LSF). Foi utilizada base de dados de ficheiros de fala alemã SVD, no entanto esta foi complementada adicionando, mais doenças e mais parâmetros, ficando com os parâmetros MFCC e com 13 coeficientes cepstrais, HNR, NHR, *autocorrelação*, *Jitta*, *Jitter*, *ShdB*, *Shim*, extraídos de 9 locuções correspondentes a 3 vogais em 3 tons (alto, normal, baixo), e a uma frase, para sujeitos com 19 patologias, mais os sujeitos de controlo. Os resultados obtidos foram bastante

bons e muito próximos dos de referência. Para o HNR, para os sujeitos de controlo, o erro Absoluto é inferior a 1,82 decibel (dB), e nos pacientes é inferior a 0,80 dB. Para a *Autocorrelação* em ambos os casos o erro Absoluto é inferior a 0,001 e para o NHR, nos sujeitos de controlo, o erro é inferior a 0,003.

No trabalho de Cordeiro & Meneses, (2019) foi feito um estudo comparativo entre três parâmetros espectrais na discriminação entre vozes saudáveis e patológicas. Os parâmetros avaliados envolvem a análise do primeiro pico espectral, a Potência Relativa da Componente Periódica, que corresponde a uma medida da relação sinal-ruído e a Inclinação Espectral da Banda Baixa. O *Low Band Spectral Tilt* é proposto como otimização do primeiro pico espectral, para solucionar os casos de erro de classificação, devido à degradação da qualidade vocal com a progressão da doença. Os três parâmetros são avaliados em dois bases de dados. O *Low Band Spectral Tilt* alcança os melhores resultados, com 100% de precisão no banco de dados da USP e 83,5% de precisão no base de dados MEEI.

No trabalho de Teixeira et al., (2020) foram experimentadas as redes neuronais artificiais ANN e SVM para classificar entre disfonia/controlo e paralisia das cordas vocais/controlo. O vetor foi composto por 4 parâmetros de *Jitter*, 4 parâmetros de *Shimmer* e uma relação Harmónico-Ruído, determinada a partir de 3 vogais diferentes em 3 tons diferentes, com um total de 81 traços. Utilizaram técnicas de seleção de variáveis e redução de dimensão, como agrupamento hierárquico, análise de regressão multilinear e PCA. Para a classificação entre disfónico e controlo obtiveram uma precisão de 100% para os grupos feminino e masculino, tanto para o ANN como para o SVM. Na classificação entre paralisia de cordas vocais e controlo foi alcançada uma exatidão de 78,9% para o grupo feminino utilizando SVM e 81,8% para o grupo masculino usando RNA.

Em Trinh, (2021) foi estudada uma abordagem de aprendizagem semi-supervisionado (SSL), aplicado num modelo generativo, para incorporar dados rotulados e não rotulados no processo de treino. Os modelos generativos explorados incluem uma rede adversarial generativa (GAN) e um *autoencoder* variacional (VAE). Foi utilizado o modelo GAN para diferenciar entre amostras de fala reais e falsas, assim como, classificar a amostra dada como saudável ou patológica. A abordagem foi testada usando três conjuntos de dados patológicos comumente usados: o conjunto de dados de doenças de *Parkinson* em espanhol (SPDD), o base de dados de voz de SVD e o base de dados de patologia de voz

árabe (AVPD). Compararam os desempenhos das abordagens baseadas em GAN e VAE, treinadas em dados rotulados e não rotulados, com uma abordagem supervisionada tradicional, baseada em uma rede neural convolucional (CNN), treinada apenas em dados rotulados. A abordagem baseada em SSL comparado com a CNN é mais preciso apenas em dados de fala patológica rotulados. De acordo com o resultado, a abordagem tem o potencial de aliviar o problema de escassez de dados rotulados na classificação patológica da fala e outras aplicações médicas onde a aquisição de dados rotulados é um desafio.

A partir dos trabalhos descritos no estado da arte conclui se que a análise acústica permite diagnosticar as patologias da voz de forma direta e não invasiva. Dos parâmetros acústicos os mais utilizados são o *Jitter*, o *Shimmer*, por serem os mais relevantes e que produzem resultados significativos. Conclui se também que apenas os parâmetros acústicos por si só não são conclusivos para o diagnóstico de patologias da voz, pelo que quando combinadas com redes neuronais artificiais tornam se conclusivos e mais precisos.



### **3. PATOLOGIA DA VOZ**

A voz é um dos principais elementos da nossa personalidade, fundamental na comunicação com os outros. É do conhecimento que as patologias da voz se tornam um transtorno preocupante na vida das pessoas, por essa razão é extrema importância o acesso a um rápido diagnóstico, tendo em consideração um tratamento eficaz e eficiente (Fernandes, 2018).

As patologias diretamente ligadas à laringe designam-se por patologias da voz ou patologias laríngeas. Existem várias lesões que podem causar estas patologias, tais como: lesões mínimas estruturais e/ou funcionais da laringe, lesões de massa localizada nas cordas vocais, alterações tecidulares da corda vocal, perturbações neurológicas e perturbações não orgânicas ou de tensão muscular (Cordeiro, 2016).

O reconhecimento de patologias vocais consiste na identificação da patologia específica que afeta a voz de um paciente. Inicialmente a avaliação vocal realizava-se de forma subjetiva através da análise preceptiva auditiva, porém, a falta de consenso entre os examinadores experientes, mesmo com o uso de diferentes escalas de alterações vocais, tornou necessária a pesquisa de uma avaliação objetiva, onde a voz fosse analisada através de aparelhos capazes de medir parâmetros acústicos (Fernandes, 2018).

#### **3.1 Paralisia de Cordas Vocais**

A paralisia das cordas vocais pode resultar de lesões ou disfunção no nível do núcleo ambíguo, seus tratos supranucleares, tronco principal do nervo vago ou nervos laríngeos recorrentes. A corda vocal esquerda é mais frequentemente paralisada do que a direita porque o nervo recorrente esquerdo tem um curso mais longo do tronco cerebral à laringe, o que fornece mais oportunidades de compressão, tração ou lesões cirúrgicas (Sasaki, 2020).

Quando a paralisia afeta apenas uma prega vocal, as cordas vocais vibram com frequências diferentes. Nestes casos, a voz apresenta um som bitonal e o paciente não consegue falar alto, perdendo o poder de amplificação vocal (Fernandes, 2018). Quando ambas as cordas vocais ficam paralisadas, a voz reduz-se em potência, embora

continue a soar normalmente. Contudo, o espaço entre as cordas paralisadas é pequeno e a via aérea é inadequada. Por esta razão, até mesmo um exercício moderado causa dificuldade em respirar e um som rouco e agudo (estridente), sempre que a pessoa respira (Sasaki, 2020).

### **3.2 Laringite Crónica**

A laringite crónica corresponde a uma inflamação persistente da mucosa laríngea, às vezes com muitos anos de evolução, normalmente provocada por infeções agudas repetidas. Afeta essencialmente as pessoas que estão em constante exposição a fatores irritantes, nomeadamente o tabaco, álcool, ambientes repletos de fumo, pó e vapores irritantes, abuso ou má utilização da voz, entre outros. A inflamação provoca, além de uma tumefação da mucosa laríngea, uma abundante produção de secreções, fatores responsáveis pelas manifestações que dão sob a forma de rouquidão tenaz, permanente, embora variável (Teixeira et al., 2011).

Os sintomas de laringite consistem numa alteração anómala da voz, como rouquidão ou diminuição do volume, ou até mesmo a perda de voz, que se desenvolve dentro de algumas horas ou de mais ou menos um dia. A garganta pode formigar ou ficar como se estivesse em carne viva e a pessoa pode sentir uma necessidade constante de limpar a garganta (Sasaki, 2020).

### **3.3 Cancro da Laringe**

O cancro da laringe ocorre predominantemente em homens acima de 40 anos e é um dos mais comuns entre os que atingem a região da cabeça e pescoço. Representa cerca de 25% dos tumores malignos que acometem essa área e 2% de todas as doenças malignas. A ocorrência pode se dar em uma das três áreas em que se divide o órgão: supraglote, glote e subglote. Aproximadamente 2/3 dos tumores surgem na corda vocal verdadeira, localizada na glote, e 1/3 atacam a laringe supraglótica (acima das cordas vocais). (INCA, 2022). No entanto, tumores supraglóticos (acima das cordas vocais) e subglóticos (abaixo das cordas vocais) quase sempre se manifestam em um estágio avançado, porque permanecem assintomáticos por longos períodos de tempo. Os sintomas para o tumor da laringe são a rouquidão, dificuldades em engolir, dificuldade em respirar, dor (às vezes correndo para o ouvido) e nódulos no pescoço (Estibeiro & Trindade, n.d.).

Praticamente todos os cânceros da laringe originam-se nas células escamosas, que revestem o interior do órgão e são chamados de carcinomas de células escamosas ou espinocelulares. Diversas doenças podem estar relacionadas a células anormais, mas a maioria não é cancerosa, como é o caso de uma alteração chamada displasia. Na maioria das vezes, principalmente quando leve, apresenta baixo risco de transformar-se em cancro e pode até mesmo desaparecer sem tratamento se o agente causador (o fumo, por exemplo) for eliminado. Já as displasias graves têm alto risco de transformação. Cânceros da laringe não costumam causar sintomas no início, a não ser quando ocorrem nas cordas vocais. (Camargo, n.d.).

### **3.4 Pólipo das Cordas vocais**

Os nódulos reativos, também denominados de pólipos, podem desenvolver-se nas cordas vocais, na grande maioria das vezes em fumantes crônicos ou em indivíduos que impõem grande esforço sobre as cordas vocais. Os adultos, principalmente os homens, são afetados com mais frequência. Devido à sua localização estratégica, estes alteram significativamente o caráter da voz e causam com frequência rouquidão progressiva (Robbins & Cotran, 2005).

### **3.5 Quistos**

Os quistos da laringe representam um grupo de lesões benignas raras, que podem causar obstrução respiratória importante e até mesmo a morte se não tratadas adequadamente. Apresentam dificuldade no seu diagnóstico, pois em regra são achados do exame objetivo, em pacientes com sintomas inespecíficos ou mesmo assintomáticos (Ribeiro et al., 2011).

Os quistos localizados nas cordas vocais são geralmente pequenos, não dificultando a respiração, produzindo antes rouquidão, tosse seca, de irritação, sensação de corpo estranho. Os localizados nas cordas aritenepiglóticas, nas falsas cordas, nos ventrículos de Morgagni, na face posterior do epiglote, geralmente volumosos e de origem embrionária, passam no início despercebidos, produzindo depois perturbações da respiração, mais ou menos intensa, de acordo com o volume e extensão do quisto. Os sintomas a serem observados variam, naturalmente, com a sede e o volume do quisto, as dimensões da laringe tais como, perturbações fonatórias, respiratórias e da deglutição

podem observar-se, ao lado da tosse, impressão de corpo estranho, salivação exagerada, regurgitação de líquidos pelas fossas nasais, etc. (Pereira, 1955).

