



UNIVERSIDADE D
COIMBRA

Ana Rita de Castro Carreiró e Sá

**TÉCNICAS DE PROCESSAMENTO DIGITAL
ÁUDIO PARA CARACTERIZAÇÃO DE
POPULAÇÃO DE CONTROLO COM VISTA À
ANÁLISE DOS EFEITOS DE TIROIDECTOMIAS**

Dissertação no âmbito do Mestrado Integrado em Engenharia Biomédica orientada pelo Professor Doutor João Manuel Rendeiro Cardoso e pelo Doutor Luís Filipe Silva e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Outubro de 2020



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Técnicas de processamento digital áudio para caracterização de população de controlo com vista à análise dos efeitos de tiroidectomias

Ana Rita de Castro Carreiró e Sá

Dissertação apresentada à
Universidade de Coimbra para obtenção do grau de
Mestre em Engenharia Biomédica

Orientadores:
Professor Doutor João Manuel Rendeiro Cardoso
Doutor Luís Filipe Silva

Coimbra, 2021

Este trabalho foi desenvolvido em colaboração com:



LIBPhys-UC



Esta cópia da tese é fornecida na condição de que quem a consulta reconhece que os direitos de autor são pertença do autor da tese e que nenhuma citação ou informação obtida a partir dela pode ser publicada sem a referência apropriada.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.



Agradecimentos

Em primeiro lugar, gostaria de agradecer ao Professor João Cardoso, pela oportunidade, pelo apoio prestado e tempo disponibilizado. Além disso, agradeço à restante equipa do LIBPhys por me ter integrado desde o primeiro dia. Gostaria também de agradecer ao Doutor Luís Filipe Silva, pela oportunidade de trabalhar com a sua equipa, embora devido às circunstâncias não tenha sido possível concretizar.

A todos os meus amigos que me acompanharam ao longo destes 5 anos, um enorme obrigado pelos momentos e memórias partilhados e que com certeza guardarei com carinho. Em especial, ao Bruno, à Inês e à Patrícia, por estarem sempre presentes, por todos os sorrisos, lágrimas e conselhos que enriqueceram o meu percurso e que levo para a vida.

À minha família, em particular ao meu pai, irmãos, prima, tios e avós, agradeço todo o carinho, incentivo e apoio incondicional que fizeram de mim a pessoa que sou hoje e que tornaram tudo isto possível.

Por último, um agradecimento a Coimbra, por ser a minha segunda casa, por todas as amizades e crescimento que me proporcionou e que recordarei com a imensa Saudade que lhe é característica!

Resumo

A tireoidectomia é um procedimento comum tanto para remover nódulos benignos como malignos que envolvam a glândula tiroide. Geralmente apresenta uma baixa taxa de complicações, todavia pode resultar no comprometimento vocal derivado de traumatismos cirúrgicos que condicionem a integridade das estruturas responsáveis pela fonação. Após a intervenção cirúrgica, as queixas de disфония são regulares e podem persistir por cerca de três a seis meses, sendo os pacientes encaminhados para tratamentos terapêuticos. A análise e diagnóstico dos respectivos tratamentos, baseia-se sobretudo em avaliações áudio-precetivas em que a subjetividade está inerente. As técnicas de avaliação subjetivas geram falta de consenso entre os profissionais, pelo que é importante procurar soluções objetivas, através das quais a qualidade vocal seja quantificada por parâmetros acústicos.

No âmbito desta dissertação é proposta uma técnica de avaliação objetiva através do processamento digital de sinais acústicos. Para tal, é implementado, em MATLAB, um conjunto de algoritmos de processamento de som durante um protocolo de teste fonético, que permite quantificar alterações nos padrões de vibração das pregas vocais. Os principais parâmetros que identificam tais alterações são: Frequência Fundamental, *jitter*, RAP, PPQ5, *shimmer*, APQ3, APQ5, APQ11 e *Harmonic-to-Noise Ratio*.

O estudo inclui uma amostra com 15 sujeitos do género feminino e 15 do género masculino, assim como a avaliação da sua qualidade vocal com recurso à produção das vogais sustentadas /a/ e /i/. A partir dos resultados obtidos é estabelecida uma comparação estatística entre os parâmetros obtidos através do *software* standard nesta caracterização (Praat) e os resultados do algoritmo desenvolvido. Além disso, são investigadas as características específicas que possam ter influência sobre a qualidade vocal de cada sujeito. Através da análise da distribuição métrica dos respectivos parâmetros de ambas as plataformas de análise, verificou-se que as diferenças foram pouco significativas na maior parte deles. A validação experimental

mostra que este método exibe resultados fiáveis, podendo ser aplicado a situações reais em ambiente clínico.

Com o presente trabalho pretendeu-se assim, dotar futuros desenvolvimentos no domínio das ferramentas de análise necessárias para a correta caracterização de uma população no estudo do efeito das tiroidectomias.

Palavras-chave: Tiroidectomia, Processamento digital de sinais, Análise Acústica, Frequência Fundamental, *jitter*, *shimmer*, *Harmonic-to-Noise Ratio*

Abstract

Thyroidectomy is a common procedure to remove both benign and malignant nodules that surround the thyroid gland. Even though it presents a low rate of complications, surgical trauma during the operation can result in vocal impairment that compromises the integrity of the structures responsible for phonation. After the surgical intervention, complaints of dysphonia are regular and may persist for about three to six months. The analysis and diagnosis of the respective treatments are mainly based on audio-preceptive evaluations. Due to the subjective nature of these techniques, there is a lack of consensus among professionals in their analysis. Therefore, it becomes important to develop unbiased solutions, through which the vocal quality can be quantified by acoustic parameters.

In this dissertation, an objective evaluation technique through the digital processing of acoustic signals is presented. For this purpose, a set of sound processing algorithms are implemented during a phonetic test protocol, using MATLAB, which allows for the quantification of changes in vocal fold vibration patterns. The main parameters that are used to identify such changes are: Fundamental Frequency, *jitter*, RAP, PPQ5, *shimmer*, APQ3, APQ5, APQ11 and *Harmonic-to-Noise Ratio*.

The study includes a sample of 15 female and 15 male subjects and the evaluation of their vocal quality through the analysis of sustained vowels /a/ and /i/. From the results obtained, a statistical comparison is established between parameters from the standard software in this characterization (Praat) and the results of the developed algorithm. Besides, specific characteristics that may influence the vocal quality of each subject are examined. Through the analysis of the metric distribution of the parameters of both analysis platforms, it was found that the differences were negligible in most of them. The experimental validation shows that the developed method can achieve reliable results, therefore showing potential for application in the clinical environment.

Overall, this work provides a step forward in the development of methods for correct characterization of the effects of thyroidectomies in a given population.

Keywords: Thyroidectomy, Digital Signal Processing, Acoustic Analysis, Fundamental Frequency, *jitter*, *shimmer*, *Harmonic-to-Noise Ratio*

Lista de Siglas e Acrónimos

F_0 Frequência Fundamental.

APQ11 *eleven-point Amplitude Perturbation Quotient.*

APQ3 *three-point Amplitude Perturbation Quotient.*

APQ5 *five-point Amplitude Perturbation Quotient.*

CAPE-V *Consensus Auditory-Perceptual Evaluation of Voice.*

EGG Eletroglotografia.

ELS Sociedade Europeia de Laringologia.

EMG Eletromiografia.

GRBAS *Grade, Roughness, Breathiness, Asthenia, Strain.*

GUI *Graphical User Interface.*

HNR *Harmonic-to-Noise Ratio.*

NLR Nervo Laríngeo Recorrente.

NLS Nervo Laríngeo Superior.

PPQ5 *five-point Period Perturbation Quotient.*

RAP *Relative Average Perturbation.*

RENLS Ramo Externo do Nervo Laríngeo Superior.

V-RQOL *Voice-Related Quality of Life Measure.*

VHI *Voice Handicap Index.*

Lista de Figuras

2.1	Aparelho fonador humano.	5
2.2	Representação esquemática das estruturas anatómicas da laringe. . .	7
2.3	Representação esquemática das 4 fases do impulso glótico.	8
2.4	Representação do <i>jitter</i> e <i>shimmer</i> na produção da vogal /a/ sustentada. 20	
3.1	Microfone <i>Zoom H4n Handy Recorder</i>	27
3.2	Picos determinados através do algoritmo <i>peakdet</i> com diferentes thresholds	29
3.3	Produção de 2 vogais sustentadas por sujeitos do género feminino . .	30
3.4	Produção de 2 vogais sustentadas por sujeitos do género masculino .	31
3.5	Interface de Análise Acústica.	33
4.1	Deteção de Picos de um sinal acústico	36
4.2	<i>Boxplots</i> dos parâmetros acústicos principais da vogal /a/ na amostra feminina.	38
4.3	<i>Boxplots</i> das medidas de <i>jitter</i> e de <i>shimmer</i> da vogal /a/ na amostra feminina.	38
4.4	<i>Boxplots</i> dos parâmetros acústicos principais da vogal /i/ na amostra feminina.	39
4.5	<i>Boxplots</i> das medidas de <i>jitter</i> e de <i>shimmer</i> da vogal /i/ na amostra feminina.	40
4.6	<i>Boxplots</i> dos parâmetros acústicos principais da vogal /a/ na amostra masculina.	41
4.7	<i>Boxplots</i> das medidas de <i>jitter</i> e de <i>shimmer</i> da vogal /a/ na amostra masculina.	41
4.8	<i>Boxplots</i> dos parâmetros acústicos principais da vogal /i/ na amostra masculina.	42
4.9	<i>Boxplots</i> das medidas de <i>jitter</i> e de <i>shimmer</i> da vogal /i/ na amostra masculina.	43

4.10	Diagramas do desvio dos parâmetros acústicos do sujeito F_13 relativamente à mediana	46
4.11	Diagramas do desvio dos parâmetros acústicos do sujeito M_15 relativamente à mediana	47

Lista de Tabelas

3.1	Caracterização da amostra Feminina.	24
3.2	Caracterização da amostra Masculina.	24
4.1	Erro relativo dos picos do MATLAB em comparação com os picos do Praat.	36
4.2	Comparação entre as 4 medidas de <i>jitter</i> do MATLAB e do Praat. . .	36
4.3	Medianas dos <i>boxplots</i> das vogais /a/ e /i/ para as amostras feminina e masculina.	44

Índice

Lista de Siglas e Acrónimos	xiii
Lista de Figuras	xv
Lista de Tabelas	xvii
1 Introdução	1
1.1 Motivação do estudo	2
1.2 Objetivos	3
1.3 Organização da dissertação	3
2 Estado de Arte	5
2.1 Mecanismos de Produção da Voz	5
2.1.1 Anatomia da Laringe e das Pregas Vocais	6
2.1.2 Fisiologia da Fonação	8
2.1.3 Voz Normal vs. Disfonia	9
2.1.4 Influência da Tireoidectomia na Produção de Voz	10
2.2 Técnicas de Avaliação da Voz	11
2.2.1 Programas de Análise Acústica da Voz	14
2.2.1.1 Praat	15
2.2.2 Parâmetros Acústicos	15
2.2.2.1 Frequência Fundamental	16
2.2.2.2 <i>Jitter</i>	16
2.2.2.3 <i>Shimmer</i>	18
2.2.2.4 <i>Harmonic-to-Noise Ratio</i> (HNR)	20
3 Metodologia	23
3.1 Amostra	23
3.2 Protocolo de Avaliação da Qualidade Vocal	25

3.3	Gravação	26
3.3.1	Equipamento de Gravação	26
3.4	Métodos de Extração dos Parâmetros	28
3.4.1	Extração dos parâmetros no Praat	28
3.4.2	Algoritmo desenvolvido no MATLAB	29
3.4.3	Interface Gráfica	32
4	Resultados	35
4.1	Precisão do Algoritmo na Determinação de Picos	35
4.2	Análise de Resultados	37
4.2.1	Vogais sustentadas nos sujeitos do género Feminino	37
4.2.2	Vogais sustentadas nos sujeitos do género Masculino	40
4.2.3	Análise Geral dos Resultados da Amostra	43
4.2.4	Análise dos Sujeitos com Restrições Vocais	45
5	Conclusão e Trabalhos Futuros	49
	Bibliografia	53

Introdução

A voz é um fenómeno extraordinariamente complexo que depende de processos fisiológicos e físicos através dos quais a vibração das pregas vocais é transformada em fala [1]. Devido à complexidade envolvida na coordenação entre os diferentes sistemas, a produção da voz é um processo altamente variável entre sujeitos e até dentro do mesmo sujeito [2]. O estudo da fisiologia deste processo é essencial para desenvolver modelos matemáticos e físicos eficientes na construção de sistemas de reconhecimento e codificação de voz. O desempenho vocal tem despertado um interesse crescente, desde o século XIX, e conseqüentemente têm surgido inúmeras pesquisas multidisciplinares relativas à avaliação e terapia da voz, incluindo vários estudos centrados nos aspetos qualitativos e quantitativos da mesma [3]. O conceito de qualidade vocal não deve ser restringindo a uma única técnica de avaliação, visto que individualmente nenhum método possibilita a caracterização abrangente de todos os aspetos da função vocal [4]. Assim, é fundamental recorrer a técnicas de avaliação que permitam um conhecimento amplo e adequado da qualidade vocal, de entre as quais se destacam a avaliação áudio-percetiva, quer pelo profissional de saúde quer pelo próprio sujeito, a videoestroboscopia, a avaliação aerodinâmica e a análise acústica [5].

A análise acústica é o tema central desta dissertação e consiste numa técnica não invasiva fundamentada pelo processamento digital do som, que oferece um diagnóstico objetivo das perturbações da voz e facilita a sua deteção precoce [6]. Esta técnica baseia-se na linearidade da teoria fonte-filtro para a produção de vogais, na qual a fonte é a vibração laríngea e o filtro é o trato vocal [7], quantificando a qualidade vocal do indivíduo através dos diferentes parâmetros acústicos que compõem o sinal. Atualmente, os parâmetros acústicos são frequentemente determinados com o recurso ao *software* Praat, amplamente utilizado em otorrinolaringologia no âmbito de diagnósticos vocais. Os parâmetros mais comuns na avaliação da integridade da voz são a Frequência Fundamental (F_0), o *Jitter*, o *Shimmer*, e a relação *Harmonic-*

to-Noise Ratio (HNR) [8]. O *jitter* é afetado principalmente pela falta de controlo da vibração das pregas vocais, o *shimmer* muda com a redução da resistência glótica e com lesões em massa nas pregas vocais e o HNR quantifica o ruído aditivo no sinal de voz [9].

Não existem definições concretas de vozes normais ou patológicas. Por norma, considera-se uma voz disfónica aquela em que a qualidade, o tom, a altura ou a flexibilidade são interpretadas como sendo desagradáveis ou inadequadas à idade e género do orador [10]. A qualidade vocal de um indivíduo pode ser afetada por uma multiplicidade de fatores com origens orgânicas ou funcionais, resultando em diferentes tipos de disfonias [11]. As cirurgias ao nível do pescoço constituem um risco para a qualidade vocal dos indivíduos e podem resultar em disfonias de origem orgânica. No que diz respeito à cirurgia da tiroide, a disfunção vocal é uma das complicações mais comuns, sendo que ocorre em cerca de 37-87% dos pacientes submetidos à mesma. Os sintomas de disfonia podem resultar de traumatismos no Nervo Laríngeo Recorrente (NLR) e no Ramo Externo do Nervo Laríngeo Superior (RENLS), responsáveis pelo controlo vocal, ou de lesões nos músculos circundantes [12]. Assim, a preservação da integridade de todas as estruturas da laringe, nomeadamente do NLR e do RENLS é fundamental para garantir um bom desempenho do processo de fonação e deve ser tida em consideração no decorrer de uma cirurgia à tiroide [13]. Contudo, nem sempre é possível evitar este tipo de lesões, pelo que é importante monitorizar a qualidade vocal antes e depois de uma tiroidectomia, aferir sobre o impacto da mesma e, se necessário, encaminhar os indivíduos para processos terapêuticos de reabilitação.

1.1 Motivação do estudo

A incidência das alterações vocais após uma tiroidectomia e a variabilidade de sintomas de que os pacientes se queixam, reforça a necessidade da realização de triagens vocais antes e depois da cirurgia e a importância de uma análise abrangente. Compreender o tipo de implicações funcionais que contribuem para a disfunção vocal é um pré-requisito na identificação antecipada das mesmas e fornece orientação na seleção entre as diferentes abordagens terapêuticas.

Na prática clínica a técnica mais comum de avaliação da qualidade da vocal, após uma tiroidectomia, é a avaliação perceptiva. No entanto, este procedimento é subjetivo e pode levar a resultados distintos dependendo do profissional envolvido. Atualmente, os especialistas procuram recursos objetivos que permitam identificar os pa-

cientes que requerem um acompanhamento mais atento no processo de recuperação. A terapêutica depende da abordagem adotada e é frequentemente mais eficaz quando os parâmetros afetados são diagnosticados objetivamente. De modo a complementar a avaliação áudio-percetiva, este estudo apresenta uma ferramenta de análise objetiva, capaz de efetuar a medição de vários parâmetros acústicos e a documentação das mudanças que ocorrem na voz.

1.2 Objetivos

O objetivo inicial do projeto consistia na implementação de técnicas de processamento digital de som para análise acústica de disfunções provocadas em cirurgias da tireoide. Além disso, também se pretendia monitorizar a eficácia dos procedimentos terapêuticos, procedendo à recolha de dados de pacientes em diferentes graus de recuperação após uma tiroidectomia. Face à situação atual da pandemia COVID-19, não foi possível ir ao CHUC recolher os sinais acústicos dos pacientes e o objetivo inicialmente proposto ficou condicionado e por cumprir. Por esta razão, foram efetuadas adaptações relativamente à amostra em análise, da qual passaram a fazer parte todos os sujeitos que se disponibilizaram a participar no estudo, em substituição dos pacientes submetidos a tiroidectomias.