### **3.6 Edema das Cordas Vocais (Reinke)**

É uma doença crónica da laringe, onde o espaço de Reinke é ocupado por muco espesso. Consoante o muco se vai acumulando, o espaço aumenta e as cordas vocais aumentam de espessura e dirigem-se para o interior da laringe. Esta doença provoca alterações na elasticidade das cordas vocais e, como consequência, a voz torna-se mais rouca e com uma tonalidade mais grave. Consequentemente, o paciente por norma faz um maior esforço vocal, provocando a abertura excessiva da glote e uma vibração assimétrica, irregular e aperiódica das cordas vocais (Cordeiro, 2016) (Martins et al., 2009).

O edema de Reinke altera, consideravelmente, as qualidades vocais, levando à diminuição acentuada da frequência fundamental. A voz da mulher torna-se masculinizada, sendo muitas vezes confundida, ao telefone, à do homem. (Koufman, et al., 2000). O tabagismo é o principal causador da doença, porém, abuso vocal, refluxo laringofaríngeo, hipotireoidismo e envelhecimento são fatores de risco associados. A voz do paciente com Edema de Reinke é caracteristicamente rouca e grave, em casos mais graves pode levar a obstrução da glote e dispneia (falta de ar) (César, n.d).

### **3.7 Granuloma**

Os granulomas, situados na laringe, são afeções orgânicas relativamente raras, que apresentam um quadro clínico e patológico bem definido. Estes podem dividir-se em dois grupos: específicos ou não específicos, sendo que, o primeiro aparecem em doenças sistémicas que apresentam manifestações laríngeas, como por exemplo, a tuberculose e a sífilis. Os granulomas específicos podem ser confundidos com o cancro da laringe. Os não específicos são tumores benignos que apresentam granulação bem definida e caracterizam-se por lesões que surgem, em geral, como uma massa localizada de tamanho variável, coloração esbranquiçada, amarelada ou avermelhada (Dieguez et al., 2010). Resultam de lesões das cordas vocais devido a refluxo gastroesofágico não controlado ou de danos causados durante uma intubação endotraqueal (inserção de um tubo de

plástico para respiração pela boca até a traqueia). Granulomas são comuns em adultos. (Sasaki, 2020).

### **3.8 Granulomas de Intubação**

Os granulomas de intubação são uma resposta inflamatória ao contato do tubo endotraqueal. Na maioria dos casos, desenvolvem-se em pacientes com intubação prolongada, mas podem ser diagnosticados em pacientes submetidos a períodos curtos de intubação, mesmo que de poucas horas (Helena et al., 2019). Outros fatores envolvidos no desenvolvimento do granuloma de intubação são: utilização de tubos traqueais de tamanho inadequado para o diâmetro da via aérea, pressões elevadas nos balonetes dos tubos traqueias, e plano de sedação inadequado, o que proporciona atrito entre o tubo e a mucosa da laringe e da traqueia (Martins & Dias, n.d.).

### **3.9 Fibroma Laríngeo**

O fibroma laríngeo é um tumor benigno localizado na laringe derivado do tecido conjuntivo, composto por fibroblastos e feixes de fibras. Os sintomas dependem da localização e tamanho da ocorrência. O tamanho do tumor é diferente, o pequeno é tamanho de grão de arroz e o maior pode bloquear o trato respiratório. Pode localizar-se em várias partes da garganta, porém, ocorre principalmente no meio das cordas vocais, podendo ser vista também na área subglótica, na zona ventricular ou no epiglote, é redonda ou oval, com superfície lisa, pedículo, acinzentada ou avermelhada e dura (Otorrinolaringologia, n.d.).

### **3.10 Tumor da Hipofaringe**

Os cancros da hipofaringe começam na parte inferior da garganta. Desenvolvem a partir de células finas e planas chamadas células escamosas, que compõem o revestimento interno dessa estrutura. Alguns fatores de risco, como o tabagismo ou o consumo abusivo de álcool, podem causar esses tipos de câncer por danificar o DNA das células que revestem o interior da laringe e hipofaringe (American Cancer Society, 2021).

### **3.11 Disfonia**

A disfonia é a alteração de algumas das qualidades acústicas da voz (intensidade, tom, timbre e duração) como consequência de um transtorno orgânico ou por um uso inadequado da voz (funcional), podendo ser permanente ou transitório (Clínica SanVicente, 2021). Embora existam muitas causas de disfonia, esta pode ser causada por uma disfunção, uso intensivo ou mau uso da voz e é mais frequente em indivíduos que usam a voz diariamente de forma abundante e incorreta. Podem ter como sintomas rouquidão, dor de garganta ou garganta seca. Um cantor ou cantora pode notar que já não é capaz de cantar em tons mais altos. Pode, ainda, ocorrer outros sintomas associados como um gotejamento contínuo na parte de trás da garganta (catarro nasal) e azia (Teixeira & Fernandes, 2015).

As disfonias podem ser classificadas etiologicamente em duas categorias vastas: os tipos orgânico e funcional. As disfonias orgânicas são aquelas causadas por alterações estruturais (como lesões de massa, inflamação das cordas vocais e malformações da laringe), alterações neurológicas (como paralisia das cordas vocais, doença de Parkinson e esclerose lateral amiotrófica) ou outros aspetos não relacionados com o uso da voz, tais como refluxo gastroesofágico/laringofaríngeo. Deste modo, quando não existe nenhuma causa estrutural ou neurológica, estamos perante uma disfonia não-orgânica, denominada de disfonia funcional (Simões Marta, 2017).

#### **3.11.1 Disfonia Funcional**

A disfonia funcional é assim definida como uma afonia ou disfonia em que não existe qualquer patologia orgânica ou, se existe, esta ou é insuficiente para explicar a natureza e gravidade do problema vocal ou é considerada secundária ao problema funcional (ex: nódulos ou pólipos) (Behlau et al., 2001).

A disfonia decorre do próprio uso da voz, quer isto dizer, quando ocorre um distúrbio do comportamento vocal. Esta pode ser causada devido ao uso incorreto da voz, uso abusivo da voz, inaptações vocais e alterações psicogénicas, que podem atuar de modo isolado ou concomitantemente (Baena, 2013).

### **3.11.2 Disfonia Hiperfuncional**

A fonação hiperfuncional é caracterizada por esforço fonatório excessivo. Este comportamento vocal inadequado pode causar alterações anatómicas e fisiológicas do aparelho vocal e por vezes trauma das cordas vocais (Caçador, 2019).

Desenvolve-se através da atividade excessiva da musculatura laríngea (intrínseca ou extrínseca) durante o processo de fonação. Esta disfonia pode ser uma resposta de compensação funcional a desalinhamentos posturais, fatores contextuais, padrões respiratórios ineficientes, tabaco, condições de saúde (alergias) e fatores compensatórios psicogénicos (Guimarães, 2007).

### **3.11.3 Disfonia Hipofuncional**

A patologia hipofuncional da laringe diz respeito a toda a patologia que se manifesta por insuficiência glótica e atrofia das cordas vocais, como alterações da mobilidade laríngea ou presbifonia. O conceito de que uma patologia Hipofuncional se pode manifestar como hiperfuncional é essencial para o correto diagnóstico destas alterações (Koufman, et al., 2000). Na alteração vocal por hipofunção há uma fraqueza dos músculos laríngeos, com fechamento incompleto da glote, devido à fraqueza muscular laríngea ou generalizada. A voz é fosca, sussurrada, sem timbre. Há um desejo de limpar a garganta, uma sensação de globus e dor no pescoço (Clínica Universidad da Navarra, 2022).

### **3.11.4 Disfonia Hipotónica**

É definida como uma alteração hiperfuncional da fonação, onde há atividade excessiva principalmente dos músculos extrínsecos da laringe (falta de força das cordas vocais), podendo levar a mudanças na sua função. Neste caso as cordas vocais não se fecham totalmente por falta de tensão muscular. A alteração da voz é restabelecida mantendo a postura adequada e a mecânica respiratória correta, reduzir a tensão muscular e eliminar os esforços laríngeos, bem como adaptar as cavidades de ressonância ao som emitido pela laringe (Clínica SanVicente, 2021).

### **3.11.5 Disfonia Espasmódica**

Disfonia espasmódica é a contração involuntária dos músculos na laringe, que controlam as cordas vocais, resultando em uma voz anormal. Na disfonia espasmódica, as pessoas podem não conseguir falar ou a voz pode soar forçada, trêmula, rouca, sussurrada, saltada, rangida, destacada ou distorcida e pode ser difícil de entender. Os médicos não sabem o que causa a disfonia espasmódica, que é mais comum entre 30 e 50 anos de idade e no sexo feminino (Sasaki, 2020).

Na disfonia espasmódica adutora, os músculos da laringe sofrem espasmo e as cordas vocais fecham-se. Geralmente quando os sons de vogais estão a ser formados, no início das palavras, geram um som espremido ou forçado. Na disfonia espasmódica abduutora, os espasmos causam a abertura excessiva das cordas vocais, gerando um som de voz fraco e sussurrado (Sasaki, 2020).

### **3.11.6 Disfonia Psicogénica**

A disfonia psicogénica consiste numa “afonia ou disfonia que ocorre como resultado de processos psicológicos alterados, em que, ocorre uma perda súbita ou intermitente do controlo volitivo, sobre o início e manutenção da fonação na ausência de patologia estrutural ou neurológica, suficiente para explicar a disfonia (Simões Marta, 2017).

As disfonias psicogénicas são aquelas que se apresentam, juntamente, com stress emocional agudo ou crónico e que surgem como manifestação de um desequilíbrio psicológico como ansiedade, depressão, perturbação da personalidade ou reação de conversão, ao ponto de o normal controlo volitivo da fonação ser perdido (Aronson, 1990).

## **3.12 Efeitos das patologias da voz no sinal de fala**

As patologias das cordas vocais e nomeadamente da laringe afetam a produção do sinal de fala (Gordon & Ladefoged, 2001).

As vozes soprosas são caracterizadas pela passagem indesejada de ar pelas cordas vocais, sem que este provoque a vibração adequada das mesmas. A voz, apesar de vozeada, apresenta algum ruído de origem soprosa devido às cordas vocais não terem capacidade

para fechar adequadamente. No limite, o vozeamento é nulo, sendo que, neste caso está-se na presença de uma voz sussurrada. Por outro lado, as cordas vocais permanecem fechadas com mais pressão e estão mais tensas que o normal, pelo que apenas a parte anterior das cordas vocais permite a passagem de ar. A parte posterior não vibra ou vibra com uma frequência diferente da parte anterior. Neste caso, está-se na presença de uma voz crepitante. As alterações na voz são perceptíveis através da qualidade vocal com alterações no timbre da voz, nomeadamente, através de uma frequência fundamental muito baixa. No limite, a glote pode mesmo permanecer fechada sem que ocorra vibração nem passagem de ar pelas cordas vocais na produção de fonemas vozeados (Cordeiro, 2016).



## 4. METODOLOGIA

Neste capítulo efetua-se a descrição das interfaces do sistema, a classificação do sistema quanto à usabilidade e a sua importância nas áreas da saúde. São também descritos os parâmetros acústicos utilizados neste trabalho e a forma de como são extraídos. Por fim, descreve-se o sistema de classificação usando redes neurais artificiais com uma arquitetura *Multi Layer perceptron*.