Na presente dissertação, mantêm-se o foco na elaboração de um algoritmo em MATLAB, capaz de determinar os parâmetros acústicos que caracterizam os movimentos das pregas vocais, Frequência Fundamental, *jitter*, RAP, PPQ5, *shimmer*, APQ3, APQ5, APQ11 e *Harmonic-to-Noise Ratio*, sendo estes quantificados em diferentes comportamentos de fala (vogais /a/ e /i/ sustentadas). Para verificar a apropriação do algoritmo desenvolvido, realiza-se uma análise comparativa entre os resultados determinados pelo algoritmo e os obtidos com o *software* Praat. Por fim, com base nas características da amostra são averiguadas as possíveis causas que podem estar na origem dos parâmetros com valores atípicos.

1.3 Organização da dissertação

Capítulo 2 Apresenta um enquadramento teórico das características anatómicas e fisiológicas das pregas vocais, essencial para compreender os processos de produção da voz, bem como os conceitos subjacentes à caracterização de vozes normais e patológicas e as técnicas de avaliação vocal usadas em contexto clínico.

Capítulo 3 Inclui a metodologia do projeto, nomeadamente a caracterização da amostra analisada, o protocolo e procedimento de gravação e a descrição detalhada do algoritmo desenvolvido em MATLAB.

Capítulo 4 Apresenta os resultados da análise comparativa entre o algoritmo do MATLAB e o Praat e uma estabelece uma relação entre as características da população e os parâmetros discrepantes da mesma.

Capítulo 5 Expõe as principais conclusões obtidas no estudo, as limitações que o condicionam e apresenta sugestões para trabalhos futuros.

2

Estado de Arte

Este capítulo apresenta um enquadramento teórico fundamental para compreender o processo de produção da voz, bem como os fatores que a podem condicionar. Além disso, são apresentadas as diferentes técnicas usadas na prática clínica para caracterizar a voz, destacando os vários aspetos da avaliação acústica. Por fim, é elaborada uma descrição detalhada dos parâmetros acústicos que vão ser analisados no presente estudo.

2.1 Mecanismos de Produção da Voz

A produção de voz caracteriza-se por um complexo mecanismo de coordenação neuromuscular entre o sistema neurológico e os diversos componentes do aparelho fonador [14]. Os principais componentes fisiológicos responsáveis pela produção de voz, são o sistema respiratório, localizado abaixo da laringe, o sistema fonatório, composto pela laringe, e o sistema de articulação e ressonância, localizado acima da laringe, como ilustrado na Figura 2.1 [15, 16].

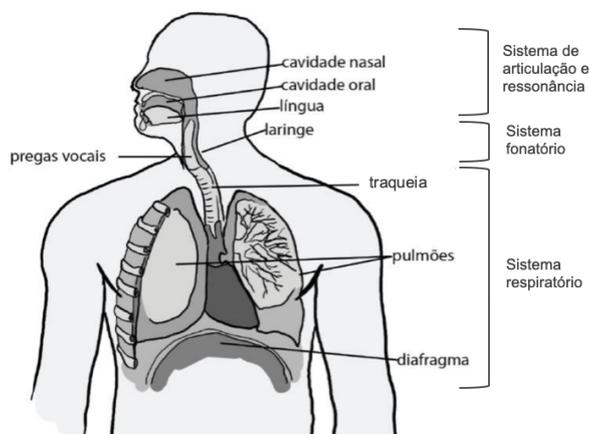


Figura 2.1: Aparelho fonador humano, dividido nos sistemas de articulação e ressonância, fonatório e respiratório, responsáveis pela fonação. Adaptado de [17]

De acordo com Lieberman *et al.* [16], o sistema respiratório, constituído pelos pulmões e pela musculatura respiratória, é responsável pelo fluxo do ar que é expelido dos pulmões por ação do diafragma. Por sua vez, o sistema fonatório converte esse fluxo, em séries quase periódicas de sopros de ar através da abertura e fecho da laringe. A alteração das configurações das cavidades de ressonância (faringe, cavidades orais e cavidades nasais) favorecem a passagem de algumas frequências e a atenuação de outras, por meio de movimentos coordenados da língua, lábios, mandíbula e restantes estruturas de articulação do trato vocal. Durante a produção da fala, os movimentos da laringe dão origem à fonte de energia acústica da fonação e os sistemas de articulação e ressonância modelam e filtram os sons provenientes dessa mesma fonte, originando a voz [18, 19].

2.1.1 Anatomia da Laringe e das Pregas Vocais

A laringe é um pequeno órgão cervical localizado no topo da traqueia, cuja principal função é proteger as vias áreas inferiores, evitando que materiais estranhos entrem nos pulmões [20]. É também responsável pela deglutição, respiração e produção de voz [2].

A estrutura rígida da laringe é formada por quatro cartilagens principais, a tiroide, a cricoide e duas aritenoides [21]. Para além destas, fazem também parte da laringe a epiglote, e dois pares de cartilagens secundárias, as corniculadas e as cuneiformes. O esqueleto cartilágneo da laringe é sustentado principalmente pelo osso hiode e unido entre si através de músculos, articulações, membranas e ligamentos [9].

A maior cartilagem da laringe é a tiroide e apresenta uma estrutura semelhante à de um escudo composto por duas lâminas retangulares que se unem na parte frontal, rodeando a laringe sem fechá-la por completo, como mostra a Figura 2.2. Na parte inferior da tiroide encontra-se a cartilagem cricoide que tem a forma de um anel circular e envolve a totalidade da via aérea [20, 9]. As aritenoides são duas cartilagens tetraédricas bilaterais e são consideradas a unidade funcional da laringe, uma vez que mudam de localização e orientação entre a fonação e a respiração [20]. Em conjunto, as cartilagens da laringe formam uma estrutura que sustenta e abriga as pregas vocais [21].

A musculatura da laringe engloba músculos intrínsecos e extrínsecos. Os músculos intrínsecos têm origem na laringe e possuem uma relação direta com a fonação [9]. Quando contraídos diferencialmente podem mover as cartilagens em relação umas às outras, resultando na abertura (abdução) ou fecho (adução) da glote, bem como na modificação das propriedades mecânicas do tecido das pregas vocais [21]. Por outro

lado, os músculos extrínsecos apresentam ligações com as cartilagens laríngeas, mas são provenientes de estruturas não-laríngeas, como o tórax, a mandíbula ou o crânio. Desempenham um papel fundamental na deglutição e atuam de forma menos direta na fonação, através da modificação do posicionamento da laringe [9].

As propriedades das pregas vocais (posição, forma, massa, tensão e constante elástica) dependem da ação da musculatura intrínseca e extrínseca da laringe [22], que por sua vez é controlada pelo sistema nervoso central. A musculatura laríngea é inervada principalmente por duas ramificações do nervo vago (X par craniano), o Nervo Laríngeo Recorrente (NLR) e Nervo Laríngeo Superior (NLS). O NLR, retratado na Figura 2.2, é responsável pela inervação motora de quase todos os músculos intrínsecos da laringe que acionam as pregas vocais. Enquanto que o ramo externo do NLS, garante a mobilidade do músculo cricotiróideo e o seu ramo interno a inervação sensitiva da região supraglótica [23, 9].

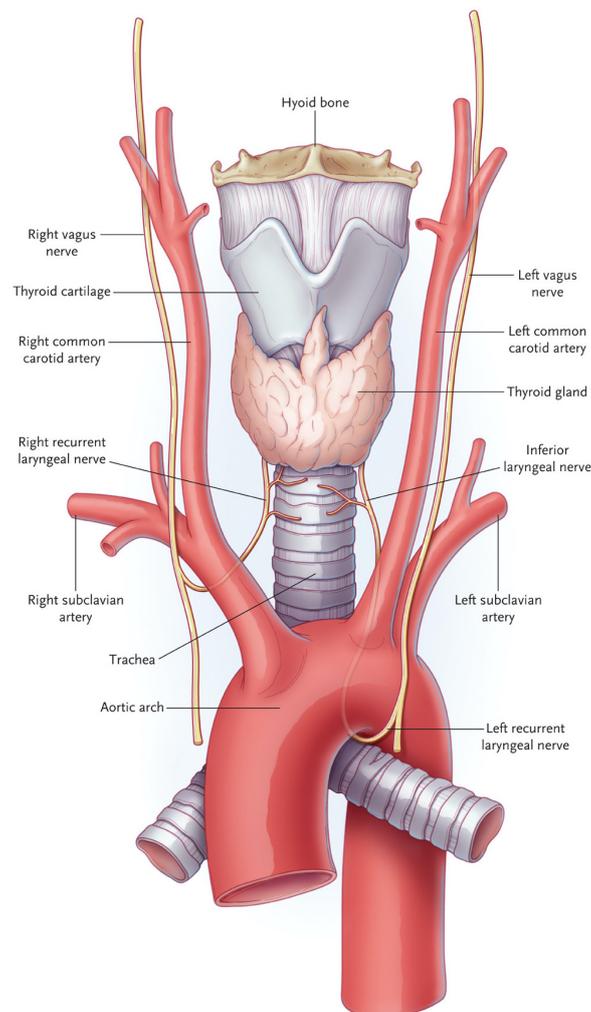


Figura 2.2: Representação esquemática das estruturas anatômicas da laringe [24].

As pregas vocais são dois músculos, que se estendem horizontalmente na laringe e estão fixados na superfície interna da cartilagem tireoide, na região de convergência de ambas as pregas [18, 9]. Na porção posterior, as pregas divergem, sendo que cada uma delas se conecta à respectiva cartilagem aritenoide [9]. O espaço livre compreendido entre as pregas vocais é chamado de glote. A glote muda de forma e comprimento várias vezes durante a fala, consoante os movimentos das pregas vocais [20].

A estrutura morfológica das pregas vocais é frequentemente descrita pelo conceito “corpo-cobertura” proposto por Hirano [22], no início dos anos 70. Embora, ao longo dos anos, tenham sido apresentados esquemas mais detalhados para descrever o modelo anatómico das pregas vocais, este esquema é particularmente útil para estudar as características vibratórias das pregas. Esta teoria sugere que as pregas vocais podem ser subdivididas em duas camadas principais, com propriedades mecânicas diferentes. A cobertura é composta por tecido mucoso, flexível, maleável e não contrátil. Enquanto que o corpo é constituído por fibras musculares e ligamentos, pelo que apresenta propriedades contráteis [22].

2.1.2 Fisiologia da Fonação

A propagação das ondas sonoras que originam a fonação, resulta da conversão do fluxo de ar constante em fluxo de ar pulsado por ação do trato vocal [25]. De acordo com a teoria mioelástica-aerodinâmica, tal é possível devido ao equilíbrio mantido entre as forças aerodinâmicas, resultantes do Efeito de Bernoulli, e as forças mioelásticas, que oferecem resistência à passagem do ar pela glote [26].

Durante a fonação, as pregas vocais realizam movimentos vibratórios cíclicos, abrindo e fechando a passagem ao fluxo de ar [20]. Na Figura 2.3 estão representadas as 4 fases características de cada ciclo glótico.

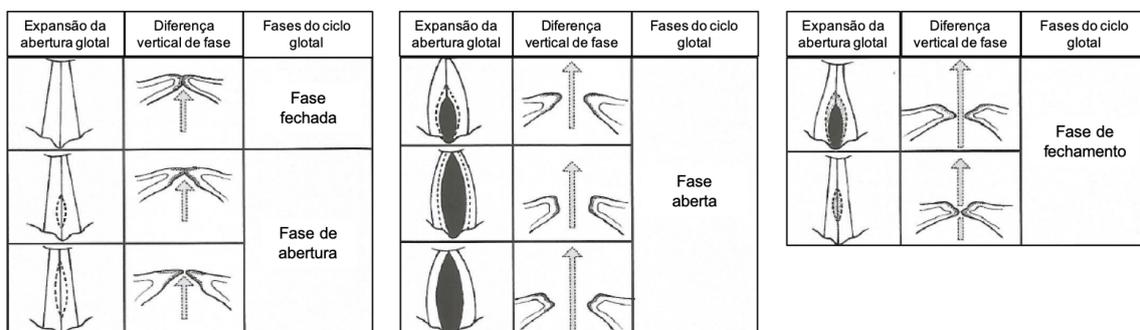


Figura 2.3: Representação esquemática das 4 fases do impulso glótico, resultantes de diferenças ao nível da pressão subglótica. Adaptado de [9]

Inicialmente as pregas vocais encontram-se unidas e a glote totalmente fechada, o que resulta num aumento da pressão subglótica. À medida que a pressão aumenta as pregas começam a afastar-se uma da outra de baixo para cima, mantendo-se somente o contacto entre as bordas superiores das mesmas. Quando a pressão é suficiente para afastar totalmente as pregas, o ar começa a sair pela glote [2]. Com base no Efeito de Bernoulli, o aumento da velocidade do fluxo de ar desencadeia a redução da pressão no ponto de constricção. A pressão negativa aliada à elasticidade da camada mucosa das pregas, promove o efeito de sucção e aproximação das comissuras inferiores das mesmas. Por fim, as pregas voltam a estar em contacto e o fluxo de ar é totalmente interrompido [2, 9]. Este ciclo é denominado por impulso ou período glótico e repete-se, durante a fonação, com uma determinada frequência fundamental [27].

A oscilação das pregas vocais é quase periódica, com registos de perturbações mínimas ao nível da frequência e da amplitude, que são quantificadas, respetivamente, pelos parâmetros *jitter* e *shimmer* [9]. Estas irregularidades surgem de uma combinação de fatores biomecânicos (assimetria das pregas vocais), neurogénicos (atividades involuntárias dos músculos laríngeos), e aerodinâmicos (flutuações do fluxo de ar e da pressão subglótica) [20]. Ambos os parâmetros são imprescindíveis na análise acústica da voz, pois permitem estabelecer uma distinção entre vozes patológicas e vozes saudáveis, como será fundamentado nas próximas secções.

2.1.3 Voz Normal vs. Disfonia

Segundo Aronson [28], é mais difícil definir o conceito de voz normal do que qualquer outra componente da fala, porque a variedade de vozes é ilimitada e os seus padrões de adequação são amplos. Assim sendo, o conceito de voz normal não é consensual, contudo foram propostos padrões gerais de normalidade de forma a caracterizar sujeitos com vozes saudáveis [9]. De entre os padrões propostos destacam-se: qualidade vocal agradável; frequência fundamental adequada à idade e ao sexo do orador; intensidade apropriada ao evento de comunicação; flexibilidade e sustentabilidade adequadas [29].

Tal como não pode ser estabelecido nenhum critério absoluto para voz normal, não existe um padrão fixo e uniforme que defina uma voz anormal. Uma voz é considerada anormal ou patológica sempre que a qualidade, a altura tonal, o timbre ou a flexibilidade na dicção, são desagradáveis ou inadequadas para a faixa etária, sexo e grupo cultural do falante [29]. Por definição, disfonia é qualquer dificuldade ou distúrbio na produção vocal que impeça a emissão natural da voz. Esta pode

manifestar-se através de diversas alterações: dificuldade em sustentar a voz; fadiga vocal; variações na frequência fundamental; rouquidão; falta de volume e projeção; perda de eficiência e de resistência vocal; e desconforto durante a emissão [11].

Dado que a voz apresenta características multidimensionais foram propostos vários critérios de classificação das disfonias. A abordagem clássica de classificação baseia-se na dicotomia funcional e orgânica. Nas disfonias de origem funcional não são observadas alterações anatômicas nas pregas vocais, as alterações que se verificam decorrem do uso da própria voz. Este tipo de disфонia pode ser causado por abuso, mau uso ou inadaptação vocal ou por alterações emocionais. No seguimento deste tipo de comportamentos de comunicação, surgem as disfonias orgânico-funcionais, que resultam de disfonias funcionais que progridem para lesões secundárias da prega vocal [11, 30]. Por sua vez, as disfonias orgânicas incluem alterações anatômicas e estruturais da laringe, que englobam lesões de massa como, nódulos e pólipos vocais, edema de Reinke, úlceras de contacto, paralisias nas pregas vocais, entre outras [11, 31].

2.1.4 Influência da Tireoidectomia na Produção de Voz

Para além das perturbações vocais que são consequência direta do uso inadequado da voz ou de patologias laríngeas, podem surgir alterações causadas por traumas físicos. As lesões traumáticas são frequentemente observadas após qualquer cirurgia que envolva anestesia geral ou manipulação laríngea [13]. Assim, resultam de entubações endotraqueais que causam hematomas/granulomas ou de danos provocados nos nervos e músculos durante procedimentos cirúrgicos. [32, 33].

A cirurgia da tiroide é um dos procedimentos com maiores consequências ao nível da produção natural da voz, sejam elas transitórias ou permanentes. Apesar da cautela dos cirurgiões durante o procedimento, lesões nos nervos laríngeos são por vezes inevitáveis [34]. Por este motivo, os estudos têm recaído sobretudo na disfunção da tiroide após a cirurgia e não tanto no impacto que as irregularidades endócrinas possuem na produção de voz [35]. A cirurgia da tiroide consiste na remoção total (tireoidectomia total) ou parcial (hemitireoidectomia) da glândula tiroide, dependendo da extensão da patologia em causa. Pode ser realizada por indicações clínicas que abrangem a remoção de nódulos benignos ou quistos, a remoção de tumores malignos, o tratamento do hipertireoidismo, a remoção de um bócio que interfira com a deglutição/respiração, ou a remoção de nódulos com análises indeterminadas na biopsia [36].