### 4.1 Interface gráfica do utilizador

A *Graphical User Interface* (GUI), (ver Figura 4) é a parte visível do *software* para o utilizador, através da qual ele comunica com o sistema para realizar tarefas. Ela possui vários elementos que são percebidos como: cor, imagem, brilho, contraste, forma, entre outros. A consistência desta percepção visual é atingida quando o utilizador pode manipular "intuitivamente" a representação visual, configurando-a na relação entre o significado pretendido pelo desenvolvedor e o significado percebido pelo utilizador (Kulpa, 2009).

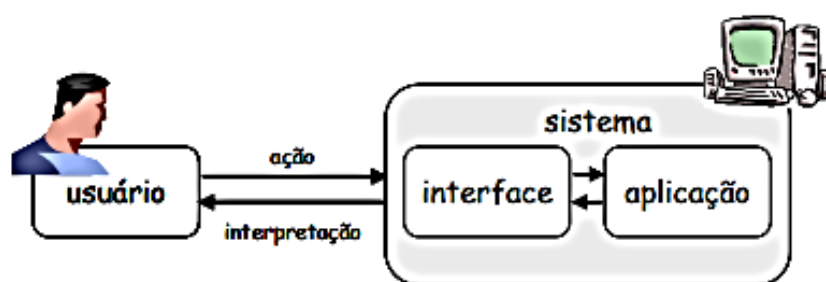


Figura 4- Processo de interação utilizador-interface (Prates & Barbosa, 2003).

#### 4.1.1 Benefícios do uso de sistemas na área da saúde.

Hoje em dia, é quase impossível imaginar um serviço de cuidados de saúde sem que este utilize um sistema computadorizado de informação, pelas inúmeras vantagens que podem trazer, tendo como primeiro objetivo apoiar os funcionários nas suas funções e, para além disso, oferecer um melhor serviço ao utente (Ammenwerth et al., 2003).

Um design de interface do utilizador bem-sucedido contribui para uma experiência do utilizador positiva, o que se reflete numa vantagem competitiva. Uma das vantagens evidentes é o uso de metáforas que são habituais para o utilizador na vida real como, por exemplo, arrastar e soltar para mover os documentos de um local para outro, ou utilizam ícones relacionados como, por exemplo, um caixote de lixo para que o utilizador entenda que está a eliminar o ficheiro, pois estes os tornam interativa para com o utilizador (Ebit, 2022). Por outro lado, a utilização dessas tecnologias contribui para apoiar os trabalhadores da área da saúde, proporcionando, por exemplo, uma maior facilidade no acesso à informação, e os utentes, pela possível melhoria e redução nos tempos de espera (Ammenwerth et al., 2003).

Num outro estudo Ammenwerth et al., (2004) mencionam que as grandes vantagens da utilização destes sistemas são, a facilidade, velocidade e quantidade de informação relativa aos utentes que os funcionários podem obter os sistemas de apoio nas decisões, que suportam as decisões clínicas.

#### **4.1.2 Métricas para a classificação de uma interface gráfica**

Os métodos de avaliação da interface diferem entre si em vários aspetos. É preciso entender as diferentes características de cada método, para se definir qual deles é o mais apropriado para se avaliar a interface de um software num determinado contexto. As principais diferenças entre os métodos são as etapas do ciclo de design do *software* em que devem ou podem ser aplicados (durante o ciclo de desenvolvimento ou após ter o produto pronto), a técnica utilizada para recolher os dados (desde entrevistas até experiências em laboratórios), os tipos de dados recolhidos (quantitativos ou qualitativos), e ainda o tipo de análise feita (o avaliador pode prever potenciais problemas ou interpretar os dados obtidos) (Prates & Barbosa, 2003).

Uma boa interação do utilizador com o sistema é um objetivo perseguido por projetos de design de interfaces, sendo uma necessidade reconhecida e divulgada. Esse objetivo vem, ao longo do tempo, ampliando-se e tornando-se cada vez mais complexo, no mesmo ritmo em que os produtos com interfaces digitais tornam-se mais populares e atendem a cada vez mais necessidades (Alves et al., n.d.). Um sistema completo, que possui todas as funcionalidades necessárias para funcionamento pleno de um hospital, pode não obter

sucesso junto aos seus utilizadores, caso as interfaces gráficas do sistema apresentem deficiências de usabilidade (Gonçalves, 2001).

No entanto a usabilidade de um sistema é considerada um fator que assegura que os produtos são fáceis de usar, eficientes e agradáveis do ponto de vista do utilizador e que a faz ser classificada nas seguintes metas: ser eficaz no uso, ser eficiente no uso, ser segura no uso, ser de boa utilidade, ser de fácil aprendizagem e fácil de lembrar como se usa (Preece et al., 2003).

### **4.1.3 Linguagem de programação**

O *Python* é uma linguagem de programação interpretada de alto nível e que suporta múltiplos paradigmas de programação: imperativo, orientado a objetos e funcional. É uma linguagem com tipagem dinâmica e gerenciamento automático de memória (Devino, 2021).

Possui um grande número de *frameworks* ou conjunto de ferramentas para construir uma GUI, indo desde a tradicional *TkInter* até as soluções de plataformas cruzadas, assim como, tecnologias nativas. Um dos seus maiores atrativos é possuir um grande número de bibliotecas, nativas e de terceiros, tornando-a muito difundida e útil numa grande variedade de setores dentro de desenvolvimento web, e, também, em áreas como análise de dados, *machine learning* e inteligência artificial (IA).

Para o desenvolvimento desta interface foi usada a linguagem de programação *Python*, a partir do ambiente *VsCode*, usando a biblioteca *AppJar*.

### **4.1.4 Leitura do código *Matlab* para *Python***

Neste trabalho efetuou-se a leitura do algoritmo para posterior extração dos parâmetros acústicos. São descritos abaixo os passos efetuados para a extração dos parâmetros a partir do *Python*.

O primeiro passo foi a tentativa de transcrever o código manualmente de *Matlab* para *Python*, mas visto que o algoritmo é muito extenso, o que torna bastante trabalhoso, por isso foi descartada essa hipótese.

No entanto surgiu a hipótese de converter o código a partir de conversores *Matlab* para *Python*, tais como *matlab2python*, *SMOP (Small Matlab and Octave to Python)*, *OMPC (Open-Source Matlab-to-Python Compiler)*, *Mat2py*, porém todos estes funcionam bem

quando se trata de códigos mais simples, que não é o caso deste algoritmo, visto que o código é bastante complexo.

A partir da pesquisa feita de como ler o algoritmo dentro do *Python* sem a necessidade de ser transcrito, foi descoberto que podia ser lido a partir do *Python* usando o módulo “*matlab.engine*” fornecido pela *Mathworks*. Este módulo foi então importado a partir do *Python* de modo a permitir a leitura do algoritmo.

Para que a leitura seja feita de forma correta foi então reescrito algumas partes do código, devido a algumas algumas incompatibilidades entre as duas linguagens.

O trecho do código abaixo mostra como é chamado o *Matlab* a partir do *Python* utilizando esse módulo:

```
import matlab.engine  
  
eng = matlab.engine.start_matlab()
```

## 4.2 Parâmetros extraídos do sinal acústico

Nesta secção são detalhados todos os parâmetros acústicos utilizados e a forma de como os mesmos foram extraídos.

A avaliação acústica ou análise acústica, quando utilizada para avaliar distúrbios vocais, permite a quantificação não invasiva das características da voz humana. Essa técnica fornece medidas objetivas, cujos os dados são extraídos automaticamente, por meio de um processo computacional apropriado, como, por exemplo, uma estimativa direta dos padrões vibratórios das cordas vocais, bem como as formas do trato vocal supraglótico e as suas respectivas modificações (Awan & Roy, 2009). Para isso, é necessário utilizar diferentes parâmetros acústicos que compõem o sinal - periodicidade, amplitude, duração e composição espectral - que caracterizam os atributos físicos da voz nos domínios do tempo, frequência e intensidade, além de outras medidas complexas, que combinam o cruzamento desses domínios (Buder, 2000).

Procedeu-se à extração desses parâmetros a partir do algoritmo desenvolvido por Gonçalves, (2015), modificado por Fernandes et al., (2021) , de forma a tornar o

algoritmo mais preciso onde também foi acrescentado os parâmetros NHR e a Autocorrelação.

O algoritmo foi usado para extrair 6 parâmetros, dos quais foram representados na interface gráfica do sistema, o *Jitter*, *Jitta*, *Shim*, *ShdB*, HNR, e Autocorrelação, para três vogais e três tons diferentes a partir de sinais de vozes reais. As vogais disponíveis são /a/, /i/ e /u/ em tons, alto, normal e baixo.

O *Jitter* e o *Shimmer*, (ver Figura 5), representam as variações que ocorrem na frequência fundamental. Enquanto o *Jitter* indica a variabilidade ou perturbação da frequência fundamental, o *Shimmer* refere-se a essa mesma perturbação, mas relacionada à amplitude da onda sonora, ou intensidade da emissão vocal.

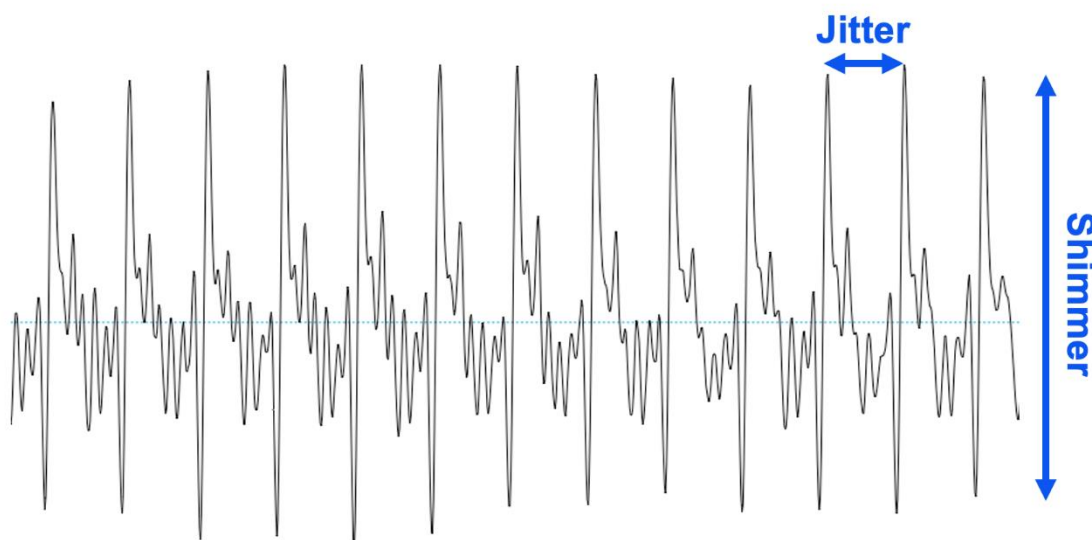


Figura 5-Representação dos parâmetros *jitter* e *shimmer* para um sinal de fala (Bäckström, 2022).

#### 4.2.1 Jitter - perturbação da frequência a curto prazo

O Jitter é uma medida de curto termo, isto é, apresenta variabilidade não voluntária na F0 ciclo a ciclo. Altera-se principalmente com a falta de controlo de vibração de cordas vocais (Teixeira et al. 2011). As vozes dos pacientes com patologia geralmente têm valor de *Jitter* maiores. A maioria dos investigadores considera como valor típico “normal” a variação entre 0,5 e os 1,0% para as fonações sustentadas em jovens adultos (Guimarães,

2007). O *Jitter* pode ser medido de quatro maneiras diferentes (Teixeira & Gonçalves, 2014) , *Jitter*, *Jitta*, *RAP* e o *PPQ5*.

#### 4.2.1.1 *Jitter Absoluto (Jitta)*

É uma variação que ocorre com a diferença absoluta média entre períodos glotais consecutivos, de ciclo a ciclo, representada na Eq.1.

$$Jitta = \frac{1}{N-1} \sum_{i=2}^N |T_i - T_{i-1}| \quad (1)$$

#### 4.2.1.2 *Jitter Relativo (Jitter)*

A sua definição é idêntica à do *Jitter Absoluto* só que, neste caso, a diferença absoluta média entre os períodos glotais consecutivos é dividida pelo período médio e expresso em percentagem, Eq.2.

$$Jitter = \frac{\frac{1}{N-1} \sum_{i=2}^N |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \times 100 \quad (2)$$

Na equação 1 e 2  $T_i$  corresponde à duração do período glotal  $i$  e o  $N$  é o número total de períodos glotais.

#### 4.2.1.3 *Jitter (RAP)*

A média relativa da perturbação (Relative Average Perturbation), é a diferença média absoluta entre um período e a média desse com os seus dois vizinhos, dividido pelo período médio, em percentagem, Eq.3.

$$RAP = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| T_i - \left( \frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} \times 100 \quad (3)$$

#### 4.2.1.4 *Jitter (PPQ5)*

O quociente de perturbação num intervalo de cinco períodos (*five point Period Perturbation Quotient*) é a média absoluta entre um período e a média desse com os seus quatro vizinhos, dividido pelo período médio, em percentagem, Eq.4.

$$PPQ5 = \frac{\frac{1}{N-3} \sum_{i=3}^{N-2} \left| T_i - \left( \frac{1}{5} \sum_{n=i-2}^{i+2} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} \times 100 \quad (4)$$

Neste trabalho, apenas são utilizadas duas dessas variações, o *Jitter* e o *jitta* visto que a análise estatística de Teixeira & Fernandes, (2015) o *Jitter* Relativo apresentou resultados semelhantes com o *RAP* e *PPQ5*.

#### 4.2.2 *Shimmer - perturbação da frequência a curto prazo*

O *Shimmer* está relacionado com a variação de magnitude ao longo dos períodos glotais. Uma redução na resistência glotal e lesões de massa nas cordas vocais, podem causar variações na magnitude glotal com a presença de ruído à emissão e com a sopro-sidade, dando origem a valores maiores de *Shimmer*. Este pode ser medido de quatro maneiras diferentes (Teixeira & Gonçalves, 2014) , *Shim*, *ShsB*, *APQ3*, e o *APQ5*.

##### 4.2.2.1 *Shimmer Relativo (Shim)*

É conhecido como a diferença média absoluta entre as magnitudes de dois períodos consecutivos, dividido pela amplitude média, em percentagem Eq.5.

$$Shim = \frac{\frac{1}{N-1} \sum_{i=2}^N |A_i - A_{i-1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (5)$$

##### 4.2.2.2 *Shimmer Absoluto (ShdB)*

É expresso como a variação da magnitude pico-a-pico em dB, ou seja, média absoluta do logaritmo de base 10 do quociente de amplitudes de dois períodos consecutivos, multiplicado por 20, Eq.6.

$$ShdB = \frac{1}{N-1} \sum_{i=2}^N |20 \times \log (A_i \div A_{i-1})| \quad (6)$$

Na equação 5 e 6, o  $A_i$  são os dados de amplitude de pico a pico extraídos e o  $N$  é o número de períodos de frequência fundamental extraídos.

#### 4.2.2.3 *Shimmer (APQ3)*

é definida como o quociente de perturbação da amplitude de três pontos (*three-points Amplitude Perturbation Quocient*) é a média absoluta entre a amplitude de um período e a média das amplitudes dos dois vizinhos, dividido pela amplitude média, Eq.7.

$$APQ3 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| A_i - \left( \frac{1}{3} \sum_{n=i-1}^{i+1} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (7)$$

#### 4.2.2.4 *Shimmer (APQ5)*

O quociente de perturbação da amplitude de cinco pontos (*five-points Amplitude Perturbation Quocient*) é a média absoluta entre a amplitude de um período e a média das amplitudes dos quatro vizinhos, dividida pela amplitude média, em percentagem, Eq.8.

$$APQ5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} \left| A_i - \left( \frac{1}{5} \sum_{n=i-2}^{i+2} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (8)$$

Neste trabalho serão utilizadas as medidas, *Shim* e o *ShdB*, isto devido a análise anterior realizada por Teixeira & Fernandes, (2015) em que ficou evidenciado que o *Shimmer* relativo tem resultados semelhantes aos *APQ3* e *APQ5*.

### 4.2.3 Parâmetros Harmônicos

As características harmônicas da voz podem ser medidas em três parâmetros, *Harmonic to Noise Ratio* (HNR), *Noise do Harmonic Ratio* (NHR) e a Autocorrelação.

#### 4.2.3.1 *Harmonic to Noise Ratio* (HNR)

A proporção harmônica-ruído, permite medir a relação entre os componentes harmônicos e de ruído de um sinal de fala, onde a periodicidade do sinal é fornecida, quantificando a relação entre a parte harmônica e o ruído. O valor HNR de um sinal pode variar, pois existem diferentes configurações do trato vocal, proporcionando diferentes amplitudes para os harmônicos (Teixeira & Fernandes, 2015). Assim, essa medida avalia a presença de ruído no sinal de voz analisado, apresentando relação direta com a qualidade vocal, determinando a percepção geral do ruído e da rouquidão no sinal vocal. Portanto, quanto maior for o HNR, melhor será a qualidade vocal (Scarpel & Fonseca, 2004).

A determinação do HNR utiliza algoritmos no domínio do tempo. O primeiro pico da Autocorrelação normalizada é considerado a energia da componente harmônica do sinal, e a energia restante é considerada como a energia do ruído. A componente ruidosa é dada pela diferença entre 1 e a energia harmônica representada na Eq.9.

$$HNR = 10 \times \log_{10} \frac{A}{1 - A} \quad (9)$$

Nesta equação,  $A$  é a componente harmônica fornecida pela energia do primeiro pico da Autocorrelação normalizada do sinal.

#### 4.2.3.2 Autocorrelação

A Autocorrelação fornece a medida das partes semelhantes do discurso, repetidas ao longo do sinal. Altos valores de Autocorrelação significam um número maior de repetições ao longo do sinal (Fernandes et al., 2019). A determinação da Autocorrelação de um sinal é feita multiplicando a Autocorrelação normalizada de um segmento do sinal de fala pela Autocorrelação normalizada de uma janela (*Hanning*). Assim, o primeiro pico do segmento de sinal é considerado a Autocorrelação. Este processo é repetido,

segmento a segmento, até o final do sinal. Finalmente, a Autocorrelação do sinal é a média de todas as Autocorrelações dos segmentos.

#### 4.2.3.3 *Noise to Harmonic Ratio (NHR)*

NHR quantifica a relação entre a componente aperiódica (ruído) e a componente periódica (parte harmónica), sendo o inverso de HNR, contudo, a medida não é feita no domínio logarítmico, nem os valores são o inverso (Boersma, 2004).

$$\text{NHR} = \frac{N}{H} = \frac{H - A}{H} = \frac{1 - A}{1} = 1 - A \quad (10)$$

O NHR é determinado tendo em conta a equação (10), em que H é a energia do sinal. Sendo determinada pela Autocorrelação normalizada, este valor é unitário.

### 4.3 Modelo de classificação MLP

Para a classificação final dos pacientes como saudável e patológico foi implementado o modelo de redes neuronais do tipo MLP, no *Python*, na plataforma Keras treinada, usando um conjunto de dados de treino, e validada usando um conjunto de dados de teste. A arquitetura de uma rede MLP consiste no arranjo topológico dos neurónios com os respetivos valores de pesos associado às conexões. O treino da rede se resume no ajustamento dos pesos sinápticos para fazer com que a resposta real da rede se aproxima da resposta desejada (Charu, 2018).

Para o treino do modelo implementado, é usado um conjunto de dados de treino, constituídos por 194 sujeitos saudáveis e 350 sujeitos patológicos distribuídos por 17 patologias na qual estão Disfonia Hiperfuncional (89 sujeitos), Paralisia de Cordas Vocais (74), Disfonia Funcional (51), Disfonia Psicogênica (31), Disfonia Espasmódica (24), Laringite Crônica (23), Pólipo de Cordas Vocais (14), Edema de Reinke (14), Disfonia Hipofuncional (9), Carcinoma de Corda Vocal (7), Tumor da Hipofaringe(4), Cisto (3), Granuloma (2), Disfonia Hipotônica (2), Tumor Laríngeo(1), Granuloma de Intubação (1), Fibroma (1).

Como cada sujeito possui 9 arquivos de som, isto é, 3 vogais 3 tons, para cada arquivo de som são extraídos 13 parâmetros (*Jitter*, *Jitta*, RAP, PPQ5, *ShdB*, *Shim*, APQ3, APQ5, frequência fundamental (F0), HNR, Autocorrelação, entropia de *Shannon* e entropia logarítmica), consumando 118 parâmetros (13 parâmetros x 9 áudios + sexo), por sujeito.

Para o treino foi utilizada 75 por cento do *Dataset*, e para o teste foi utilizada 25 por cento.

A matriz de entrada ficou composta por 118 nós na camada de entrada, 1 nó na camada de saída, 250 nós na camada escondida, a função de ativação da camada escondida é uma ReLU, e a função de ativação da camada de saída é sigmoideal. Para otimizar os valores dos parâmetros de modo a reduzir o erro cometido pela rede, foi utilizado o *Stochastic gradient descent optimizer* (SGD), a função de perda foi a *binary\_crossentropy* normalmente utilizada para modelos de classificação binária. Para avaliar o desempenho do modelo foram utilizadas as métricas, *Mean Squared Error* (MSE), calculado como a média das diferenças quadráticas entre os valores previstos e reais, *Mean Absolute Error* (MAE), a média da diferença absoluta entre os valores reais e previstos, e a *accuracy*. O modelo foi treinado em 300 épocas



## 5. RESULTADOS

Neste capítulo são apresentados os resultados das várias interfaces do sistema de ajuda ao diagnóstico de patologias da fala, desenvolvida neste trabalho, bem como, os resultados referentes ao modelo de classificação, usado para a classificação final do paciente. A primeira interface do sistema pode ser explicada na Figura 6.