A glândula tiroide localiza-se na base do pescoço, na região anterior à traqueia e inferior à cartilagem tiroide, como apresentado anteriormente na Figura 2.2. Encontra-se envolvida pela musculatura cervical e inervada pelas ramificações do nervo vago. Estas estruturas podem estar diretamente envolvidas pela patologia da glândula tiroide e sofrer sequelas durante a tiroidectomia [13]. O local de lesão mais comum é no Nervo Laríngeo Recorrente (NLR), cuja principal função é controlar a mobilidade das pregas vocais. Lesões em qualquer ponto da trajetória do nervo vago, desde a sua origem no tronco encefálico até as suas ramificações e inserções na laringe, podem levar ao surgimento de paralisias uni ou bilaterais das pregas vocais. Embora menos frequentemente, durante a cirurgia também podem ocorrer lesões no Nervo Laríngeo Superior (NLS), em particular no Ramo Externo do Nervo Laríngeo Superior (RENLS), que prejudicam a capacidade de afinação e projeção vocal [13]. Uma técnica cirúrgica delicada pode evitar tais complicações, pelo que o cirurgião deve ter um conhecimento profundo do curso e variações de curso destes nervos.

Outras possíveis causas de disfonia pós-cirúrgica incluem lesões da musculatura extrínseca e intrínseca da laringe, irritação laríngea, edemas, hemorragias e danos nos tecidos moles [37]. Além disso, a remoção da glândula tiroide modifica o fornecimento vascular e a drenagem venosa da laringe, provocando alterações na sua mucosa que podem originar perturbações vocais [38].

A disfunção vocal pode ter consequências significativas na vida profissional e social do indivíduo, sendo que as principais queixas são dificuldade em produzir um tom alto, rouquidão, fadiga vocal e dificuldade em engolir [39]. A análise acústica computadorizada de doentes submetidos a tiroidectomias pode ajudar a determinar objetivamente as possíveis alterações vocais [37].

2.2 Técnicas de Avaliação da Voz

A fonação pode ser afetada por uma complexa diversidade de fatores, pelo que a produção de voz é um processo altamente variável, não apenas de pessoa para pessoa, mas também dentro da mesma pessoa [1]. Assim sendo, a avaliação da qualidade vocal baseia-se tradicionalmente num protocolo multidimensional que engloba vertentes subjetivas e objetivas [5].

Conforme proposto pelo Comité de Foniatria da Sociedade Europeia de Laringologia (ELS) [5], a avaliação da voz inclui cinco abordagens distintas: avaliação do impacto psicossocial, avaliação áudio-percetiva, videoestroboscopia, avaliação aerodinâmica e avaliação acústica [5]. Antes de cada procedimento é comum o clínico

realizar uma entrevista ao paciente, que resulta na história clínica do mesmo [23]. Adicionalmente, a fisiologia da laringe também pode ser caracterizada por meio de exames de Eletromiografia e Eletroglotografia. Individualmente nenhum método é suficiente para classificar de forma abrangente e precisa a qualidade vocal. Todas as abordagens são complementares, possibilitando um conhecimento amplo da função laríngea e uma correta definição do diagnóstico clínico e terapêutico [23].

História Clínica: Este tipo de avaliação tem como objetivo, estabelecer a história cronológica do problema vocal, procurar fatores etiológicos associados e determinar a motivação do paciente para a resolução do problema. Para tal, é realizado um questionário que permite recolher informações que contribuem para uma visão holística acerca dos antecedentes da patologia e das possíveis causas da mesma [40, 23].

Autoavaliação do paciente: Esta avaliação é realizada pelo próprio paciente e, embora subjetiva, é de grande importância na prática clínica. O paciente convive diariamente com a sua voz e tem um conhecimento íntimo de todos os seus aspetos e do seu impacto a níveis sociais, culturais e profissionais [5]. O índice de perturbação da voz pode ser calculado com base nas respostas do paciente a uma lista de perguntas cuidadosamente selecionadas. Os instrumentos utilizados para esse efeito são o *Voice Handicap Index* (VHI) [41] e o *Voice-Related Quality of Life Measure* (V-RQOL) [42].

Avaliação Áudio-percetiva: A avaliação perceptiva sugere que o médico avalie uma amostra vocal produzida pelo falante. Por norma, a avaliação centra-se na produção de vogais sustentadas, na leitura de frases ou textos e no discurso espontâneo [23]. No entanto, este método apresenta algumas limitações nomeadamente o facto de não existir uma definição universal de voz normal e o facto de depender da qualificação e experiência do avaliador [43]. Na tentativa de minimizar as variações das avaliações inter e intra avaliadores, foram desenvolvidas inúmeras escalas de avaliação perceptiva, de entre as quais se destacam a escala *Grade, Roughness, Breathiness, Asthenia, Strain* (GRBAS) [44] e a escala *Consensus Auditory-Perceptual Evaluation of Voice* (CAPE-V) [45].

Avaliação Videoestroboscópica: A videoestroboscopia é o principal instrumento clínico de diagnóstico etiológico das perturbações vocais. Permite avaliar a fisiologia laríngea e a qualidade da vibração da prega vocal, bem como a eficácia dos tratamentos, médicos ou cirúrgicos [5]. Este exame pode ser efetuado durante a fonação, recorrendo a um endoscópio rígido ou flexível ligado a um videogravador, de forma a visualizar diretamente o padrão de vibração das pregas vocais e gravar as respetivas imagens para referência futura [40].

Avaliação Aerodinâmica: Vários métodos, tanto diretos como indiretos, têm sido utilizados para avaliar quantitativamente os volumes de ar, fluxos de ar e pressões de ar envolvidas na produção de voz. O método direto consiste na medição da pressão de ar subglótica através da inserção de uma agulha hipodérmica no espaço cricotiróideo e caracteriza-se por ser particularmente invasivo e doloroso [46]. Por outro lado, as estimativas indiretas da pressão de ar subglótica podem ser obtidas, de forma não invasiva, por meio da determinação da pressão intraoral. Para tal, é necessária a utilização de uma máscara facial de Rothenberg [47], conectada a transdutor de pressão.

Eletroglotografia (EGG): A EGG é uma técnica não invasiva, que permite caracterizar o comportamento muco-ondulatório das pregas vocais [46]. Este procedimento envolve a colocação de dois elétrodos, de ambos os lados das lâminas tiroideias, pelos quais passa uma pequena corrente elétrica à medida que as pregas vocais vibram. Os elétrodos medem a resistência variável proporcionada pelos movimentos vibratórios e traduzem essa informação na área de contacto estabelecida entre as pregas vocais durante a fonação [40].

Eletromiografia (EMG): A EMG é um procedimento invasivo que se baseia na medição direta dos potenciais elétricos da atividade muscular laríngea. O exame é realizado com elétrodos bipolares em forma de agulha que são inseridos por via percutânea nos músculos tirearitenóideo e cricotiróideo, refletindo o padrão de atividade dos nervos laríngeo recorrente e laríngeo superior, respetivamente [9].

Avaliação Acústica: A análise acústica é uma técnica não invasiva baseada no processamento digital de som. Trata-se de uma ferramenta comum na prática clínica, geralmente usada como complemento à avaliação áudio-percetiva e à videoestroboscopia [34, 6].

A maioria das medidas acústicas resulta de amostras de vogais sustentadas e com menos frequência de amostras de discurso corrente [48]. A justificação para esta preferência prende-se com o facto de que as vogais sustentadas representam uma fonação estável e não registam efeitos de coarticulação, sendo relativamente imunes às influências linguísticas [23]. Além disso, em contraste com a fala contínua, as vogais sustentadas são de fácil compreensão, registo e análise, uma vez que não contêm fonemas não vocalizados e não são afetadas pela velocidade da fala, pelo contexto fonético ou por questões de cariz emocional (stress) [48]. Embora a produção de vogais sustentadas seja atrativa por várias razões, não se pode confiar exclusivamente neste contexto de voz, por não ser representativo dos padrões diários de comunicação e por poder encobrir os efeitos da disфонia [48].

Uma das principais vantagens desta técnica é a de proporcionar um diagnóstico objetivo e mensurável da qualidade vocal, através da determinação de propriedades físicas da onda sonora (periodicidade, amplitude, duração e composição espectral) [6]. Todos os resultados e gráficos obtidos podem ser arquivados e exportados para a maioria dos *softwares* disponíveis no mercado [49]. Além disso, é uma técnica útil como instrumento de triagem, dado que o conceito de voz normal é definido por valores de referência acima dos quais a voz é considerada patológica [9]. É importante salientar que a maioria das análises acústicas depende de formas de onda quase-periódicas, o que dificulta a sua aplicação em vozes excessivamente irregulares e ruidosas [50].

As interações entre o *hardware* de gravação, o ambiente da sala e o *software* de extração dos parâmetros podem ser extremamente complexas, pelo que a interpretação correta do sinal requer cuidados especiais [25]. Para esse efeito Behlau [9], sugere quatro categorias de normas a seguir para obter um registo de áudio adequado: quanto às condições de registo, quanto aos protocolos de gravação, quanto aos sistemas de gravação e análise, e quanto à análise qualitativa do sinal acústico.

Um estudo realizado por Solomon *et al.* [34] revelou que a análise acústica de vogais sustentadas é apropriada para monitorizar alterações vocais em pacientes submetidos a tiroidectomias à medida que recuperam da cirurgia. Nesta dissertação, também vai ser elaborada uma análise acústica de vogais sustentadas, uma vez que o objetivo inicial do projeto consistia no auxílio ao pré-diagnóstico e acompanhamento terapêutico dos indivíduos submetidos a tiroidectomias. A comparação das medidas acústicas antes e depois de um tratamento pode servir dois propósitos. Antes do tratamento, os parâmetros acústicos fornecem provas indiretas da gravidade do problema vocal, apesar de não identificarem a etiologia da patologia em causa. Durante o tratamento, estas medidas podem ajudar a avaliar o progresso do paciente nas diversas fases da terapia [40].

2.2.1 Programas de Análise Acústica da Voz

A avaliação acústica da voz por meio de *softwares* têm despertando interesse em contextos clínicos, pedagógicos e de investigação, nomeadamente no que diz respeito à caracterização da fisiologia do aparelho fonador [51]. As principais utilidades dos *softwares* incluem: feedback visual e auditivo da gravação; determinação de parâmetros quantitativos, objetivos e fiáveis, que podem ser armazenados para apreciações futuras; e avaliação e diagnóstico da intervenção terapêutica de patologias vocais [51].

Existem vários *softwares* de livre acesso ou comercializados, sendo que os recursos disponíveis dependem do tipo de *software* escolhido. Apesar de todos implementarem técnicas de processamento digital de som, diferem entre eles em termos de potencialidades, velocidade, compatibilidade e custo [23]. A literatura apresenta um grande número de estudos, cada um com configurações próprias no que diz respeito à acústica da sala, tipo e colocação do microfone, algoritmo de análise e *software* implementado. Isto limita a comparação direta entre estudos e influencia a fiabilidade e validade dos parâmetros acústicos para os diferentes *softwares* disponíveis [52, 53].

No âmbito desta dissertação, foi desenvolvido um algoritmo em MATLAB capaz de extrair alguns dos parâmetros de qualidade vocal que se obtêm nos diferentes *softwares*. De forma a avaliar o seu desempenho foi estabelecida uma análise comparativa com o *software* Praat, por ser bastante conceituado em meio científico nas áreas de fonética e fonologia. Posto isto, é elaborada uma breve introdução a este *software* na secção seguinte.

2.2.1.1 Praat

O Praat é um programa de computador de livre acesso que possibilita a análise, sintetize e manipulação da fala [54]. Tem sido desenvolvido desde 1992 por Paul Boersma e David Weenink no Instituto de Ciências Fonéticas da Universidade de Amsterdão [55].

Trata-se de um *software* muito flexível e maioritariamente utilizado no meio académico e de investigação. Oferece uma vasta gama de procedimentos que abrangem análises espectrográficas, das quais é possível extrair a periodicidade, a intensidade e parâmetros qualitativos, como o *jitter*, o *shimmer* e o HNR [56]. Para cada um dos parâmetros estão disponíveis ficheiros de auxílio que explicam os respetivos algoritmos implementados.

2.2.2 Parâmetros Acústicos

A maioria das abordagens encontradas na literatura recorre à deteção automática de perturbações vocais por meio de técnicas de processamento digital de sinais de longa duração, como é o caso das vogais sustentadas [6]. Os parâmetros mais relevantes para a análise acústica da voz são a Frequência Fundamental (F_0), o *jitter* e o *shimmer* e a *Harmonic-to-Noise Ratio* (HNR) [57].

2.2.2.1 Frequência Fundamental

A Frequência Fundamental (F_0) é o parâmetro físico resultante da taxa de vibração das pregas vocais, que corresponde à velocidade à qual a forma da onda acústica se repete por unidade de tempo [9, 40]. A nível fisiológico, cada repetição da onda equivale à duração de um período glótico (T), como exemplificado na Figura 2.4. Uma vez determinado o período glótico, a F_0 pode ser calculada a partir do seu inverso, como representado na equação 2.1.

$$F_0(\text{Hz}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{T_i} \quad (2.1)$$

Onde T_i é a duração de um período glótico i e N é o número de períodos glóticos.

A F_0 é uma medida da eficiência do sistema fonatório, que reflete a interação da biomecânica laríngea com a pressão subglótica, assim como o resultado natural do comprimento e espessura das pregas vocais de um indivíduo [9, 23]. Os principais mecanismos que alteram a frequência vocal são a regulação do comprimento, tensão e massa das pregas vocais, e o ajuste dos fatores aerodinâmicos [20].

Há vários fatores envolvidos nas modificações da frequência vocal, desde a idade, o sexo, os hábitos quotidianos (álcool, tabaco) ou a existência disfonias. O valor de F_0 também difere de acordo com o comportamento vocal, sendo que no caso de vogais sustentadas varia consoante a vogal produzida. Por norma, vogais altas, como o /i/, apresentam F_0 mais elevada que vogais baixas, como o /a/ [23].

A informação normativa quanto aos valores de F_0 apresenta uma grande variabilidade entre estudos. Contudo, é possível chegar a um consenso e atribuir valores de referência para o português europeu. Tipicamente, a F_0 das mulheres é mais elevada que a dos homens, sendo que numa voz feminina varia entre 150 e 350 Hz e numa voz masculina entre 80 e 200 Hz [58]. Relativamente à faixa etária, as crianças são as que apresentam maiores valores de F_0 podendo variar entre 200 e 400 Hz [59].

2.2.2.2 Jitter

O *jitter* é o parâmetro que reflete a variabilidade da Frequência Fundamental ciclo-a-ciclo, pelo que é considerado um índice de perturbação de F_0 a curto prazo [9]. O princípio do *jitter* encontra-se representado na Figura 2.4. Este parâmetro está correlacionado com a qualidade e com a regularidade da vibração das pregas vocais e permite determinar o grau de estabilidade do sistema fonatório [57, 23].

O valor de *jitter* pode ser extraído através de diversas medidas: *jitter* local, *jitter* absoluto, *jitter* RAP e *jitter* PPQ5. A determinação destes parâmetros pode ser efetuada de acordo com as equações propostas por Boersma [54], apresentadas em seguida.

***Jitter* (local, absoluto):** O *Jitter* absoluto ou *Jitta* representa a média da diferença absoluta entre dois períodos consecutivos, expressa em segundos. O valor limiar para detetar patologias é 83.200 μs [54].

$$jitta(s) = \frac{1}{N-1} \sum_{i=2}^N |T_i - T_{i-1}| \quad (2.2)$$

Onde T_i é a duração de um período glótico i e N é o número de períodos glóticos.

***Jitter* (local):** O *Jitter* relativo representa a média da diferença absoluta entre períodos consecutivos, dividida pelo período médio, expressa em percentagem. O valor limiar para detetar patologias é 1.04% [54].

$$jitter(\%) = \frac{\frac{1}{N-1} \sum_{i=2}^N |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \times 100 \quad (2.3)$$

***Jitter* (RAP):** A Perturbação Média Relativa (*Relative Average Perturbation*) representa a média da diferença absoluta entre um período e a média desse período com os seus dois vizinhos, dividida pelo período médio, em percentagem. O valor limiar para detetar patologias é 0.680% [54].

$$RAP(\%) = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| T_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} \times 100 \quad (2.4)$$

***Jitter* (PPQ5):** O Quociente de Perturbação num intervalo de 5 períodos (*five-point Period Perturbation Quotient*) representa a média da diferença absoluta entre um período e a média desse período com os seus quatro vizinhos, dividida pelo período médio, em percentagem. O valor limiar para detetar patologias é 0.840% [54].

$$PPQ5(\%) = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} \left| T_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} \times 100 \quad (2.5)$$

É de esperar que indivíduos normais apresentem perturbações de F_0 , influenciadas pelas características biomecânicas das pregas vocais e por fatores de ordem neurológica e emocional. Assim como em F_0 , os valores de *jitter* também variam de acordo com a idade e sexo do falante, sendo que a frequências mais altas correspondem perturbações menores [23].

Embora este parâmetro não defina a etiologia da disfonia, é útil na distinção entre voz “normal” e disfónica, uma vez que a sua magnitude reflete a extensão da alteração encontrada [9]. O *jitter* altera-se, sobretudo, devido à instabilidade da vibração das pregas vocais. Indivíduos com disfonias de origem neurológica, por vezes advindas de tiroidectomias, não controlam a vibração das pregas e tendem a apresentar valores de *jitter* mais elevados [60, 9].

2.2.2.3 *Shimmer*

O *shimmer* é o parâmetro que traduz a variabilidade da amplitude do sinal ciclo-a-ciclo, isto é, quantifica alterações mínimas de amplitude entre ciclos consecutivos, como ilustrado na Figura 2.4 [23, 57]. Tal como o *jitter*, é um índice de perturbação a curto prazo que permite avaliar a estabilidade da fonação [9].

Este parâmetro pode ser calculado de quatro maneiras diferentes: *shimmer* local, APQ3, APQ5 e APQ11. São novamente apresentadas as equações sugeridas por Boersma [54].

Os métodos usados para determinar estas medidas são idênticos ao usados para determinar as medidas de *jitter*. A principal diferença é que nos cálculos de *jitter* são considerados os períodos e nos de *shimmer* são tidas em conta as amplitudes do sinal [8].

***Shimmer* (local):** O *shimmer* relativo é a média da diferença absoluta entre as amplitudes de dois períodos consecutivos, dividida pela amplitude média, em percentagem. Cujo limiar de patologia é 3.810% [54].

$$shimmer(\%) = \frac{\frac{1}{N-1} \sum_{i=2}^N |A_i - A_{i-1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (2.6)$$

Onde A_i é a amplitude de um período glótico i e N é o número de períodos glóticos.

***Shimmer* (apq3):** O Quociente de Perturbação da Amplitude de três pontos (*three-point Amplitude Perturbation Quotient*) é a média da diferença absoluta entre a amplitude de um período e a média das amplitudes desse período e dos seus dois

vizinhos, dividida pela amplitude média, em percentagem [54].

$$APQ3(\%) = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| A_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (2.7)$$

Shimmer (apq5): O Quociente de Perturbação da Amplitude de cinco pontos (*five-point Amplitude Perturbation Quotient*) é a média da diferença absoluta entre a amplitude de um período e a média das amplitudes desse período e dos seus quatro vizinhos, dividida pela amplitude média, em percentagem [54].

$$APQ5(\%) = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} \left| A_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (2.8)$$

Shimmer (apq11): O Quociente de Perturbação da Amplitude de onze pontos (*eleven-point Amplitude Perturbation Quotient*) é a média da diferença absoluta entre a amplitude de um período e a média das amplitudes desse período e dos seus dez vizinhos, dividida pela amplitude média, em percentagem. Cujo limiar de patologia é 3.070% [54].

$$APQ11(\%) = \frac{\frac{1}{N-10} \sum_{i=6}^{N-5} \left| A_i - \left(\frac{1}{11} \sum_{n=i-5}^{i+5} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (2.9)$$

O *shimmer* é inversamente proporcional à intensidade vocal média, independentemente da tarefa vocal realizada e da existência ou não de patologia [61]. Além disso, representa a energia total da voz e oferece uma percepção indireta do ruído na produção vocal. Assim, esta medida tende a ser tanto maior quanto maior for a quantidade de ruído de uma emissão e menores forem a frequência e a intensidade da mesma [59, 9].

De modo análogo ao *jitter*, indivíduos normais também apresentam irregularidades na amplitude, introduzidas pelas características das pregas vocais. Indivíduos com distúrbios laríngeos tem tendência a apresentar valores de *shimmer* mais elevados, principalmente quando a resistência glótica diminui ou quando surgem lesões de massa nas pregas vocais [9].

Há algumas evidências de que as medidas do *shimmer* podem ser mais sensíveis à percepção do ruído vocal e um melhor indicador de patologia laríngea do que as medidas de *jitter* [62, 63].

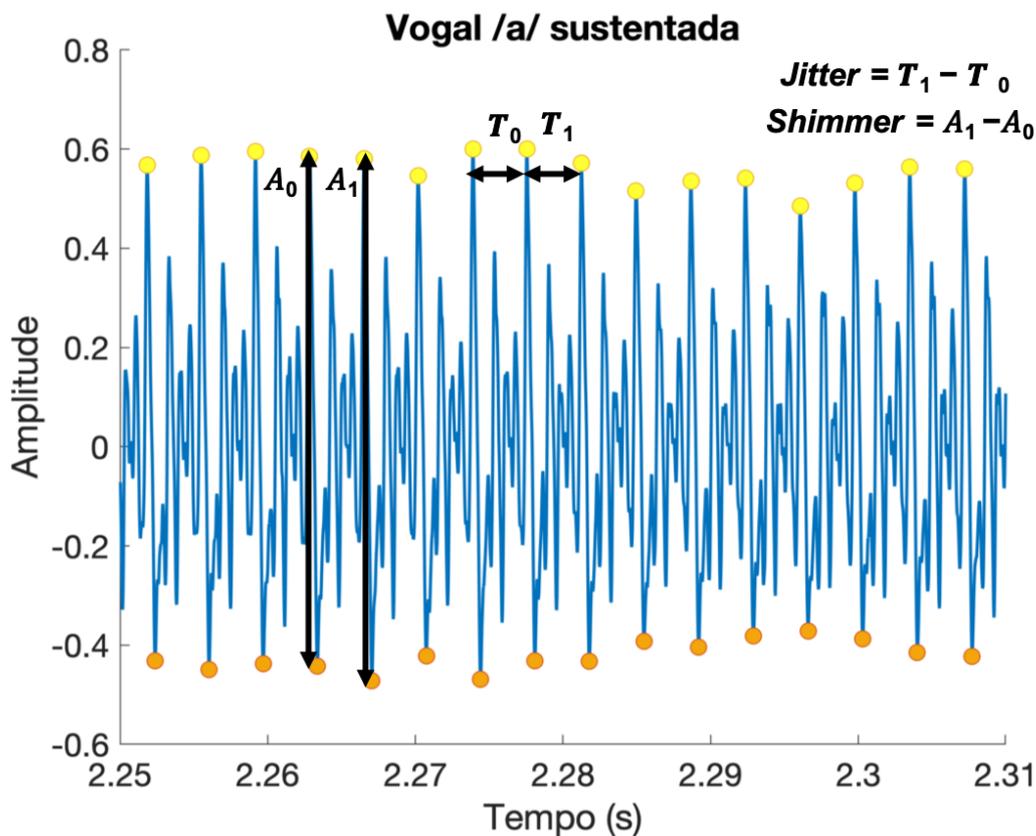


Figura 2.4: Representação do *jitter* e do *shimmer* na produção da vogal /a/ sustentada. Os máximos locais são ilustrados pelos pontos amarelos e os mínimos locais pelos pontos laranja e indicam os respectivos instantes entre os quais é calculado o *jitter* e as respectivas amplitudes entre as quais é calculado o *shimmer*.

2.2.2.4 Harmonic-to-Noise Ratio (HNR)

A *Harmonic-to-Noise Ratio* é uma medida de ruído que indica a distribuição da energia acústica e o ruído gerado pela turbulência resultante da passagem do ar pela glote, durante a fonação [49]. Matematicamente, corresponde à relação entre a componente periódica (harmónica) e aperiódica (ruído glótico) da onda acústica de um segmento de fala sustentada [64]. Assim, a primeira componente traduz a vibração das pregas e a segunda o ruído glótico. A relação entre ambas reflete o desempenho do processo de fonação [65].

Existem diversos métodos para determinar o parâmetro HNR. O método implementado no Praat consiste num algoritmo desenvolvido por Boersma [66]. A sua abordagem baseia-se na deteção da periodicidade acústica, por meio de num método de autocorrelação normalizada do sinal de voz. Após o cálculo da autocorrelação, para determinar o HNR basta medir a energia do primeiro pico, correspondente à componente harmónica, e considerar a restante energia como a energia do ruído.

Desta forma, a energia do ruído pode ser determinada pela diferença entre 1 e a energia harmónica, como representada na equação 2.10 [60].

$$HNR(dB) = 10 \times \log_{10} \frac{H}{1 - H} \quad (2.10)$$

Onde H corresponde ao componente harmónico dado pela energia do primeiro pico da autocorrelação normalizada do sinal.

Tal como para os parâmetros anteriores, as vozes normais também apresentam uma quantidade de ruído expectável, relacionada com as perturbações naturais da produção de voz [9]. Uma voz saudável apresenta um HNR elevado, associado à perceção de voz sonora e harmónica, enquanto que um valor baixo de HNR evidencia uma voz asténica e disfónica [65].

De um modo geral, as mulheres apresentam valores de HNR mais elevados que os homens, dado que quanto mais elevada é a frequência maior é o componente harmónico da onda acústica [9]. Contudo, os valores de HNR relatados na literatura não são totalmente consensuais e variam de acordo com o *software* de análise acústica. Os principais fatores responsáveis pelas discrepâncias encontradas podem ser: diferentes algoritmos utilizados, diferentes procedimentos de registo, diferentes instrumentos de gravação, e diferentes amostras vocais [67].

De acordo com Boersma [54], um indivíduo saudável pode produzir as vogais /a/ ou /i/ com uma harmonização de cerca de 20dB. E um estudo realizado Yumoto *et al.* [64], pressupõe que valores de HNR inferiores a 7dB podem ser considerados patológicos.

O HNR é mais sensível à qualidade vocal que os parâmetros anteriormente descritos, uma vez que engloba as perturbações da frequência, da amplitude e da morfologia da forma de onda [29]. Por esse motivo, é indicado na monitorização da evolução dos tratamentos de disfonias e de tiroidectomias [9].

Metodologia

Neste capítulo é elaborada a caracterização da amostra submetida à avaliação acústica, no decorrer do projeto. É também apresentada uma breve descrição do protocolo que foi utilizado nessa mesma avaliação e do equipamento usado para o seu registo. Por último, procede-se à explicação dos métodos de extração dos parâmetros e do algoritmo desenvolvido para esse propósito.

3.1 Amostra

A amostra do presente estudo é composta por um total de 30 indivíduos, sendo 15 do género masculino e os restantes 15 do género feminino. Os indivíduos do género masculino apresentam idades compreendidas entre os 11 e os 61 anos, com uma média de idades de 30.2. A amostra da população feminina apresenta idades compreendidas entre os 15 e os 53 anos, com uma média de idades de 27.3. Com o intuito de caracterizar a amostra antes de cada avaliação, foi realizado um questionário simplificado, em substituição da entrevista clínica. Este consistiu no levantamento de informação relevante como a idade, os hábitos tabágicos e os históricos clínicos de entubação endotraqueal, de patologias relacionadas com a tiroide e de patologias respiratórias, como asma ou infeções respiratórias. A caracterização das amostras feminina (F) e masculina (M) encontra-se nas Tabelas 3.1 e 3.2, respetivamente. Em nenhum momento foram registados dados passíveis de identificar os pacientes inequivocamente, sendo as amostras sempre tratadas com uma referência ao sujeito na forma F_X e M_X, com F e M correspondentes ao género do sujeito e com X de 1 a número de sujeitos.

Para constituir a amostra não foram propostos critérios de inclusão nem de exclusão, uma vez que o objetivo principal passou a ser a análise comparativa entre os valores obtidos com o *software* Praat e os determinados com o algoritmo desenvolvido no MATLAB. Como se trata de um estudo comparativo, a componente de maior inte-

3. Metodologia

resse passa a ser a semelhança entre os resultados e os restantes critérios deixam de ser significativos.

Tabela 3.1: Caracterização da amostra Feminina.

Sujeito	Idade	Hábitos tabágicos	Histórico de entubação	Antecedentes de doenças da tiroide	Problemas respiratórios/vocais
F_1	15	Não	Não	Não	Não
F_2	16	Não	Não	Não	Não
F_3	21	Sim	Não	Não	Não
F_4	21	Sim	Não	Não	Não
F_5	22	Não	Não	Não	Não
F_6	22	Não	Não	Não	Não
F_7	22	Não	Não	Não	Não
F_8	22	Não	Não	Não	Asma
F_9	22	Não	Não	Não	Não
F_10	23	Não	Não	Não	Asma
F_11	23	Não	Não	Não	Não
F_12	25	Não	Não	Não	Não
F_13	50	Não	Não	Nódulos	Não
F_14	52	Não	Não	Nódulos	Não
F_15	53	Não	Não	Não	Não

Tabela 3.2: Caracterização da amostra Masculina.

Sujeito	Idade	Hábitos tabágicos	Histórico de entubação	Antecedentes de doenças da tiroide	Problemas respiratórios/vocais
M_1	11	Não	Não	Não	Não
M_2	13	Não	Não	Não	Não
M_3	17	Não	Não	Não	Não
M_4	21	Sim	Não	Não	Não
M_5	22	Não	Não	Não	Não
M_6	22	Não	Não	Não	Não
M_7	22	Não	Não	Não	Não
M_8	23	Não	Não	Não	Não
M_9	23	Sim	Não	Não	Não
M_10	31	Não	Não	Não	Não
M_11	32	Não	Não	Não	Não
M_12	47	Não	Não	Não	Não
M_13	52	Não	Não	Não	Não
M_14	56	Não	Não	Não	Não
M_15	61	Sim	Anestesia Geral (\approx 6 meses)	Não	Não

Por conveniência, foram incluídos todos os sujeitos que se disponibilizaram para participar no estudo. Em comum, todos os envolvidos partilham o facto de serem nativos da língua portuguesa e de não apresentarem queixas de disfonia, isto é, disporem de vozes que à partida são consideradas “normais”.

No entanto, para considerar a amostra como representativa de uma população de controlo, deveriam ser eliminados os sujeitos F_13 e F_14, que já foram diagnosticados com nódulos na tiroide, o sujeito M_15 que foi submetido a uma anestesia geral e é fumador, os sujeitos M_4, F_3 e F_4 que são fumadores e possivelmente os sujeitos F_8 e F_10, que são asmáticos.

3.2 Protocolo de Avaliação da Qualidade Vocal

Para definir um protocolo de avaliação apropriado ao âmbito da dissertação, foram tidos em consideração vários fatores. A medição da qualidade da voz a partir de vogais sustentadas apresenta vários argumentos a favor, como já foi referido anteriormente. Em particular, serem relativamente invariantes no tempo; livres de influências do contexto fonético e, portanto, não serem afetadas por condicionantes como a entoação, stress ou velocidade de fala; serem fáceis de obter, de produzir e de analisar; serem facilmente controladas; e serem menos afetadas pelo dialeto do orador ou do ouvinte, do que a fala contínua [68].

A inclusão de duas vogais com diferenças ao nível da produção, como as vogais /a/ e /i/, justifica-se dado que cada uma pode revelar alterações específicas da qualidade vocal. Embora a fonte sonora seja a mesma, certas alterações no trato vocal podem influenciar determinadas características da fonação. A vogal /a/ é uma vogal relaxada, enquanto que a vogal /i/ é uma vogal tensa. Além disso, a vogal /i/ apresenta como vantagem o facto de ser muitas vezes utilizada durante a videoestroboscopia, o que facilita a comparação acústica com os resultados da imagem laríngea [34].

A avaliação acústica deve incluir múltiplas repetições da mesma vogal, por forma a considerar a variabilidade individual e a estabelecer um desempenho estável. O número de repetições deve representar adequadamente o comportamento da fala, sendo que, ocasionalmente, o primeiro ensaio de um indivíduo pode afastar-se do seu registo de produção habitual devido à falta de familiaridade com a tarefa [40].

Com base no que é proposto na literatura, foi elaborado um protocolo de avaliação da qualidade vocal para a análise acústica da voz. Os parâmetros acústicos são avaliados com recurso às seguintes tarefas:

1. Produzir a vogal /a/ de modo sustentado e estável, mantendo níveis de intensidade confortáveis e tentando aproximar o mais possível à voz habitual. Esta tarefa deve ser executada 2 vezes com durações de 3 e 5 segundos, respectivamente. Em seguida, repetir a tarefa substituindo a vogal /a/ pela vogal /i/.
2. Ler as seguintes frases, uma de cada vez, como se estivesse a conversar:
 - “A asa do avião andava avariada.”
 - “O Tiago comeu quatro pêras.”
3. Por fim, repetir a primeira tarefa.

É importante salientar que foram incluídas duas frases após a duas primeiras repetições das vogais sustentadas, para que de certa forma o sujeito se apercebesse de como as pronuncia em contexto de conversa. Em seguida, volta se a repetir a primeira tarefa com a qual o sujeito já está familiarizado. As frases foram retiradas de uma adaptação do CAPE-V para o português europeu [69]. Contudo, não são analisadas acusticamente no presente estudo, apenas têm o propósito de desviar a atenção do sujeito à tarefa principal e involuntariamente torná-la menos forçada.

3.3 Gravação

Antes de iniciar as gravações foi solicitado aos indivíduos que permanecessem sentados confortavelmente a uma distância de aproximadamente 10/15 cm do microfone. Quando se encontravam corretamente posicionados, dava-se início à gravação, eram realizadas as tarefas do protocolo e a gravação só terminava quando todas as tarefas tivessem sido executadas.

3.3.1 Equipamento de Gravação

Para o processo de gravação foi utilizado um gravador portátil de alta qualidade, o *Zoom H4n Handy Recorder* representado na Figura 3.1. Este está equipado com 2 microfones dispostos num padrão de captação XY, cujos ângulos de configuração (90° e 120°) podem ser ajustados para direcionar a captação de áudio. Neste caso, foi escolhido o ângulo de 90° por ser ideal na captação de sinais que são produzidos por um único sujeito posicionado diretamente de frente para o microfone. O gravador foi programado para modo *stereo*, de forma a obter sinais acústicos de alta qualidade. O nível da gravação foi definido para 75, uma vez que para volumes de gravação muito elevados o próprio aparelho tem tendência a introduzir ruído no

signal. Os sinais acústicos foram armazenados em formato *.wav* com uma frequência de amostragem de 44.1 kHz e resolução de 24 bits. A principal vantagem deste tipo de formato é que os ficheiros não são comprimidos, pelo que contém todo o sinal sem que nenhuma informação seja removida/compactada e por isso a qualidade das gravações é superior [70].