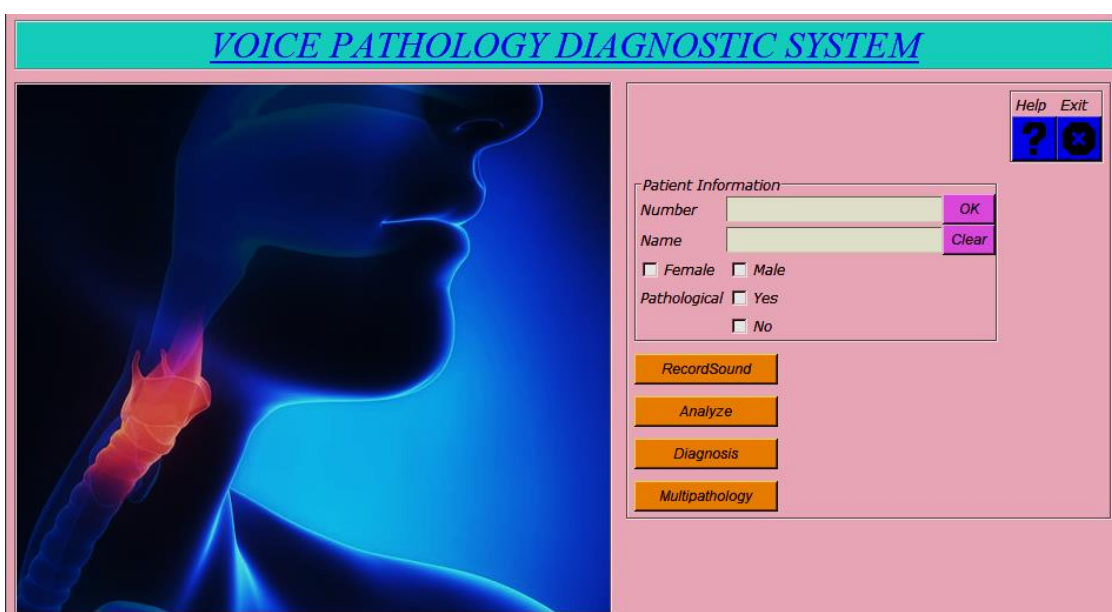


Figura 6- Representação do *Front-end* do sistema.

O *Front-end* do sistema apresenta, um *label* “*Patient Information*”, onde são introduzidas as informações do paciente já existentes na base de dados, incluindo nome, número, género, e se a pessoa já possui alguma patologia.

Apresenta também 4 botões, o “*RecordSound*”, “*Analyze*”, “*Diagnosis*”, e a “*Multipathology*”, e ainda dois botões “*Exit*”, “*Help*”.

Ao clicar no botão “*Exit*” o sistema perguntara se deseja terminar a sessão, e se for pressionado o botão “*Help*” o sistema fornecerá ajuda, sobre a funcionalidade para uma melhor utilização, Figura 7.

A funcionalidade dos restantes botões é apresentada a seguir.

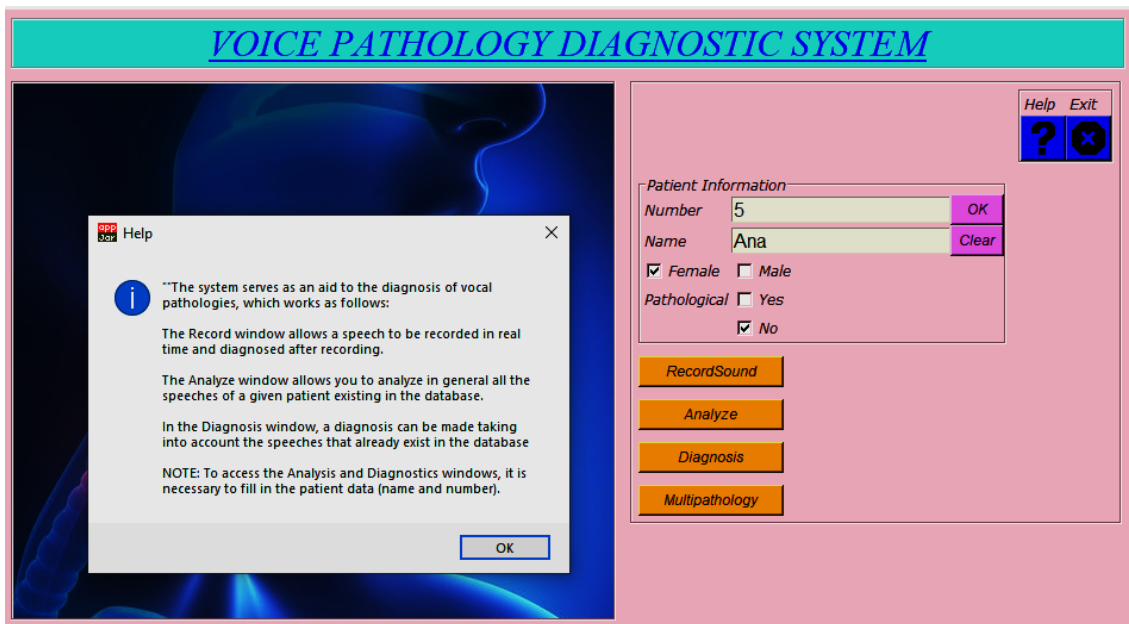


Figura 7- Front-end do sistema quando pressionado o botão *Help*.

### 5.1 Janela *RecordSound*

A janela “*RecordSound*” (Figura 8 e 9) é usada para gravar uma fala. Antes de se dar início à gravação, é necessário introduzir os dados do paciente no *label* “*Patient Information*” situada dentro desta mesma janela, tais como nome, número, idade, género, pois, são considerados obrigatórios. Caso estes dados não estejam preenchidos, o sistema não avança para a gravação. Caso o sujeito seja patológico, existe a possibilidade de ser selecionada a patologia no botão “*Pathology*” ou então escrever a patologia na caixa “*others Pathology*”.

O botão ícone interrogação ajudará o utilizador a perceber o funcionamento da janela

O botão “*Submit*” permite adicionar o paciente à base de dados.

O botão “*Clear*” limpará todas as caixas de dados, caso algum dado tenha sido introduzido erradamente. E o botão “*Cancel*” cancelará a gravação, caso o pretenda fazer.

A janela possui nove botões de gravação isto é cada botão pertence a um vogal (“a”, “i”, “u”) em tons (alto, baixo e normal). Quando se clica em um desses botões, o sistema iniciará a gravação em simultâneo com a representação do mesmo, (ver Figura 9).

No botão “*Exemple*” pode-se ouvir um áudio exemplar, referente ao sinal que quer gravar. Após a gravação do sinal, o mesmo pode ser reproduzido, escolhendo a gravação que quer ouvir e, de seguida, selecionar o botão “*Play Record*”.

A seta “*back*” fechará a janela *RecordSound* e retornará à janela principal do sistema.

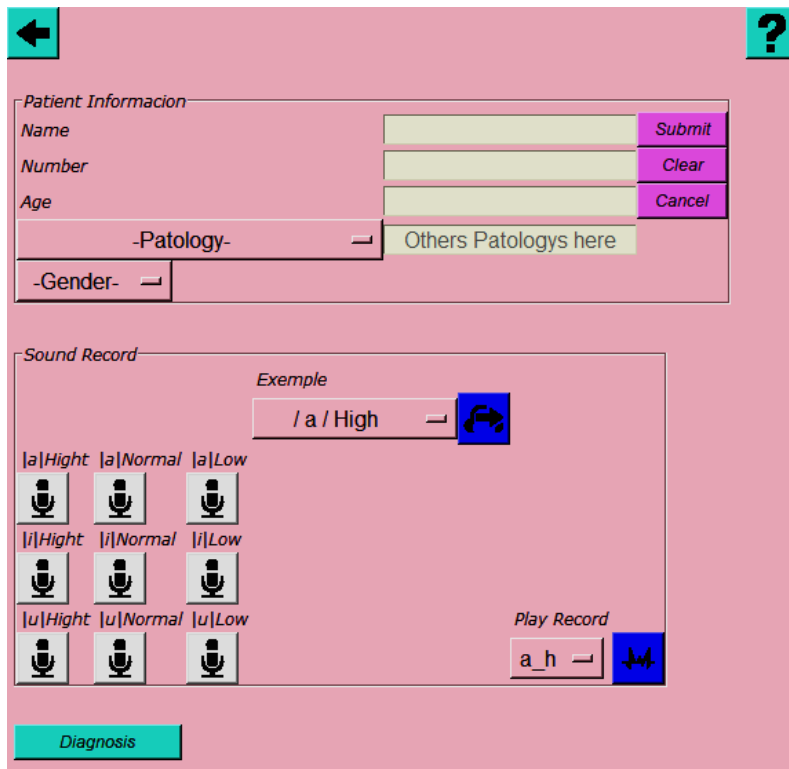


Figura 8- Janela *RecordSound*.

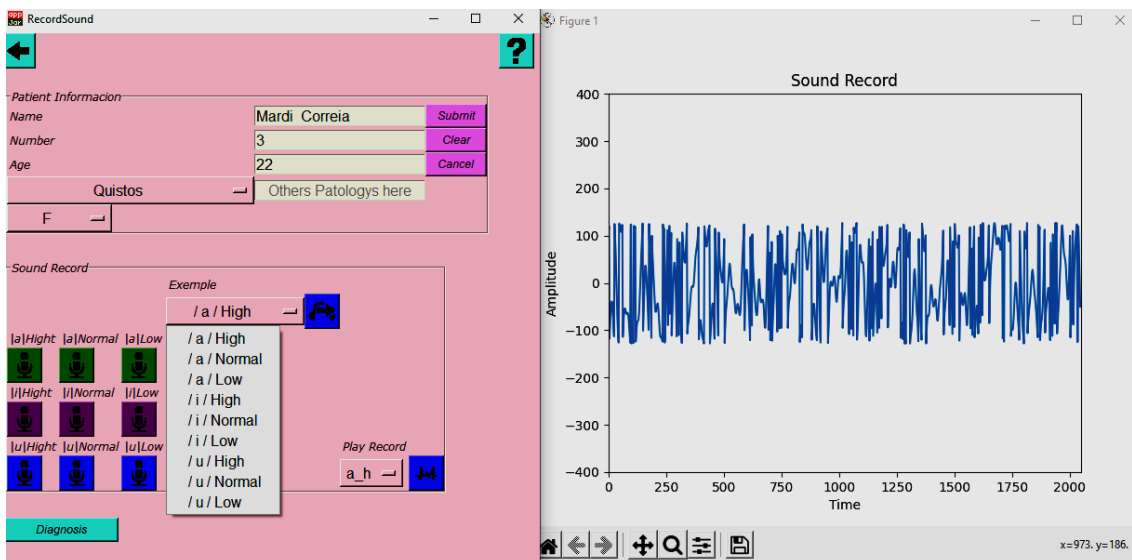


Figura 9- Janela *RecordSound*, ao efetuar uma gravação.