Figura 3.1: Microfone *Zoom H4n Handy Recorder*, usado na captação dos sinais acústicos da população da amostra [71].

Em seguida, são sumariadas as especificações e as limitações envolvidas no processo de gravação dos sinais acústicos:

- **Especificações do Equipamento de Gravação**

- Modo de Gravação: *Stereo*;
- Microfones incorporados: Condensador Unidirecional (Ganho: +7_+47 dB);
- Ângulo de configuração dos microfones: 90°;
- Nível de entrada do sinal: 75;
- Formato do ficheiro: *.wav*;
- Frequência de Amostragem: 44,1 kHz;
- Resolução: 24 bits.

- **Especificações do Registo de Gravação**

- Ambiente: Silencioso;

- Posicionamento: Sujeito sentado a 10/15 cm do microfone;
- Duração: Tempo necessário para realizar todas as tarefas descritas no protocolo.

- **Principais Limitações**

- Todas as gravações foram realizadas em ambientes silenciosos, no entanto, não foram todas realizadas no mesmo espaço e nenhum deles era isolado acusticamente;
- A distância entre os sujeitos e o microfone foi delimitada através do uso de uma régua, não havendo um suporte fixo para o microfone e não sendo possível garantir que o seu posicionamento tenha sido exatamente o mesmo entre sujeitos.

3.4 Métodos de Extração dos Parâmetros

Para caracterizar a qualidade vocal e o padrão vibratório das pregas foi desenvolvida uma ferramenta computacional em MATLAB que permite extrair o conjunto de parâmetros: *jitter* relativo, absoluto, RAP e PPQ5; *shimmer* relativo, APQ3, APQ5 e APQ11; e HNR. O seu desempenho foi avaliado através da comparação com o *software* Praat, que é usado para a mesma finalidade. De forma a garantir a validade da comparação, o algoritmo desenvolvido segue todos os princípios propostos pelo Praat.

3.4.1 Extração dos parâmetros no Praat

O primeiro passo para a análise acústica dos sinais consistiu na divisão da gravação de cada sujeito num total de 8 ficheiros, sendo 4 correspondentes às vogais /a/ sustentadas e os outros 4 às vogais /i/ sustentadas. Os ficheiros foram carregados no programa Praat e analisados como *mono*.

Segundo Martens *et al.* [50] é recomendado que a avaliação de vogais sustentadas seja realizada numa porção relativamente estável do sinal. Para tal, os primeiros e últimos 250 ms devem ser removidos, eliminando o *onset* e *offset* do sinal. Tendo isto em conta, todos os ficheiros foram analisados na sua porção mais estável, numa janela temporal de 2 segundos.

Solomon *et al.* [34] afirma que os parâmetros acústicos diferem significativamente entre as vogais /a/ e /i/, indicando que as avaliações devem especificar a vogal em

estudo e os resultados não devem ser comparados entre vogais. Assim, para cada ficheiro foram extraídos os respetivos parâmetros e elaborada a média dos mesmos para as 4 vogais iguais de um mesmo sujeito. Este processo repetiu-se no algoritmo desenvolvido em MATLAB, sendo que cada um dos ficheiros foi analisado no mesmo intervalo de tempo em ambos *softwares*.

3.4.2 Algoritmo desenvolvido no MATLAB

O algoritmo inicia-se com a leitura do sinal de áudio que se pretende analisar e a informação é apresentada numa matriz de 2 colunas, correspondentes aos canais direito e esquerdo do gravador, resultantes da gravação em modo *stereo*. De forma a analisar o sinal como um todo, foi criada uma versão *mono* fazendo a média dos dois canais do sinal. Em seguida, o tempo foi delimitado apenas para o segmento que se pretendia analisar, sendo igual ao selecionado no decorrer da análise no Praat. O ponto de partida da análise acústica consiste na determinação da frequência fundamental de um dado sinal. Para esse efeito, inicialmente recorreu-se à função *peakdet* [72] que consistia na determinação dos máximos e mínimos do sinal, com base num *threshold* definido a partir do qual estes eram contabilizados (Figura 3.2).

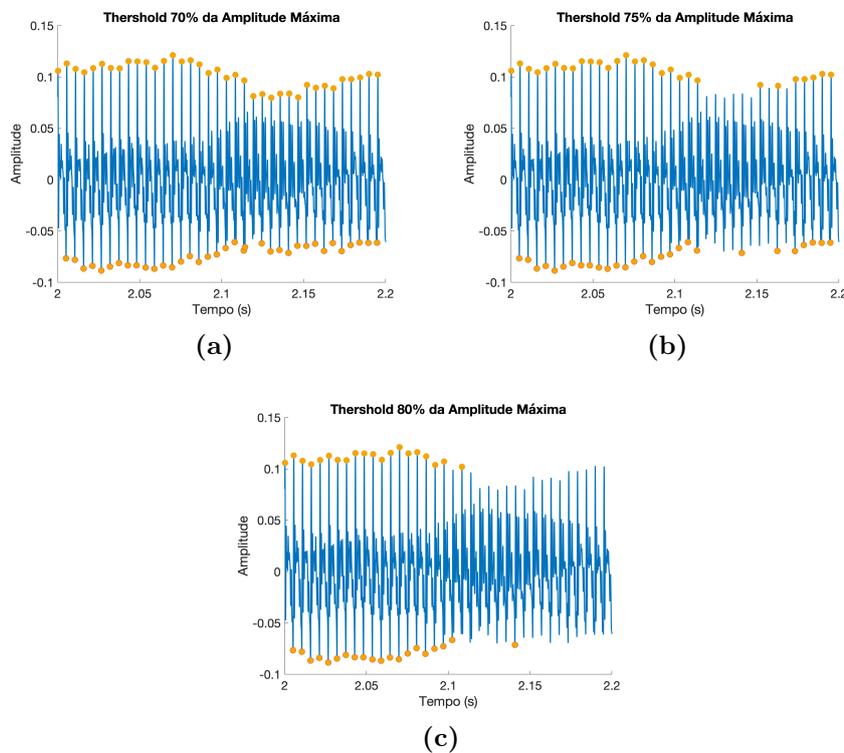


Figura 3.2: Picos determinados através do algoritmo *peakdet* com diferentes thresholds, no intervalo de [2-2.2] s. (a) Threshold 70% da Amplitude Máxima (b) Threshold 75% da Amplitude Máxima (c) Threshold 80% da Amplitude Máxima

Definiram-se *thresholds* específicos consoante a vogal produzida e o género do orador. Contudo, após várias tentativas com diferentes *thresholds* chegou-se à conclusão de que esta forma não era viável, pois a diversidade de sinais continuava a ser muito grande e o algoritmo não era capaz de identificar com sucesso os picos em todos os períodos glóticos, como ilustrado na Figura 3.2.

Para solucionar o problema substituiu-se a função anterior pela função *pitch* do MATLAB, que retorna uma estimativa da frequência fundamental ao longo do segmento em análise. Uma vez que todos os sinais analisados consistem na produção de vogais sustentadas, não se verificam variações abruptas da frequência fundamental ao longo dos mesmos e por isso foi possível determinar a duração dos períodos glóticos através do inverso da frequência fundamental dada pela função *pitch*. Posteriormente, procedeu-se à determinação dos picos máximos e mínimos do sinal. Para tal, foi implementado um ciclo que percorre a totalidade do sinal em segmentos com duração igual à de um período glótico e para cada segmento, os picos são determinados como o valor máximo e mínimo. Através deste processo foi possível determinar com sucesso os picos para todos os sinais contidos na amostra. A título de exemplo são apresentados 2 gráficos, nas Figuras 3.3 (a) e 3.3 (b), correspondentes às vogais sustentadas /a/ e /i/ produzidas por sujeitos do género feminino, e outros 2 gráficos, nas Figuras 3.4 (a) e 3.4 (b), correspondentes à produção das mesmas vogais por sujeitos do género masculino.

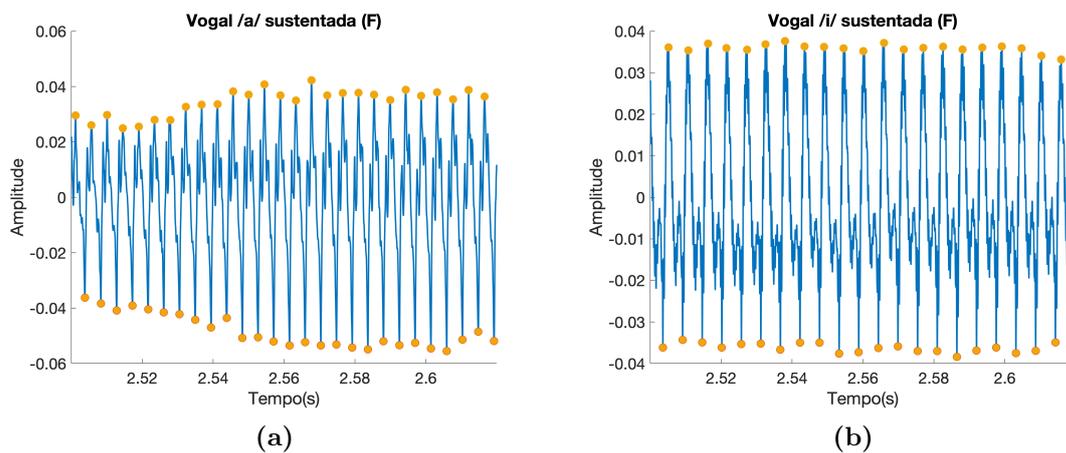


Figura 3.3: Produção de 2 vogais sustentadas por sujeitos do género feminino, no intervalo [2.5-2.62] s. (a) Produção da vogal /a/ sustentada. (b) Produção da vogal /i/ sustentada.

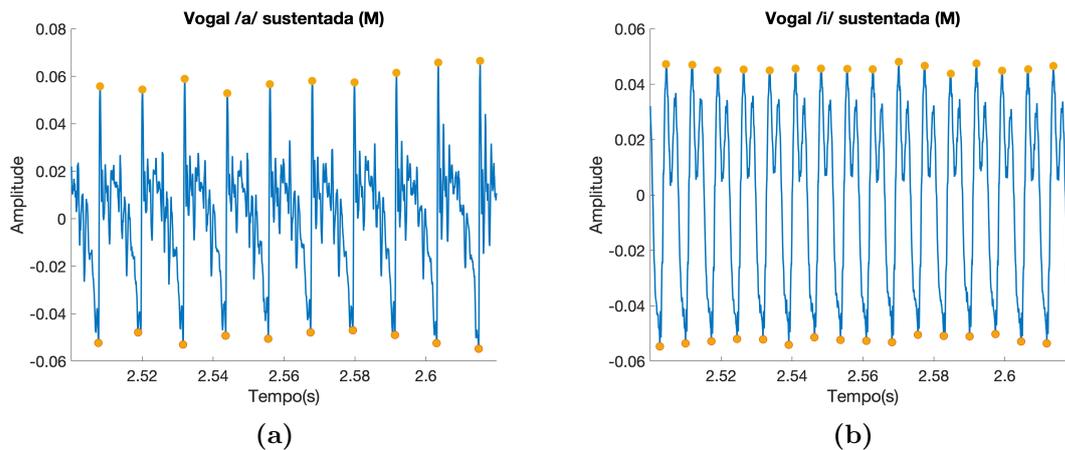


Figura 3.4: Produção de 2 vogais sustentadas por sujeitos do género masculino, no intervalo [2.5-2.62] s. **(a)** Produção da vogal /a/ sustentada. **(b)** Produção da vogal /i/ sustentada.

Visto que o objetivo consistia numa análise comparativa com o *software* Praat, ao longo do algoritmo foram implementadas as mesmas condições que esse *software* impõe para a determinação de cada parâmetro.

Frequência Fundamental: Com os picos detetados, calcularam-se novamente os períodos glóticos, desta vez recorrendo à diferença entre os picos mínimos consecutivos. A escolha dos mínimos prende-se com o facto de serem os que o Praat usa para determinar este parâmetro. Em seguida, voltou a calcular-se a frequência fundamental através destes períodos glóticos, como descrito na equação 2.1.

Jitter: Antes de iniciar o cálculo de *jitter* foram removidos todos os períodos inferiores a 0.0001 e superiores a 0.02, tal como indica no Praat que deve ser feito [54]. Para a determinação desta medida de perturbação, teve-se em consideração que representa a variação entre períodos glóticos e foram determinadas as diferenças entre os períodos consecutivos. De acordo com o algoritmo do Praat, quando a razão entre dois períodos consecutivos é superior a 1.3, a diferença entre os períodos deve ser descartada nos cálculos seguintes [54]. Posto isto, foram implementadas as equações 2.2 a 2.5 e determinadas as 4 medidas de *jitter*.

Shimmer: Para o cálculo do *shimmer*, com os valores dos máximos e mínimos determinados anteriormente procedeu-se à determinação das amplitudes correspondentes a cada período glótico. Dado que esta medida de perturbação reflete a variação entre amplitudes, em seguida, foram calculadas as diferenças entre as amplitudes sucessivas do sinal. Neste caso, o Praat também impõe um fator de correção, sendo que se a razão entre amplitudes consecutivas for superior a 1.6, o par de amplitudes

é ignorado nos cálculos seguintes [54]. Implementando as equações 2.6 a 2.9 foram determinadas as 4 medidas de *shimmer*.

HNR: Por fim, para determinar o HNR foi efetuada a deteção da periodicidade acústica com base no método de autocorrelação normalizada proposto por Boersma [66]. Conforme é descrito no Praat, este parâmetro deve ser analisado em segmentos e não na totalidade do sinal, tendo em conta que janelas de 6 períodos glóticos permitem detetar de forma fiável valores de HNR até cerca de 60 dB, mas o algoritmo torna-se mais sensível a mudanças dinâmicas no sinal [54]. Como se pretendem determinar os parâmetros em sinais acústicos de vogais sustentadas, não ocorrem mudanças dinâmicas significativas ao longo dos mesmos, pelo que neste algoritmo optou-se por determinar os valores de HNR em janelas sucessivas de 6 períodos glóticos. Uma vez que a duração dos períodos glóticos varia consoante o sinal que está a ser analisado, para determinar de forma mais precisa a duração de um intervalo de HNR recorreu-se novamente ao período glótico determinado no início do algoritmo e multiplicou-se por 6. A determinação do HNR de cada segmento foi realizada com auxílio da função *xcorr* com a especificação *'normalized'*. Para encontrar a relação HNR é necessário medir a energia do primeiro máximo local da autocorrelação e aplicar a equação 2.10. Este processo foi repetido sucessivamente utilizando sempre como referência o comprimento de 6 períodos glóticos. O valor final de HNR corresponde à média do HNR de todos os segmentos sucessivos.

3.4.3 Interface Gráfica

Com o objetivo de agilizar o processo de análise foi elaborada uma *Graphical User Interface* (GUI) em MATLAB, adaptada a partir do código de [73], de modo a satisfazer as necessidades da dissertação. A interface de análise acústica é ilustrada na Figura 3.5. Esta permite selecionar qualquer ficheiro *.wav* que se pretenda analisar, sendo que o seu gráfico é imediatamente exibido na secção “Sinal Acústico”. Para estabelecer o tempo do segmento a analisar, basta introduzir o seu valor nos espaços disponíveis para tal e através do botão “Segmentar o sinal” o seu gráfico surge na secção “Segmento selecionado”. Adicionalmente, também é possível ouvir o respetivo segmento premindo o botão “Play”. Por fim, procede-se à identificação das características do sinal relativas ao género e ao comportamento de fala e através do botão “Analisar segmento” são apresentados os 4 parâmetros mais comuns da análise acústica. Os limiares de patologia associados a cada um dos parâmetros são apresentados como informação de referência, facilitando o processo de classificação da qualidade vocal.

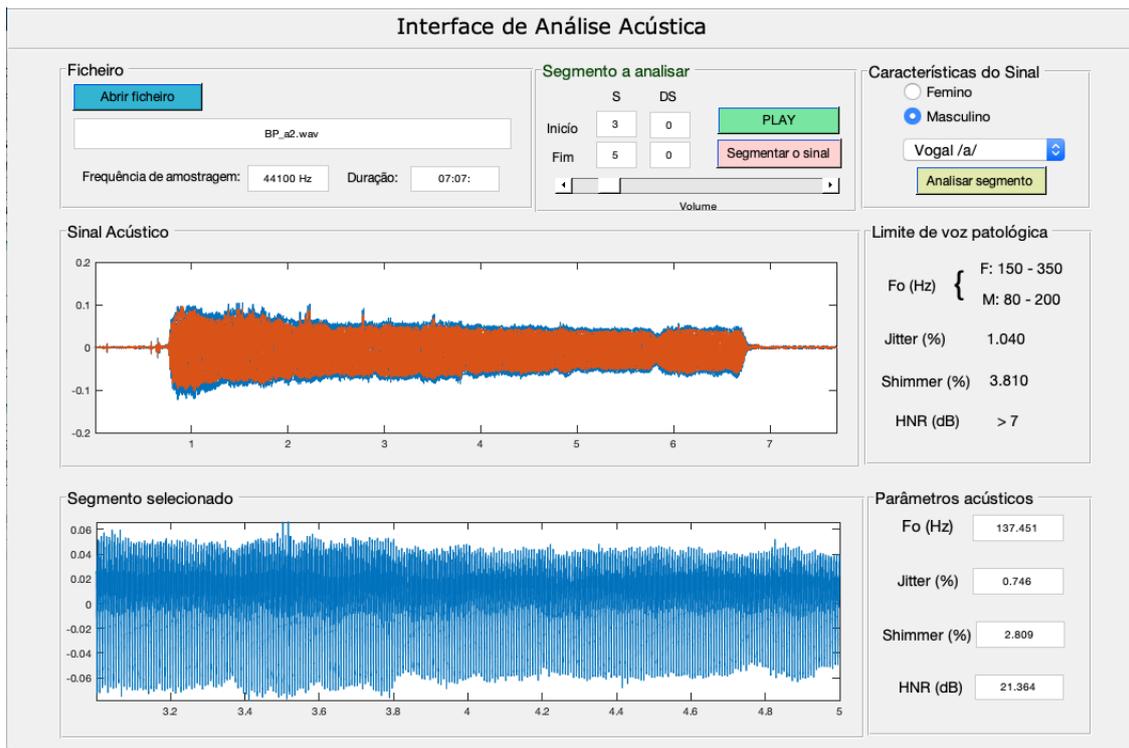


Figura 3.5: Interface de Análise Acústica usada para determinar automaticamente os 4 principais parâmetros acústicos de qualquer ficheiro *.wav*, no segmento de tempo que se pretenda analisar.

Resultados

Nesta secção é efetuada uma análise comparativa entre o *software* Praat e o algoritmo desenvolvido no MATLAB, com o intuito de provar a sua eficácia na determinação dos vários parâmetros acústicos. De forma a validar a precisão do algoritmo na deteção dos picos, é apresentada uma análise preliminar da mesma na Secção 4.1. Por fim, é realizada uma breve avaliação da relação entre as características da população e os valores dos respetivos parâmetros.

4.1 Precisão do Algoritmo na Determinação de Picos

A fim de verificar a precisão do algoritmo na deteção dos picos, foi realizada uma experiência com base na comparação entre os valores obtidos pelo algoritmo e os obtidos com o programa Praat. Para tal, foi selecionado um dos áudios ao acaso e analisado no intervalo de 2.1 a 2.16 segundos, em ambas as plataformas de processamento (MATLAB e Praat). Nas Figuras 4.1 (a) e 4.1 (b), encontram-se assinalados os picos detetados pelo MATLAB e pelo Praat, respetivamente. Na Tabela 4.1 encontram-se registados os erros relativos que cada um dos picos do MATLAB apresenta quando comparado com o respetivo pico do Praat. Analisando a Tabela 4.1 conclui-se que os valores de ambos os algoritmos são semelhantes, porém a maioria apresenta modificações ao nível da 4^a casa decimal. Para determinar a influência que essas pequenas diferenças introduzem nos resultados finais, foram calculadas as medidas de *jitter* para cada uma das situações recorrendo ao algoritmo desenvolvido. Os resultados obtidos encontram-se na Tabela 4.2.

4. Resultados

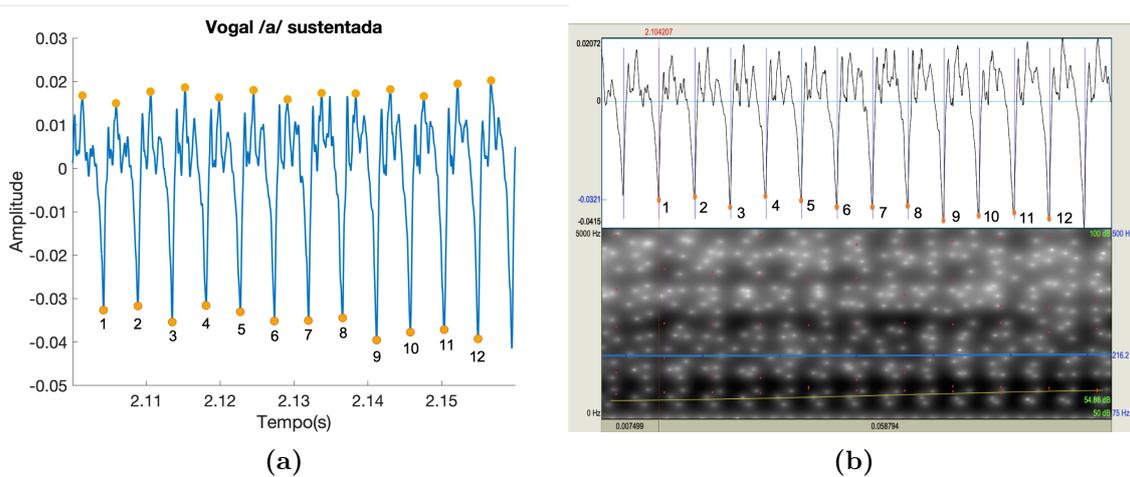


Figura 4.1: Detecção de Picos de um sinal acústico no intervalo de tempo de [2.1-2.16] s. (a) Picos detetados pelo MATLAB. (b) Picos detetados pelo Praat.

Tabela 4.1: Erro relativo dos picos do MATLAB em comparação com os picos do Praat.

Nº do Pico	Erro Relativo (%)
1	8.06E-06
2	1.27E-04
3	1.64E-04
4	1.55E-03
5	1.94E-04
6	6.16E-04
7	5.45E-04
8	3.39E-04
9	5.58E-04
10	1.69E-03
11	1.71E-03
12	2.03E-03

Tabela 4.2: Comparação entre as 4 medidas de *jitter*, determinadas com os picos de 1 a 12, do MATLAB e do Praat.

	Praat	MATLAB
<i>Jitta</i> (μs)	20.078	29.478
<i>Jitter</i> (%)	0.4350	0.6392
RAP (%)	0.2937	0.4371
PPQ5 (%)	0.4068	0.5901

A partir da tabela 4.2, verifica-se que os resultados determinados com os picos do Praat foram ligeiramente inferiores aos obtidos com os picos do MATLAB. É de esperar que a diferença entre os resultados dos *softwares* fique ainda mais acentuada quando se analisam segmentos de tempo maiores, pois aumenta o número de picos. No entanto, isso não invalida a correta deteção dos mesmos.

Nesta secção não foi realizada uma comparação entre as medidas de *shimmer*, mas de modo análogo podem ocorrer variações ao nível das casas decimais que influenciem o resultado final.

4.2 Análise de Resultados

O algoritmo foi testado nos sinais acústicos de 15 sujeitos do género feminino e 15 do género masculino, cuja caracterização se encontra nas Tabelas 3.1 e 3.2, respetivamente. Para analisar os sinais procedeu-se à sua segmentação, sendo que em cada sinal foram considerados os seus 2 segundos mais estáveis. É de salientar que foram seleccionados exatamente os mesmos intervalos de tempo no algoritmo que se pretende testar e no Praat, para tornar a comparação o mais fiável possível. Tendo em conta as considerações apresentadas na secção 3.4.1, as vogais são analisadas separadamente e é realizada a média dos parâmetros dos 4 sinais correspondentes à mesma vogal num mesmo sujeito. Além disso, os resultados foram também divididos entre géneros, uma vez que as características das pregas vocais diferem entres eles.

Para esta análise utilizaram-se diagramas *boxplot*, que apresentam a distribuição métrica dos parâmetros nestes conjuntos de vozes e facilitam a visualização da semelhança entres os valores obtidos pelas diferentes plataformas de processamento acústico.

4.2.1 Vogais sustentadas nos sujeitos do género Feminino

Na Figuras 4.2 e 4.4 são apresentadas as distribuições dos principais parâmetros da análise acústica, F_0 , *jitter*, *shimmer* e HNR, obtidos para a amostra de população feminina na produção das vogais /a/ e /i/, respetivamente. Do mesmo modo, nas Figuras 4.3 e 4.5 são exibidas as distribuições das respetivas medidas de *jitter*, RAP e PPQ5, e de *shimmer*, APQ3, APQ5 e APQ11.

Vogal /a/ sustentada

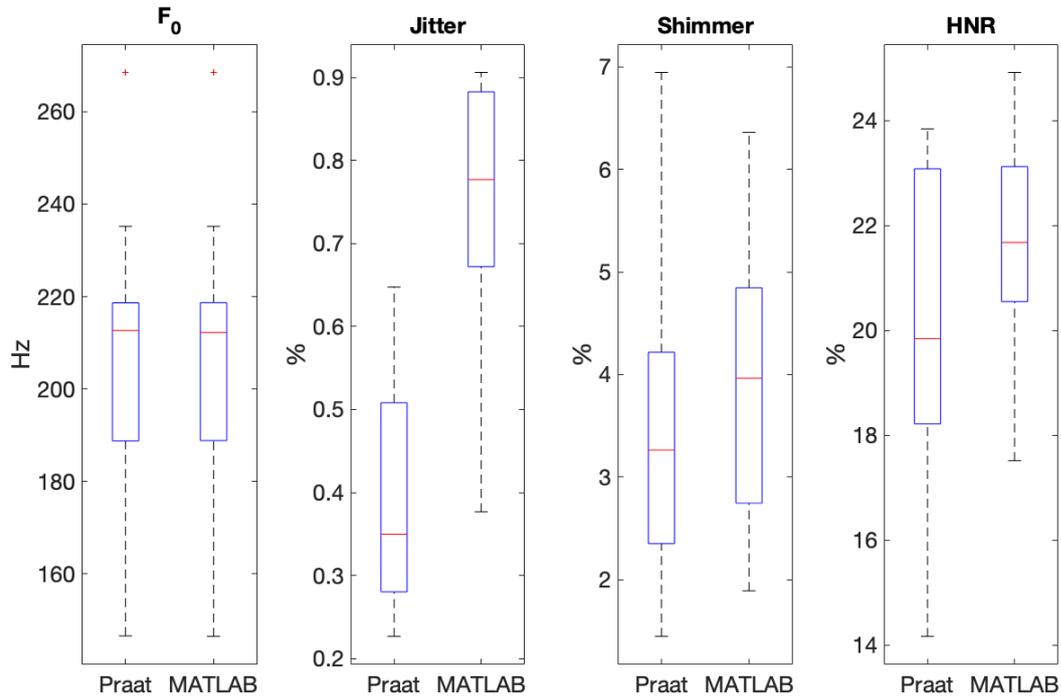


Figura 4.2: *Boxplots* da distribuição métrica dos parâmetros acústicos principais da vogal /a/ na amostra feminina.

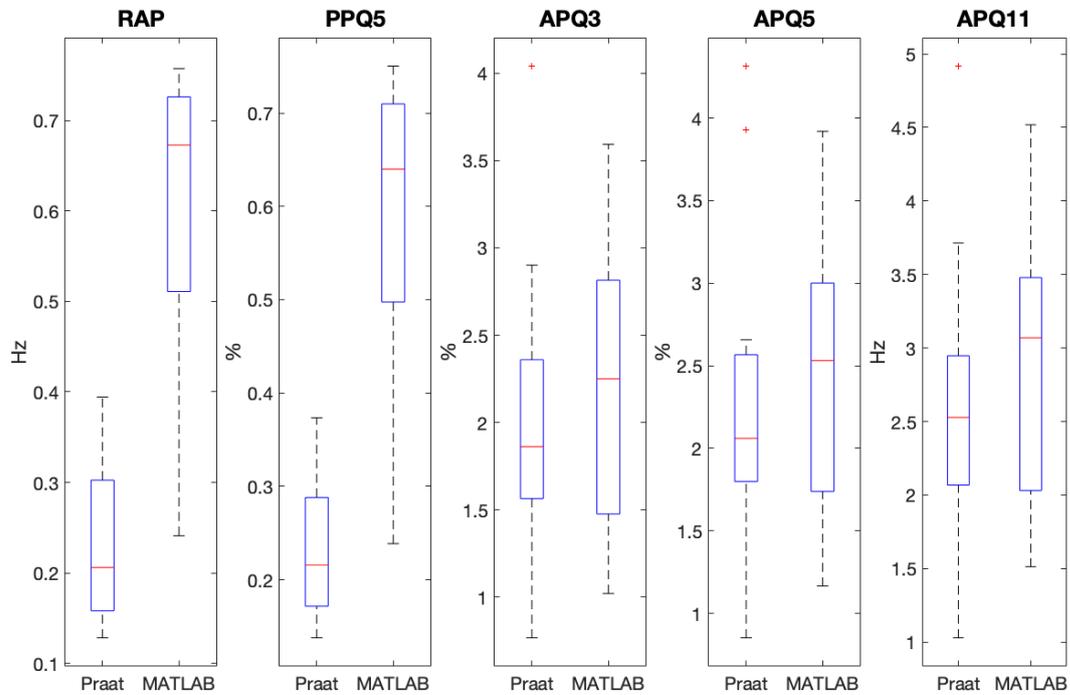


Figura 4.3: *Boxplots* da distribuição métrica das medidas de *jitter* e de *shimmer* da vogal /a/ na amostra feminina.

Analisando a distribuição dos parâmetros da vogal /a/, presentes na Figura 4.2, verifica-se que a F_0 é parâmetro com melhor correlação entre os *softwares*. Apesar de se observarem ligeiras diferenças no *shimmer* e no HNR, estas não são consideradas estatisticamente significativas. As principais discrepâncias ocorrem ao nível do *jitter*, uma vez que os quartis do Praat e do MATLAB não partilham a mesma linha horizontal. Os valores determinados pelo algoritmo são superiores, mas não ultrapassam o limite de patologia e estas desigualdades são corroboradas pelo exemplo apresentado na secção 4.1. No que diz respeito à Figura 4.3, observa-se que os resultados são coerentes com os da Figura anterior (4.2), uma vez que apresenta discrepâncias da mesma ordem nas medidas de *jitter* e correlações razoáveis nas medidas de *shimmer*.

Vogal /i/ sustentada

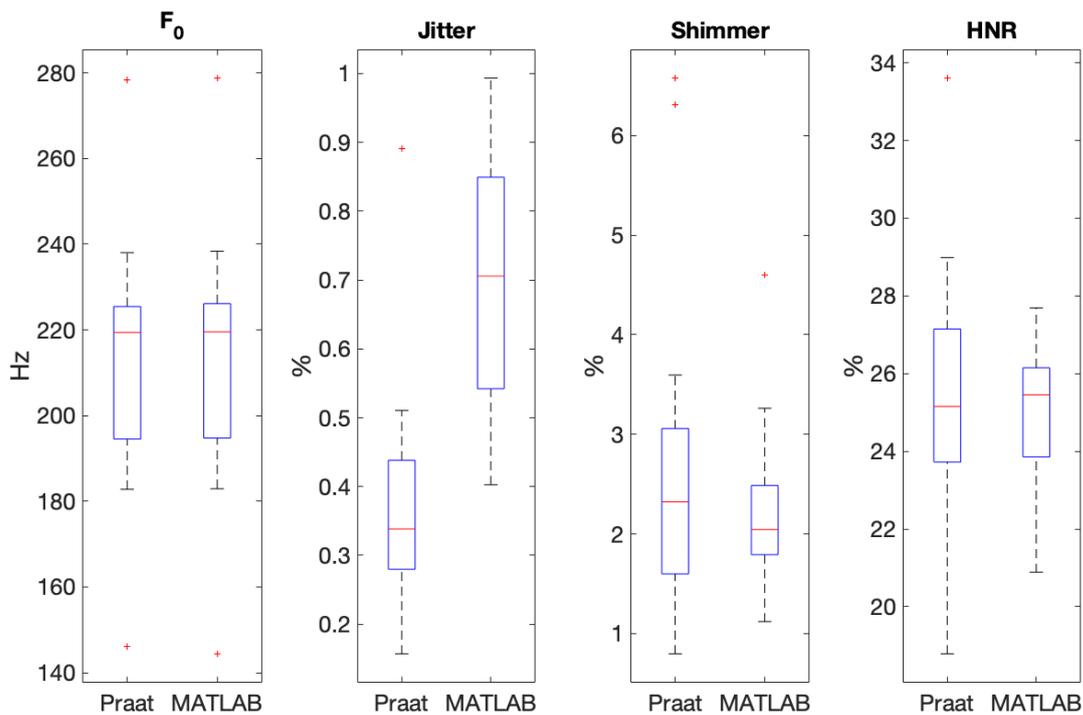


Figura 4.4: *Boxplots* da distribuição métrica dos parâmetros acústicos principais da vogal /i/ na amostra feminina.