E por fim dentro deste mesmo subwindow temos a janela, “*Diagnosis*”, representado na Figura 10, que faz o diagnóstico com base na determinação dos parâmetros acústicos das vogais gravadas, usando o classificador MLP. Pois o mesmo será feito após a gravação

de todos os vogais e tons. Para ter acesso aos parâmetros de cada vogal gravado, basta clicar no botão referente à vogal.

O botão “*Google Search*” permite ser direcionado ao google, caso o utilizador necessite pesquisar alguma coisa.

O resultado da classificação do paciente é apresentado em percentagem, apesar de o valor de saída da função não corresponder diretamente a uma percentagem, pois a sua função de ativação é sigmoïdal, nesta fase de desenvolvimento considera-se que a saída corresponde a uma percentagem, em versões posteriores, este valor deverá ser convertido numa percentagem usando a função *softmax* e tendo 2 nós na camada de saída da rede neuronal.

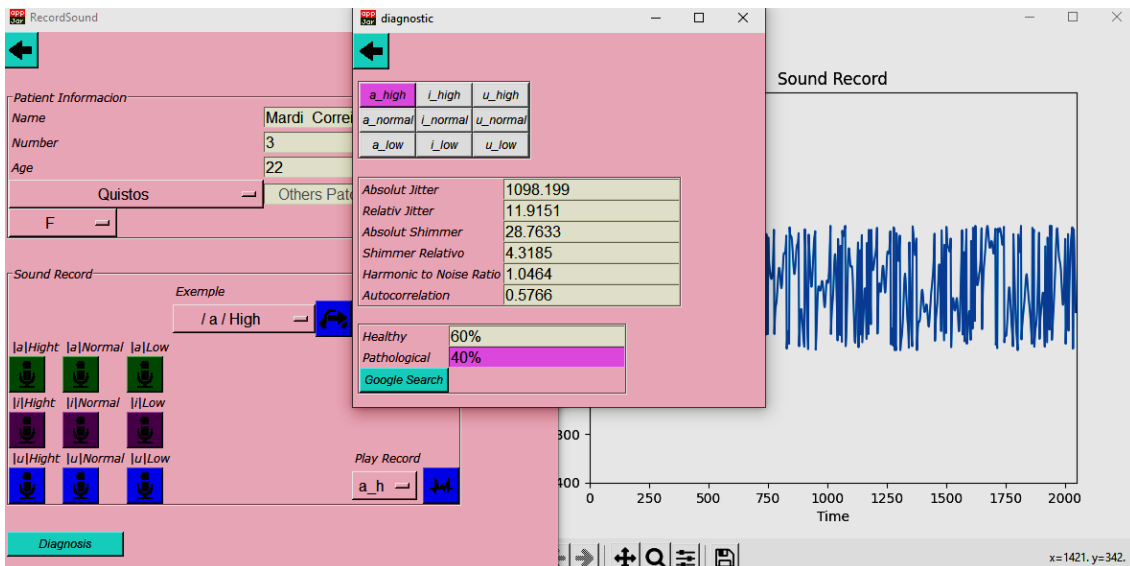


Figura 10- Janela *RecordSound* ao selecionar o botão *Diagnosis*.

## 5.2 Janela *Analysis*

Para ter acesso à janela “*Analysis*” (Figura 11), primeiro há a necessidade de preencher o *label* “*Information*” representado no *Front-end* do sistema, e caso o paciente se encontra na base de dados, a janela é acessada.

Ao clicar no botão com seta “*Play*” todas as gravações dos pacientes são carregadas e determina-se os parâmetros acústicos do sinal de cada gravação, como pode ser visível na Figura 12. Se o paciente apresentar alguma patologia, e essa informação já constar na base de dados, a mesma é apresentada na caixa “*Pathology*”.

Ao selecionar o botão ícone “*Clear*” todas as entradas são limpadas.

A seta “back” fechará a janela e retornará à janela principal do sistema.

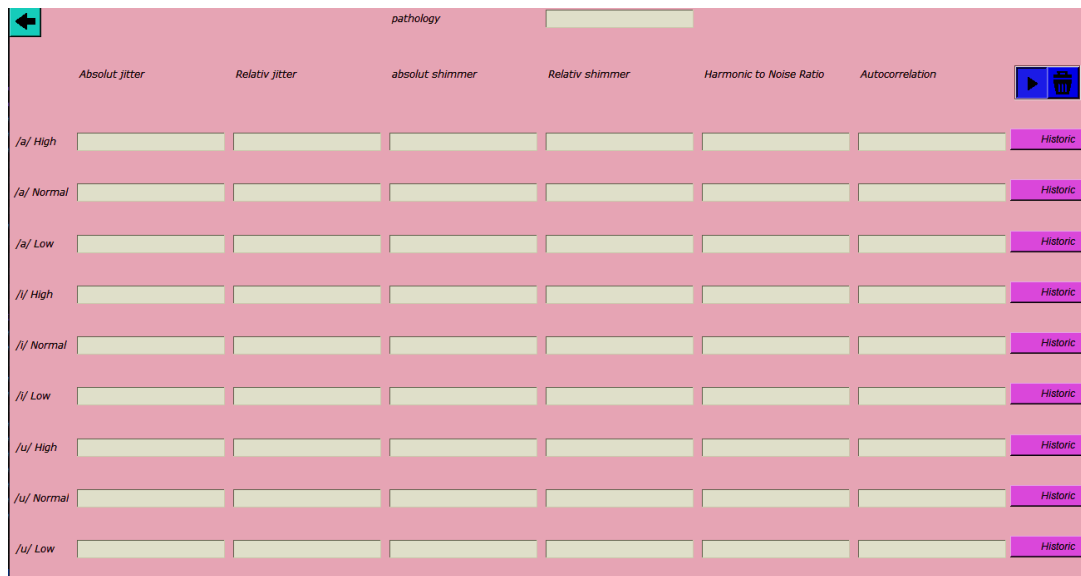


Figura 11- Janela *Analysis*.

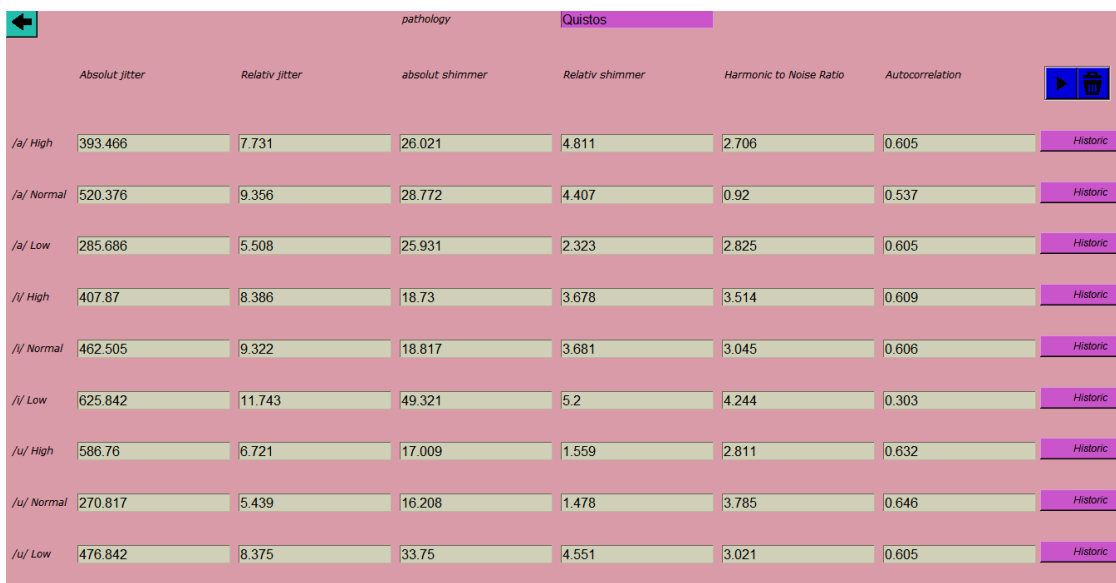


Figura 12- Janela *Analysis* ao clicar no botão com seta *Play*.

Ainda dentro desta mesma janela, “analysis” temos 9 botões referentes a históricos de cada sinal gravado, existentes na base de dados, onde ao clicar no botão “Show”, é visível a data em que o mesmo foi gravado e os seus respectivos parâmetros acústicos, (ver Figura 13).

Recording date	Absolute Jitter	Relative jitter	Absolute shimmer	Relative shimmer	Harmonic to Noise Ratio	Autocorrelation
2022-09-27	1094.9883	11.7177	31.4373	4.5236	2.4039	0.6315
2022-06-30	1890.8602	24.4748	95.4015	10.5737	15.2869	0.8837
<input type="button" value="Show"/>						

Figura 13- Janela *Históric*.

### 5.3 Janela *Diagnosis*

Além de poder fazer o diagnóstico dentro da janela “*RecordSound*”, isto é, após a gravação da voz, também é possível diagnosticar uma voz de um dado paciente existente na base de dados através da janela “*Diagnosis*”, representado na Figura 14.

O botão “*Google Search*” permite fazer pesquisas a partir do google.

Cada botão referente a um vogal e um tom que for selecionado, são determinados os seus respectivos parâmetros. O diagnóstico, para classificação do paciente, é apresentado em percentagem.

The screenshot shows the 'Diagnosis' window interface. At the top left is a back arrow button. Below it is a 3x3 grid of buttons for selecting vowels and tones: a\_high, i\_high, u\_high; a\_normal, i\_normal, u\_normal; a\_low, i\_low, u\_low. The 'a\_high' button is highlighted in pink. Below the grid is a 'Parameters' section with a list of acoustic parameters and their values: Absolute Jitter (1094.9883), Relativ Jitter (11.7177), Absolut Shimmer (31.4373), Relativ Shimmer (4.5236), Harmonic to Noise Ratio (2.4039), and Autocorrelation (0.6315). Below that is a 'Diagnóstico' section showing 'Healthy' at 60% and 'Pathological' at 40%. The 'Pathological' result is highlighted in pink. At the bottom left of this section is a 'Google Search' button.

Figura 14- Janela *Diagnosis*.

Embora não fosse primordial nesse trabalho, foi feita uma comparação entre o *Software Praat* e o algoritmo, de modo a verificar a exatidão do algoritmo na determinação dos parâmetros acústicos, pois os mesmos são usados como entrada da rede neuronal para

posterior classificação, esta comparação é feita com base, nos parâmetros, *Jitter* Absoluto, *Jitter* Relativo, *Shimmer* Absoluto, *Shimmer* Relativo, *Harmonic to Noise Ratio* (HNR), e a Autocorrelação. Como exemplo desta comparação foi utilizado um sinal de voz referente à vogal “a”, em tom alto, de um determinado sujeito. A unidade do *Jitter* é expressa em Microsegundos ( $\mu$ s), o *Jitter* em porcentagem (%), *ShdB* em decibel (dB), *Shim* em (%), e o HNR em dB.

Tabela 1: Valores dos parâmetros obtidos através *software Praat* e através do algoritmo correspondentes à vogal a.

	<i>vogal Tones</i>	<i>Jitter</i> Absoluto ( $\mu$ s)	<i>Jitter</i> Relativo (%)	<i>Shimmer</i> Absoluto (dB)	<i>Shimmer</i> Relativo (%)	<i>Harmonic to Noise Ratio</i> (dB)	Autocorrelação
<i>Praat</i>	<i>a_high</i>	13,793	0,206	0,189	2,165	26,453	0,9969
Algoritmo	<i>a_high</i>	24,855	0,371	0,234	2,619	24,917	0,9969

Apesar de não ser possível fazer uma conclusão fidedigna utilizando apenas um sujeito e uma vogal, observando a Tabela 1 é possível perceber que os resultados obtidos pelo algoritmo são próximos dos valores de referência. Esta análise não foi estendida uma vez que no trabalho de Fernandes et al., (2021), e no trabalho de Teixeira & Gonçalves, (2016) já foi realizada esta análise para os parâmetros *Jitter* e *Shimmer* respectivamente, e em Fernandes, (2018) para os parâmetros HNR, e Autocorrelação.

#### 5.4 Modelo de classificação

O Modelo MLP, utilizado para classificar o sujeito como sendo ou não patológico, apresentou como resultado do treino, a função perda (loss) de 0.0013, o MAE de 0.0013, o MSE de 0.174, e teve uma exatidão de 100%. E como resultado do teste, a função perda de 0.1132, MAE de 0.0161, MSE de 0.0117, e a exatidão de 98.6%.

A exatidão é medida no pós processamento da saída da rede, em que o resultado são valores entre 0 e 1. Foi utilizado como valor de referencia o valor 0.6, ou seja se o valor da saída da rede for superior a 0.6 é considerado 1, e significa que o paciente é patológico, e se for inferior a 0.6 ou seja 0 o paciente é saudável. Na figura 15 é possível observar o comportamento do modelo, assim como a exatidão obtida.

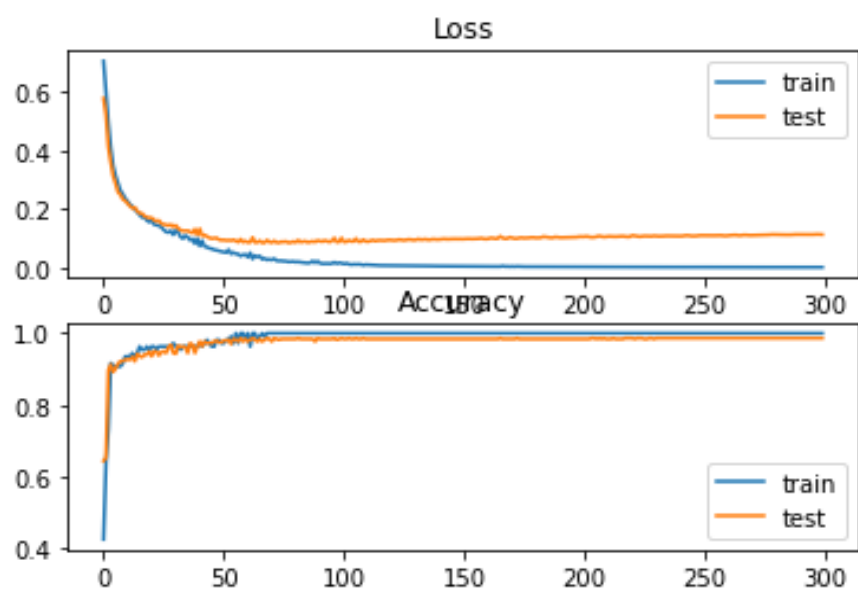


Figura 15-Exatidão do modelo MLP.



## 6. DISCUSSÃO DOS RESULTADOS

Nesta secção são discutidos os principais resultados, relativamente ao sistema desenvolvido, o porquê de ser desenvolvido em linguagem *Python* e outras considerações quanto ao mesmo.

O sistema foi desenvolvido numa linguagem *Python* devido à sua interação, versatilidade, simplicidade e acessibilidade. Uma das principais vantagens de usar a linguagem *Python* é a sua capacidade de desenvolver soluções simples, bem como, projetos mais complexos, com muitos dados, e utilizar inteligência artificial *machine learning*. Existe o benefício de a linguagem *Python* ser, hoje em dia, uma das mais utilizadas entre os desenvolvedores, tendo cada vez mais adeptos na sua utilização, apresentando vários *frameworks* e ferramentas que possibilitam criar interfaces gráficas.

Para a gravação dos sinais, recomenda-se a utilização de microfones profissionais, que captam de forma correta o som, e que a mesma seja feita num local apropriado sem barulhos e ruídos. Pois a forma como o sinal é captado, influenciará o resultado dos parâmetros acústicos, e conseqüentemente leva a erros de diagnóstico.

Para melhores resultados na determinação dos parâmetros acústicos, seria de grande importância, efetuar um pré-processamento do sinal, que eliminasse os períodos de pausa inicial e final, medindo os parâmetros, apenas, nas zonas onde ocorre fala.

Comparando com outras interfaces desenvolvidas anteriormente, por exemplo, a interface desenvolvida na linguagem MatLab por Fernandes et al., (2021), além de ser numa linguagem diferente, teve alterações quanto à sua representação visual, acrescentando mais opções e funcionalidades ao sistema. Por exemplo, para efetuar a gravação, são recolhidos mais dados ao ser introduzido o sujeito na base de dados, tal como género, idade, se é patológico ou não ter o acesso ao histórico de cada sinal gravado de um determinado sujeito, diagnosticar um sujeito logo após efetuar uma gravação, bem como, se o sujeito já se encontrar na base de dados acedendo a janelas diferentes, bem como outros botões e funcionalidades.



## 7. CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho foi desenvolvido um sistema de auxílio ao diagnóstico de patologias da fala.

A interface do utilizador é considerada a chave de entrada de um sistema computacional e torna se essencial a sua interação entre o homem e a máquina, contendo as funções necessárias para a sua operação. O utilizador vê a interface como o próprio sistema. Para ele a interface juntamente com os seus algoritmos, arquitetura e modelo de dados determina a qualidade do sistema. Por este motivo, muita das vezes, os utilizadores possuem dificuldade em aceitar um sistema, quando este não possui uma interface que seja fácil, agradável de usar e que garanta a eficiência do uso.

O sistema desenvolvido neste trabalho servirá como um método complementar no pré-diagnóstico de patologias da fala. Com a ajuda deste sistema-interface para o diagnóstico de patologias da voz, é possível um reconhecimento com rápido e fácil, e de forma não invasiva, tornando vantajoso numa primeira análise de rastreio, principalmente, em regiões com acesso limitado a profissionais clínicos.

Apesar de o sistema desenvolvido não ter sido avaliado por profissionais de saúde, afim de testar a interface quanto à usabilidade, incluindo fatores como, eficiência no desempenho e performance, eficácia na assertividade do sistema, quanto a satisfação do utilizador, facilidade de uso e a segurança na utilização. É reconhecível que esta interface apresente essas métricas de avaliação.

O modelo MLP, foi utilizado na classificação dos sujeitos entre dois grupos, saudáveis e patológicos, e atingiu uma taxa de exatidão de 98,6%, que, tendo em consideração o estado da arte, este valor encontra-se dentro dos valores para os melhores resultados. A possibilidade de inserção deste sistema, em hospitais e clínicas contribuirá para melhorias do mesmo, por meio de familiarização com os profissionais da área.

Como trabalhos futuros recomendo o desenvolvimento e o funcionamento da janela “*Multipathology*”. A classificação de uma entre várias patologias já se encontra em

investigação no grupo, pois é importante, não só classificar se o sujeito é patológico, mas também saber qual a patologia associada.

Uma melhoria ao sistema, seria a possibilidade de remover pacientes da base de dados.

Criar um painel de utilizadores entre terapeutas da fala e médicos de otorrino, afim de poder identificar pontos de melhoria.



## 8. REFERENCIAS BIBLIOGRAFICAS

- Almeida, N. (2010). Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor de Suporte. [Tese de Mestrado]. Universidade Federal do Rio Grande do Norte.
- Alves, N. (2016). Diagnóstico Inteligente de Patologias da Laringe. [Tese de Mestrado]. Instituto Politécnico de Bragança, Bragança-Portugal.
- American Cancer Society. (2021, January 19). About laryngeal and hypopharyngeal cancer. The American Cancer Society Medical and Editorial Content Team.
- Ammenwerth, E., Brender, J., Nykänen, P., Prokosch, H., Rigby, M., & Talmon, J. (2004). Visions and strategies to improve evaluation of health information systems Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. *International Journal of Medical Informatics*, 73(6), Pp., 479–491.
- Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., & König, J. (2003). Evaluation of health information systems—problems and challenges. *International Journal of Medical Informatics*, 71(2-3), Pp., 125–135.
- Aronson, A. E. (1990). *Clinical voice disorders*. New York: Thieme Medical Publishers.
- Awan, S. N., & Roy, N. (2009). Outcomes Measurement in Voice Disorders: Application of an Acoustic Index of Dysphonia Severity. *Journal of Speech, Language, and Hearing Research*, 52(2), 482–499.
- Bäckström, T. (2022, January 27). Jitter and shimmer. Aalto University Wiki.
- Baena, A. (2013, June 17). Disfonias: um problema de saúde pública. <https://www.douradosnews.com.br/saude-e-bem-estar/leia-disfonias-um-problema-de-saude-publica-de-ademir-baena/513264/>.

- Behlau, M., Azevedo, R., Pontes, P., & Brasil, O. (2001). Disfonias funcionais. Em: M.A. Behlau (Ed), *Voz: O Livro Do Especialista* (Pp. 247-294). Rio de Janeiro: Revinter.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Institute of Phonetic Sciences, University of Amsterdam, Proceedings 17* (1993), 97-110, 17, 97–110.
- Boersma, P. (2004). *Stemmen meten met Praat*. Universiteit van Amsterdam.
- Braga, A., & Oliveira, M. (2014). *Curso de Licenciatura em Letras-Língua Portuguesa*. EditAedi Belem-Pa, 11.
- Buder, E. (2000). Acústic Analysis of Voice Quality: A Tabulation of Algorithms 1992-1990. *Voice Quality Measurement*, 119–244.
- Caçador, M. C. R. C. (2019). *Voz e postura*. [Tese de mestrado em Cirurgia e Morfologia Humana]. Universidade Nova de Lisboa- Faculdade de ciências médicas.
- Camargo, A. C. (n.d.). *Laringe*. Cancer Center.
- César, A. (n.d.). *Edema de Reinke*. Sociedade Brasileira de Cirurgia de Cabeça e Pescoso. Retrieved July 12, 2022, from <http://dralexandrecesar.com.br/edema-de-reinke/>
- Charu, C. (2018). *Neural networks and deep learning*. Springer.
- Clínica SanVicente. (2021). *Alterações da Voz: Disfonia*. Retrieved 12/10/2022, from <https://www.clinicasanvicente.es/pt-pt/servicios/alteracoes-da-voz-disfonia/>
- Clínica Universidad da Navarra. (2022). *Disfonia Hipofuncional*. Retrieved 12/07/2022, from <https://www.cun.es/diccionario-medico/terminos/disfonia-hipofuncional>