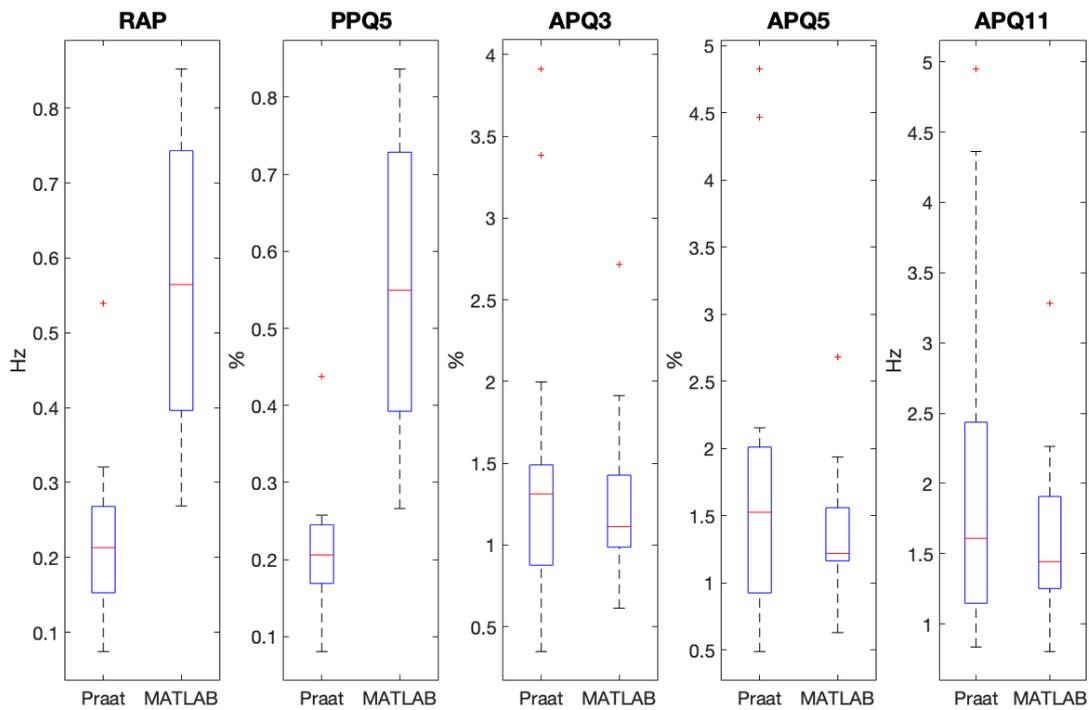


Figura 4.5: *Boxplots* da distribuição métrica das medidas de *jitter* e de *shimmer* da vogal /i/ na amostra feminina.

Relativamente às Figuras 4.4 e 4.5, de um modo geral a distribuição dos parâmetros da vogal /i/ apresenta um comportamento semelhante ao da distribuição dos parâmetros da vogal /a/. Contudo, na produção da vogal /i/ os valores de *shimmer* e de HNR apresentam uma correlação superior entre os *softwares*. Os valores de *jitter* e das suas medidas permanecem díspares entre o algoritmo e o Praat.

4.2.2 Vogais sustentadas nos sujeitos do género Masculino

As Figuras 4.6 e 4.8 exibem a distribuição dos principais parâmetros da análise acústica obtidos para a amostra de população masculina na produção das vogais /a/ e /i/, respetivamente. As Figuras 4.7 e 4.9 apresentam as distribuições das respetivas medidas de *jitter* e *shimmer*, nessas mesmas vogais.

Vogal /a/ sustentada

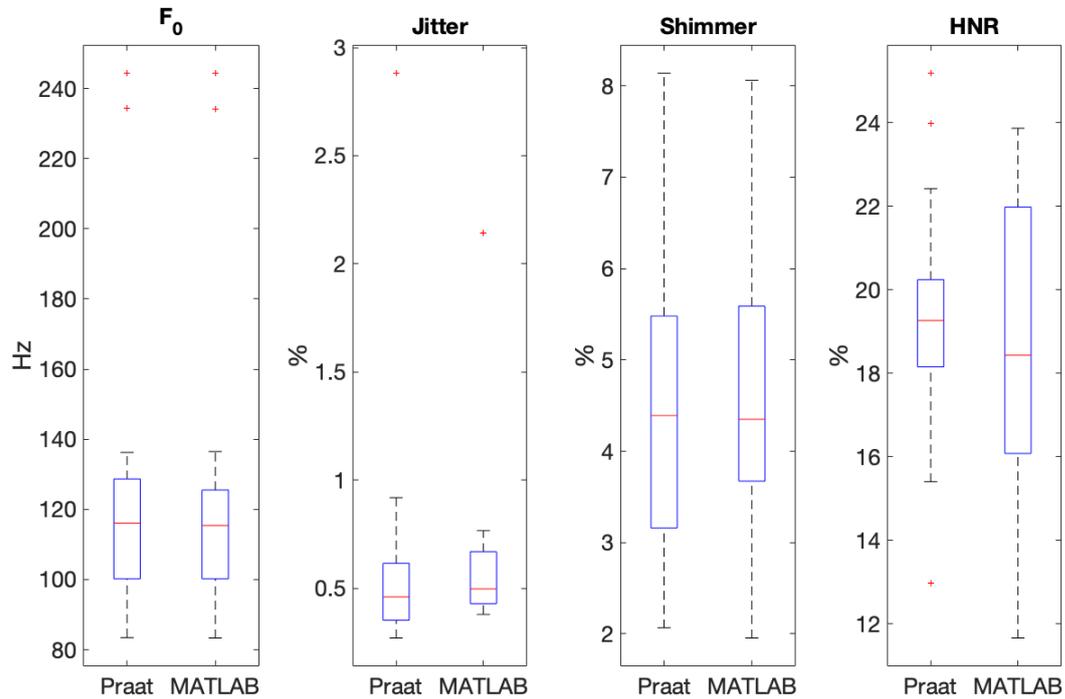


Figura 4.6: *Boxplots* da distribuição métrica dos parâmetros acústicos principais da vogal /a/ na amostra masculina.

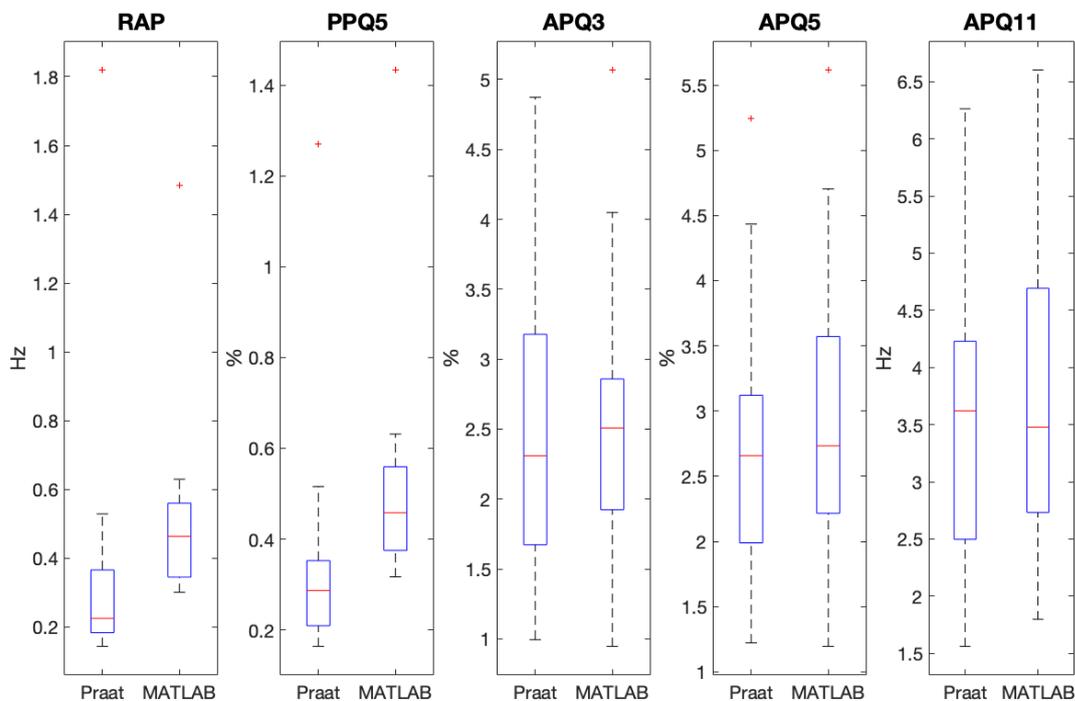


Figura 4.7: *Boxplots* da distribuição métrica das medidas de *jitter* e de *shimmer* da vogal /a/ na amostra masculina.

Observando as Figuras 4.6 e 4.7, conclui-se que na produção da vogal /a/ por sujeitos do género masculino, o parâmetro com a melhor correspondência entre o Praat e o MATLAB é a F_0 , tal como na amostra feminina. Contrariamente ao que sucede na amostra feminina as discrepâncias entre os valores de *jitter* não são tão acentuadas, apenas são ligeiramente evidentes nas medidas RAP e PPQ5. O parâmetro *shimmer*, bem como, as suas respetivas medidas APQ3, APQ5, APQ11, e o parâmetro HNR apresentam uma correlação superior comparativamente à correlação dos mesmos na produção da vogal /a/ na amostra feminina.

Vogal /i/ sustentada

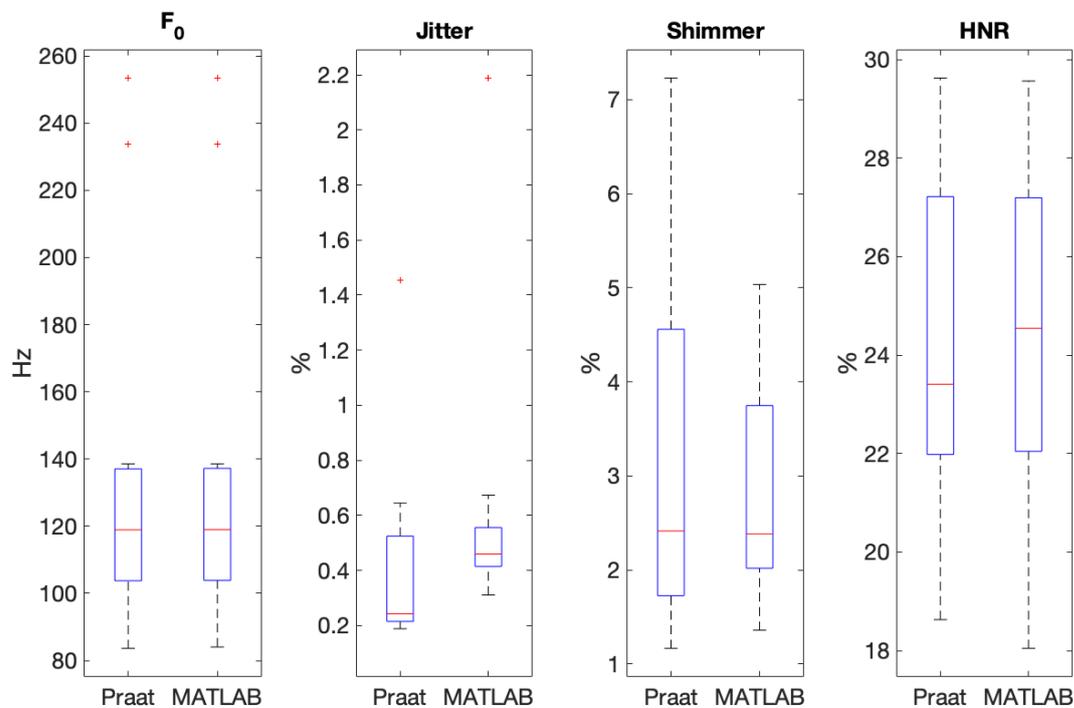


Figura 4.8: *Boxplots* da distribuição métrica dos parâmetros acústicos principais da vogal /i/ na amostra masculina.

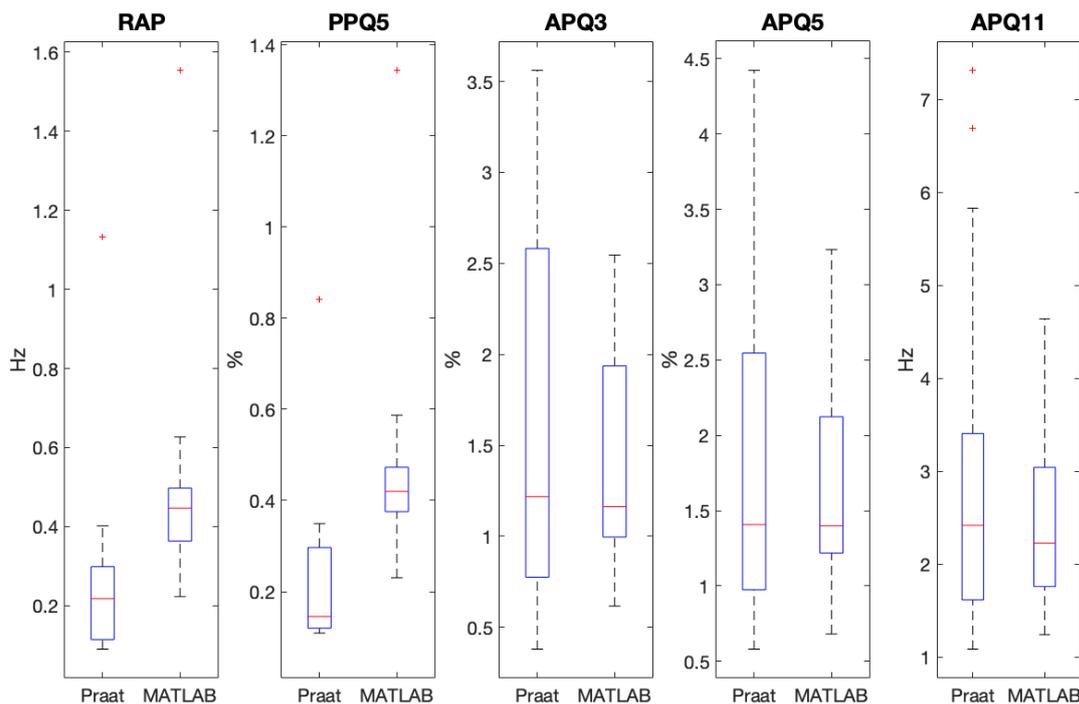


Figura 4.9: *Boxplots* da distribuição métrica das medidas de *jitter* e de *shimmer* da vogal /i/ na amostra masculina.

Através das Figuras 4.8 e 4.9, verifica-se que, a distribuição dos parâmetros da vogal /i/ na população masculina apresenta um desempenho semelhante à distribuição dos parâmetros da vogal /a/ nessa mesma população. Neste caso, a correlação de *jitter* e das suas respetivas medidas é ligeiramente mais fraca do que no caso da vogal /a/. Os restantes parâmetros exibem correlações satisfatórias, sendo que, mais uma vez, a melhor é a de F_0 .

4.2.3 Análise Geral dos Resultados da Amostra

De acordo com Mehta *et.al.* [74], a validade e fiabilidade das medidas acústicas utilizadas na prática clínica depende inerentemente da confiança depositada na determinação exata da Frequência Fundamental. Analisando os *boxplots* das secções 4.2.1 e 4.2.2, é possível concluir que o algoritmo desenvolvido é eficaz na determinação dos parâmetros em causa, dado que a distribuição dos valores de F_0 para o MATLAB é muito semelhante à distribuição apresentada para o mesmo parâmetro no Praat.

As medianas correspondentes aos *boxplots* que representam a distribuição dos parâmetros para ambas as vogais e para ambos os géneros, encontram-se resumidas na Tabela 4.3.

Tabela 4.3: Medianas resultantes da distribuição métrica dos parâmetros apresentados nos *boxplots* das vogais /a/ e /i/ para as amostras de população feminina (F) e masculina (M).

	Vogal /a/ F		Vogal /i/ F		Vogal /a/ M		Vogal /i/ M	
	Praat	Matlab	Praat	Matlab	Praat	Matlab	Praat	Matlab
F0(Hz)	212.6680	212.2407	219.4035	219.5288	116.1052	115.4098	118.9213	118.9793
Jitter(%)	0.3498	0.7769	0.3385	0.7057	0.4602	0.4971	0.2425	0.4591
Shimmer(%)	3.2630	3.9634	2.3222	2.0460	4.3905	4.3488	2.4152	2.3842
HNR(dB)	19.8445	21.6775	25.1545	25.4538	19.2602	18.4306	23.4097	24.5450
RAP(%)	0.2060	0.6729	0.2130	0.5645	0.2250	0.4635	0.2177	0.4470
PPQ5(%)	0.2157	0.6398	0.2060	0.5494	0.2873	0.4583	0.1455	0.4198
APQ3(%)	1.8620	2.2506	1.3125	1.1120	2.3095	2.5076	1.2183	1.1630
APQ5(%)	2.0605	2.5310	1.5245	1.2173	2.6580	2.7330	1.4083	1.3995
APQ11(%)	2.5285	3.0699	1.6093	1.4438	3.6215	3.4789	2.4190	2.2279

A partir da análise da Tabela 4.3, destacam-se as seguintes considerações:

- As pequenas diferenças que se observam entre o algoritmo desenvolvido e o Praat não são consideradas significativas e podem ser justificadas por alterações ao nível das casas decimais, tal como é explícito na secção 4.1.
- A maior parte dos resultados obtidos com o algoritmo estão dentro da gama de valores correspondentes a vozes saudáveis. O *shimmer* e o APQ11 são os únicos parâmetros que não se encontram dentro dos intervalos de normalidade, no caso da produção da vogal /a/ por ambos os géneros. No entanto, para a amostra masculina no Praat, estes parâmetros também apresentam valores acima dos limites patológicos. Tendo em conta que tanto o algoritmo como o Praat apresentam valores relativamente superiores ao expectável, julga-se que estes podem ter sido aumentados artificialmente por distorções introduzidas pelo instrumento de gravação.
- A Frequência Fundamental é o parâmetro com maior grau de semelhança em todos os casos. Tal como previsto na literatura, a F_0 é superior na amostra feminina e inferior na amostra masculina.

- O *jitter* e as suas respetivas medidas não são comparados entre géneros, por se tratarem de medidas relativas. No entanto, apresentam diferenças acústicas entre as vogais. De um modo geral, é possível constatar que para ambos os géneros são atribuídos valores superiores de *jitter* e *shimmer* à produção da vogal relaxada (/a/) e valores inferiores à produção da vogal tensa (/i/), mas podem haver exceções, particularmente ao nível do *jitter*.
- Relativamente ao HNR, verifica-se que a vogal /a/ tende a apresentar valores inferiores quando comparada com vogal /i/. Destaca-se também que a população feminina apresenta valores de HNR mais elevados que a população masculina, tal como previsto na literatura.