- Cordeiro, H. (2016). Reconhecimento de Patologias da Voz usando Técnicas de Processamento da Fala. [Tese de Doutorado]. Faculdade de ciências e tecnologia, universidade nova de Lisboa.
- Cordeiro, H., & Meneses, C. (2019). Spectral features of healthy and pathological voices: Results comparison between two databases. Iberian Conference on Information Systems and Technologies, CISTI, 2019-June.
- Costa, W. C. de A., Costa, S. L. do N. C., de Assis, F. M., & Neto, B. G. A. (2013). Classificação de sinais de vozes saudáveis e patológicas por meio da combinação entre medidas da análise dinâmica não linear e codificação preditiva linear. *Revista Brasileira de Engenharia Biomedica*, 29(1), 3–14.
- Devino, B. (2021, October 15). Python - Uma Introdução à Linguagem. Alura. Retrieved 12/10/2022, from <https://www.alura.com.br/artigos/python-uma-introducao-a-linguagem>
- Dibazar, A. A., Narayanan, S., & Berger, T. W. (2002). Feature analysis for automatic detection of pathological speech. Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings of the Second Joint EMBS/BMES Conference Houston, TX, USA., 1, 182–183.
- Dieguez, F., Barbosa, M., Almeida, A., RaMaciel, N., & Souza, V. (2010). Granuloma Laríngeo: relato de caso (Vol. 5). *Revista Científica da FMC*.
- Ebit. (2022, May 23). O que é uma interface gráfica de utilizador. Ebit.
- Estibero, H., & Trindade, C. (n.d.). Tumores da Laringe e Faringe. Instituto Português de Oncologia de Lisboa Francisco Gentil, E.P.E.
- Fernandes J. (2018). Determinação da Autocorrelação, HNR e NHR para Análise Acústica Vocal. [Tese de Mestrado]. Instituto Politécnico de Bragança, Bragança- Portugal.
- Fernandes, J., Borghi, P. H., Freitas, D. S., & Teixeira, J. P. (2021). Optimization of Glottal Onset Peak Detection Algorithm for Accurate

Jitter Measurement. In *Communications in Computer and Information Science*: Vol. 1488 CCIS.

Fernandes, J. F. T., Freitas, D., & Teixeira, J. P. (2021). Voice Pathologies: The Most Common Features and Classification Tools. *Iberian Conference on Information Systems and Technologies, CISTI*.

Fernandes, J., Silva, L., Teixeira, F., Guedes, V., Santos, J., & Teixeira, J. P. (2019). Parameters for Vocal Acoustic Analysis - Cured Database. *Procedia Computer Science*, 164, 654–661.

Fezari, M., Amara, F., & El-Emary, I. M. M. (2014). Acoustic Analysis for Detection of Voice Disorders Using Adaptive Features and Classifiers. *Proceedings of the 2014 International Conference on Circuits, Systems and Control* ISBN: 978-1-61804-216-3112.

Forero M., L. A., Kohler, M., Vellasco, M. M. B. R., & Cataldo, E. (2016). Analysis and Classification of Voice Pathologies Using Glottal Signal Parameters. *Journal of Voice*, 30(5), 549–556.

Fry, D. B. (1976). *Acoustic Phonetics: a course of basic readings*. Cambridge: Cambridge University Press.

Gama, A. C. C. (1997). *Análise Acústica de Voz: Estudo comparativo entre mulheres sem queixa vocal e com disfonia [Tese de Mestrado]*. Escola Paulista de Medicina.

Gonçalves, A. A. (2015). *Patologias da Laringe com Análise Acústica Vocal: Tese Mestrado*. Instituto Politécnico de Bragança, Bragança- Portugal.

Gonçalves, E. (2001). *A interação com o usuário na validação do software oficina de relatórios. [Tese de Mestrado]*. Universidade Federal de Santa Catarina.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4), 383–406.

Guimarães, I. (2007). *A ciência e a arte da voz humana (Edição da ESSA)*.

- Helena, R., Martins, G., Dias, N. H., Segundo, C., Soares, P., Cristina, A., Gramuglia, J., & Martins, H. G. (2019). Treatment of Laryngeal Granulomas. *Int Arch Otorhinolaryngol*, 23, 322–324.
- INCA. (2022, April 25). Tipos de cancer. Instituto Nacional de Cancer. Retrieved 12/07/2022, from <https://www.inca.gov.br/tipos-de-cancer/cancer-de-laringe>
- Koufman, J. A., Amin, M. R., & Panetti, M. (2000). Prevalence of reflux in 113 consecutive patients with laryngeal and voice disorders. *Otolaryngol Head Neck Surg* 2000; 123:385-88.
- Koufman, J., Postma, G., Cummins, M., & Blalock, P. (2000). Vocal fold paresis. *Otolaryngol Head Neck Surg*. 2000; 122: 537-41.
- Kulpa, Cí. C. (2009). A contribuição de um modelo de cores na usabilidade das interfaces computacionais para usuários de baixa visão [Tese de Mestrado]. Universidade Federal do Rio- Rio Grande do Sul.
- Lopes, J. (2008). Ambiente de análise robusta dos principais parâmetros qualitativos da voz. [Tese de Mestrado]. Faculdade de engenharia da universidade do Porto.
- Martínez, C. E., & Rufiner, H. L. (2000). Acoustic analysis of speech for detection of laryngeal pathologies. *Annual International Conference of the IEEE Engineering in Medicine and Biology -Proceedings of the 22"d Annual EMBS International Conference, July 23-28,2000, Chicago IL*, 3, 2369–2372.
- Martins, R., & Dias, N. (n.d.). Complicações das vias aéreas relacionadas à intubação endotraqueal. In Botucatu. Faculdade de Medicina de Botucatu - Unesp.
- Martins, R., Domingues, M., Fabro, A., Dias, N., & Santana, M. (2009). Edema de Reinke: estudo da imunexpressão da fibronectina, da laminina e do colágeno IV em 60 casos por meio de técnicas imunoistoquímicas. *SCIELO*, 75.

- Michael P Underbrink, M. (2019, October 26). Vocal cord paresis (paralysis): symptoms, causes, and treatment options. Houston ent and allergy blog.
- Muhammad, G., Altuwaijri, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., & Al-Nasheri, A. (2016). Automatic voice pathology detection and classification using vocal tract area irregularity. *Biocybernetics and Biomedical Engineering*, 36(2), 309–317.
- Murphy, P., & Akande, O. (2005). Cepstrum-Based Estimation of the Harmonics-tonoise Ratio for Synthesized and Human Voice Signals. In *Nonlinear Analyses and Algorithms for Speech Processing*. Barcelona, LNAI 3817, Springer.
- Neves, A., Sarmiento, R., & Vitorino, D. (n.d.). Design de interação: além do alcance individual. In *Book: Design e Inovação Social* (Pp.239-256).
- Otorrinolaringologia. (n.d.). Fibroma Laríngeo. YBSITE. Retrieved October 12, 2022, from <https://pt.ybsite.org/disease/h-7744.html>
- Panek, D., Skalski, A., Gajda, J., & Tadeusiewicz, R. (2015). Acoustic Analysis Assessment in Speech Pathology Detection. *International Journal of Applied Mathematics and Computer Science*, 25(3), 631–643.
- Pereira, C. (1955). Quistos da laringe. In *BJORL- Brazilian journal of otorhinolaryngology* (5th ed., Vol. 23, pp. 184–200).
- Prates, R., & Barbosa, S. (2003). Avaliação de Interfaces de Usuário-Conceitos e Métodos. *Jornada de Atualização Em Informática Do Congresso Da Sociedade Brasileira*.
- Preece, J., Rogers, Y., & Sharp, H. (2003). *Design de interação: além da interação homem-computador*. Porto Alegre: Bookman; 2003.
- Ribeiro, R., Vilarinho, S., Ferreira, J. P., & Sousa, C. A. (2011). Mucopiocele supraglótico - causa de paragemcardio-respiratória. *Revista Portuguesa de Otorrinolaringologia e Cirurgia Cérvico-Facial*, 49.
- Robbins, S. L., & Cotran, R. S. (2005). *Robbins & Cotran Patologia - Bases Patológicas das Doenças*. Rio de Janeiro, Brazil-. Elsevier Editora Ltda.

- Sasaki, C. T. (2020, January). Distúrbios de ouvido, nariz e garganta / Distúrbios da Laringe / Disfonia espasmódica. Manual MSD -Versão Saúde Para Família.
- Scarpel, R. D., & Fonseca, M. D. L. (2014). Parâmetros acústicos de vozes de mulheres na pós-menopausa. *Revista Brasileira de Geriatria e Gerontologia*, 17(4), 741–750.
- Simões Marta, P. (2017). Disfonia Psicogénica. [Tese de Mestrado]. Faculdade de Medicina Lisboa- Portugal.
- Stojanovic, S., & Belic, B. (2013). Laryngeal Manifestations of Rheumatoid Arthritis. In *Innovative Rheumatology*. InTech.
- Teixeira, J. P., (1995). Modelização paramétrica de sinais para sistemas de conversao texto-fala. [Tese de Mestrado]. Universidade do Porto.
- Teixeira, J. P., Alves, N., & Fernandes, P. O. (2020). Vocal acoustic analysis: Ann versus SVM in classification of dysphonic voices and vocal cords paralysis. *International Journal of E-Health and Medical Communications*, 11(1), 37–51.
- Teixeira, J. P., & Fernandes, A. P. (2013). Didactic speech synthesizer: Acoustic module. In *proceedings of 6th International Joint Conference on Biomedical Engineering Systems and Technologies*, 356–359.
- Teixeira, J. P., Fernandes, J., Teixeira, F., & Fernandes, P. O. (2018). Acoustic analysis of chronic laryngitis statistical analysis of sustained speech parameters. *BIOSIGNALS 2018 - 11th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 11th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2018*, 4, 168–175.
- Teixeira, J. P., & Fernandes, P. O. (2015). Acoustic Analysis of Vocal Dysphonia. *Procedia Computer Science* 64 (2015) 466 – 473, 64, 466–473.

- Teixeira, J. P., Barbosa Ferreira, D., & Moreira Carneiro, S. (2011). Análise acústica vocal -determinação do Jitter e Shimmer para diagnósticos de patologia da fala. Atas do VI Congresso Luso Moçambicano de Engenharia; Maputo, Moçambique, ISBN: 978-972-8826-24-6
- Teixeira, J. P., & Gonçalves, A. (2014). Accuracy of Jitter and Shimmer Measurements. *Procedia Technology*, 16, 1190–1199.
- Teixeira, J. P., & Gonçalves, A. (2016). Algorithm for Jitter and Shimmer Measurement in Pathologic Voices. *Procedia Computer Science*, 100, 271–279.
- Trinh, N. H. (2021). Semi-Supervised Learning with Generative Models for Pathological Speech Classification. master's science desertion. Dublin City University.
- Vale, S. (2011). “Emissão vocal. Uma visão física, fisiológica e psicológica das pregas vocais” [Tese de Mestrado]. Universidade Católica Portuguesa-Porto.