4.2.4 Análise dos Sujeitos com Restrições Vocais

Como referido na secção 3.1, para construir a amostra não foram impostos critérios de exclusão e foram incluídos sujeitos com restrições (hábitos tabágicos, nódulos na tiroide, asma, histórico de entubação) que podem influenciar a qualidade vocal. Com o objetivo de constatar se existia alguma restrição significativa, procedeu-se à análise da relação entre a história clínica dos sujeitos e os parâmetros com valores atípicos.

Na amostra feminina, verificou-se que o sujeito F_13, diagnosticado com nódulos na tiroide, foi o que revelou parâmetros com maior afastamento relativamente à mediana da amostra a que pertence. Este sujeito apresentou:

- Os valores de F_0 mais baixos para ambas as vogais;
- Os valores de *jitter*, RAP e PPQ5 mais altos para a vogal /i/;
- Os valores de *shimmer*, APQ3, APQ5 e APQ11 mais elevados tanto para a vogal /a/ como para a vogal /i/;
- Os valores de HNR mais baixos para ambas as vogais.

Na Figura 4.10 é ilustrado um gráfico que exhibe as disparidades existentes entre os valores obtidos para os parâmetros do sujeito F_13 e a mediana desses parâmetros na amostra da população feminina.

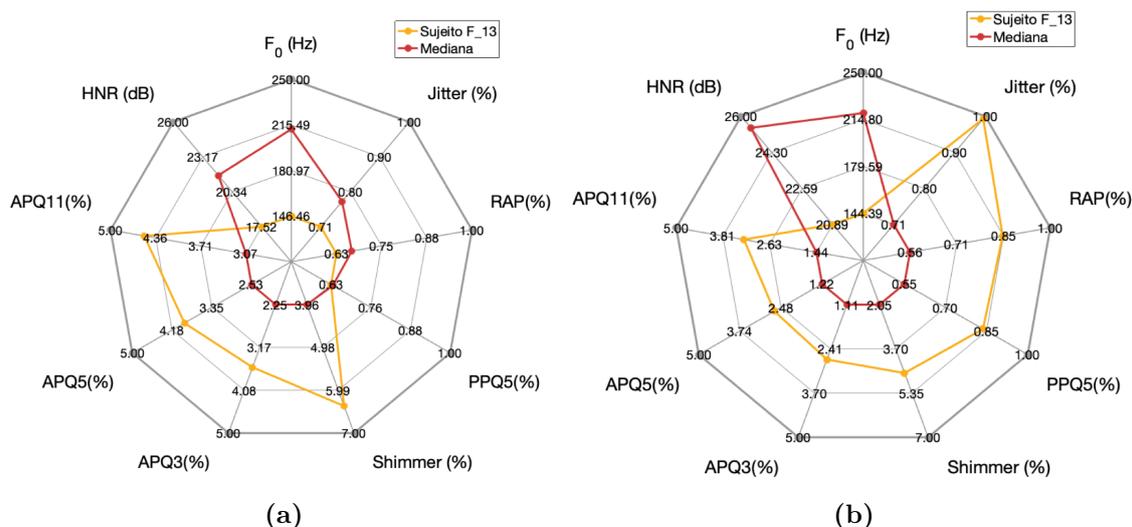


Figura 4.10: Diagramas representativos do desvio dos parâmetros acústicos do sujeito F_13 relativamente à mediana da amostra feminina (a) Parâmetros correspondentes à produção da vogal /a/ (b) Parâmetros correspondentes à produção da vogal /i/.

Analisando a Figura 4.10, confirma-se que apenas o parâmetro *jitter* da vogal /a/ apresenta valores mais ou menos semelhantes aos da generalidade da amostra, sendo que contrariamente à maior parte da amostra exibe valores de *jitter* para a vogal /i/ superiores aos que exibe para vogal /a/. Todos os outros parâmetros correspondem aos piores valores da amostra feminina, dos quais se destaca o *shimmer* que ultrapassa os limiares de patologia para ambas as vogais. As medidas de *jitter*, RAP, PPQ5 e as medidas de *shimmer*, APQ3, APQ5 e APQ11, apresentam o mesmo padrão de comportamento que os respetivos parâmetros principais relativamente à mediana. Apesar de exibirem resultados redundantes podem ser úteis em casos idênticos em que seja necessária uma discriminação fina.

Considerando os restantes sujeitos com restrições nesta amostra, é de notar que o sujeito F_4, que é fumador, demonstrou os segundos valores de *shimmer* mais altos da amostra para ambas as vogais. Por outro lado, o sujeito F_14, que já foi diagnosticado com nódulos, o sujeito F_3, que é fumador, e os sujeitos com asma, não apresentaram nenhum parâmetro com diferenças relevantes.

No que diz respeito à amostra masculina, o sujeito M_15, que é fumador e foi submetido a uma anestesia geral há cerca de 6 meses, apresentou parâmetros com valores relativamente distintos aos da restante amostra, particularmente no caso do *jitter*.

Este sujeito apresentou:

- Valores de F_0 intermédios para ambas as vogais;
- Os valores de *jitter*, RAP e PPQ5 mais altos tanto para a vogal /a/ como para a vogal /i/;
- Valores de *shimmer*, APQ3, APQ5 e APQ11 elevados para ambas as vogais;
- Valores de HNR intermédios para ambas as vogais;

Na Figura 4.11 é apresentando um gráfico que ilustra as disparidades existentes entre os valores obtidos para os parâmetros do sujeito M_15 e a mediana desses parâmetros na amostra da população masculina.

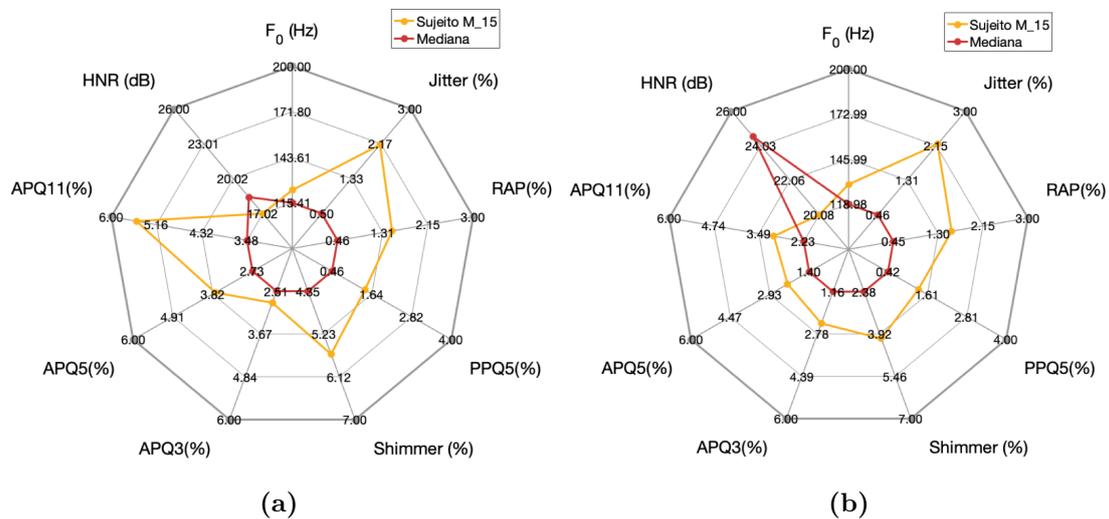


Figura 4.11: Diagramas representativos do desvio dos parâmetros acústicos do sujeito M_15 relativamente à mediana da amostra masculina (a) Parâmetros correspondentes à produção da vogal /a/ (b) Parâmetros correspondentes à produção da vogal /i/.

Analisando a Figura 4.11, conclui-se que os valores de *jitter* apresentam resultados consideravelmente superiores aos da mediana, sendo que ultrapassam o limite patológico para ambas as vogais. Os valores de *shimmer*, embora não correspondam aos mais elevados da amostra, também exibem resultados insatisfatórios relativamente aos limiares patológicos. Mais uma vez, as medidas de *jitter* e *shimmer* são concordantes com os respetivos parâmetros. Os restantes parâmetros encontram-se na gama de valores esperada, porém destaca-se que os valores de HNR são inferiores aos da mediana da amostra.

Relativamente aos restantes sujeitos com restrições, é de salientar que o sujeito M_4, que é fumador, apresentou os valores de *shimmer* mais elevados da amostra para ambas as vogais. Enquanto o sujeito M_3, que também é fumador, apresentou todos os parâmetros com comportamento semelhante aos da amostra a que pertence.

O presente estudo não dispõe de população suficiente para tirar conclusões acerca da influência que as restrições apresentadas têm na qualidade vocal. Seria necessário reunir mais população nas mesmas condições para determinar o significado concreto das restrições apresentadas nesta secção. No entanto, confirma-se que pelo menos os sujeitos F_13, F_4, M_15 e M_4 não podem fazer parte da caracterização de uma população de controlo. Existem também sujeitos sem qualquer tipo de restrição com valores superiores aos estipulados pela literatura, nomeadamente no parâmetro *shimmer*. Isto reforça o facto de que, dadas as características multidimensionais da voz, apenas os parâmetros acústicos não são suficientes para aferir com rigor sobre a qualidade da mesma.

Conclusão e Trabalhos Futuros

A cirurgia da tireoide consiste na remoção total ou parcial da glândula, dependendo do número e tipo de nódulos que a envolvem. É um procedimento com uma baixa taxa de complicações, mas existem casos de lesões no nervo laríngeo que provocam alterações vocais. Na prática clínica, várias técnicas têm sido utilizadas para avaliar a qualidade vocal de pacientes submetidos a tireoidectomias. Contudo, este procedimento carece de metodologias objetivas das alterações que surgem no decorrer do processo de recuperação da voz. Os principais parâmetros quantitativos envolvidos diretamente no processo de produção de fala são, a Frequência Fundamental, o *jitter*, o *shimmer* e o HNR.

O objetivo inicial deste projeto consistia na caracterização da qualidade vocal pré e pós tireoidectomia, através de um algoritmo capaz de determinar os parâmetros acústicos mais relevantes. Face à impossibilidade de realizar gravações de voz em ambiente hospitalar, o objetivo principal passou a ser a realização de uma análise comparativa entre o algoritmo desenvolvido em MATLAB e o *software* Praat, amplamente difundido e conceituado no meio clínico.

Para verificar a adequação do algoritmo proposto, foram analisados sinais acústicos de 30 indivíduos e realizada uma comparação da dispersão métrica dos resultados obtidos em ambos os *softwares*. Os resultados sugerem:

1. Boas correlações entre os *boxplots* do Praat e do MATLAB para a maioria dos parâmetros, à exceção do parâmetro *jitter* e respetivas medidas (RAP e PPQ5) para o caso da produção de ambas as vogais na amostra feminina, justificado pelo exemplo da secção 4.1;
2. Valores de *shimmer* relativamente elevados, tendo em conta os valores definidos como limiares de patologia. No entanto, como aparecem elevados em ambos os *softwares*, exclui-se a possibilidade de se tratar de um erro introduzido pelo algoritmo;

3. Na amostra feminina, verificou-se que os valores mais discrepantes, relativamente à mediana dos parâmetros de ambas as vogais, pertenciam ao sujeito F_13 que foi diagnosticado com nódulos na tiroide.
4. Na amostra masculina, o sujeito M_15 que é fumador e que já foi submetido a anestesia geral, apresentou valores com alguma discrepância em relação à mediana. Em particular no parâmetro *jitter*, foi o único sujeito a apresentar valores acima dos limiares patológicos para ambas as vogais.
5. Na generalidade da amostra, registaram-se valores de *shimmer* bastante elevados em 2 sujeitos fumadores (F_4 e M_4).
6. Os restantes sujeitos com restrições não apresentaram discrepâncias relevantes.

De um modo geral, os resultados obtidos provam que este algoritmo permite determinar os parâmetros com valores suficientemente próximos aos do Praat.

Alguns dos dados obtidos para sujeitos com vozes consideradas saudáveis demonstraram que as medidas de perturbação de F_0 e da amplitude são especialmente suscetíveis à influência do tipo, colocação e ângulo do microfone; do tipo de algoritmo de processamento e das configurações do *software*; do ruído ambiente e dos efeitos de ruídos introduzidos pelo próprio instrumento de gravação.

Dadas as circunstâncias que restringiram o processo de gravação, o número da amostra em estudo foi reduzido. A esta limitação acresce o facto de que nem todos os sujeitos analisados podem fazer parte de uma população de controlo. No que diz respeito aos sujeitos que apresentam restrições, face à disparidade de fatores envolvidos, não é possível fazer generalizações que permitam tirar conclusões específicas. Contudo, destaca-se que os parâmetros *jitter* e *shimmer* parecem ser adequados para um primeiro diagnóstico de algumas alterações associadas ao aparelho fonador.

Em relação a trabalhos futuros, sugere-se que sejam colmatadas as limitações referidas, através da melhoria das condições de gravação, do alargamento da amostra analisada e da imposição de critérios de inclusão e exclusão. Além disso, para caracterizar melhor a amostra, podem ser avaliadas outras vogais ou sinais acústicos em contexto de fala e determinados outros parâmetros.

Alguns dos parâmetros determinados no presente estudo dependem de uma estimativa precisa da Frequência Fundamental, sendo que esta pode ser uma tarefa complexa na presença de certas patologias ou irregularidades vocais. Assim, é imprescindível avaliar diversas vozes pré e pós tiroidectomia para averiguar a apropriação e robustez do algoritmo num contexto real. De forma a corroborar tais resultados, é

sugerida a sua comparação com imagens laríngeas adquiridas em videoestroboscopia ou com outras técnicas objetivas de avaliação vocal.

Bibliografia

- [1] A. Giovanni and S. Duffo, “Physiology of voice production,” in *Surgery of Larynx and Trachea*, pp. 1–9, Springer, 2009.
- [2] J. A. Seikel, D. G. Drumright, and D. W. King, *Anatomy & physiology for speech, language, and hearing*. Nelson Education, 2015.
- [3] F. L. Wuyts, M. S. D. Bodt, G. Molenberghs, M. Remacle, L. Heylen, B. Millet, K. V. Lierde, J. Raes, and P. H. V. d. Heyning, “The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach,” *Journal of speech, language, and hearing research*, vol. 43, no. 3, pp. 796–809, 2000.
- [4] M. Hirano, “Objective evaluation of the human voice: clinical aspects,” *Folia Phoniatrica et Logopaedica*, vol. 41, no. 2-3, pp. 89–144, 1989.
- [5] P. H. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. Van De Heyning, M. Remacle, and V. Woisard, “A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques,” *European Archives of Oto-rhino-laryngology*, vol. 258, no. 2, pp. 77–82, 2001.
- [6] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, “Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters,” *IEEE transactions on biomedical engineering*, vol. 53, no. 10, pp. 1943–1953, 2006.
- [7] G. Fant, *Acoustic theory of speech production*. No. 2, Walter de Gruyter, 1970.
- [8] J. P. Teixeira, C. Oliveira, and C. Lopes, “Vocal acoustic analysis—jitter, shimmer and hnr parameters,” *Procedia Technology*, vol. 9, pp. 1112–1122, 2013.
- [9] M. Behlau, “Voz - livro do especialista, vol. 1,” *Rio de Janeiro: Revinter*, 2001.

- [10] P. Enderby and J. Emerson, “Speech and language therapy: does it work?,” *BMJ*, vol. 312, no. 7047, pp. 1655–1658, 1996.
- [11] P. Behlau, Mara Pontes, *Avaliação e tratamento das disfonias*. Editora Lovise, 1995.
- [12] J. Lee, K. Y. Na, R. M. Kim, Y. Oh, J. H. Lee, J. Lee, J.-S. Lee, C.-H. Kim, E.-Y. Soh, and W. Y. Chung, “Postoperative functional voice changes after conventional open or robotic thyroidectomy: a prospective trial,” *Annals of surgical oncology*, vol. 19, no. 9, pp. 2963–2970, 2012.
- [13] S. S. Chandrasekhar, G. W. Randolph, M. D. Seidman, R. M. Rosenfeld, P. Angelos, J. Barkmeier-Kraemer, M. S. Benninger, J. H. Blumin, G. Dennis, J. Hanks, *et al.*, “Clinical practice guideline: improving voice outcomes after thyroid surgery,” *Otolaryngology—Head and Neck Surgery*, vol. 148, no. 6_suppl, pp. S1–S37, 2013.
- [14] E. C. Thompson-Ward and B. E. Murdoch, “Instrumental assessment of the speech mechanism,” *Dysarthria: A Physiological Approach*. Cheltenham: Stanley Thornes Ltd, pp. 68–101, 1998.
- [15] M. Morrison, L. Rammage, H. Nichol, B. Pullan, P. May, and L. Salkeld, “Anatomy and physiology of voice production,” in *The Management of Voice Disorders*, pp. 161–200, Springer, 1994.
- [16] P. Lieberman and S. E. Blumstein, *Speech physiology, speech perception, and acoustic phonetics*. Cambridge University Press, 1988.
- [17] “Sons da fala.” https://moodle.ufsc.br/pluginfile.php/905248/mod_resource/content/1/Sons_da_fala_letras.pdf. Acedido em: 2020-09-27.
- [18] J. Laver, “The phonetic description of voice quality,” *Cambridge Studies in Linguistics London*, vol. 31, pp. 1–186, 1980.
- [19] C. Van Riper, *Speech correction*. Prentice-Hall, 1972.
- [20] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer handbook of speech processing*. Springer, 2007.
- [21] B. H. Story, “Mechanisms of voice production,” *The handbook of speech production*, pp. 34–58, 2015.
- [22] M. Hirano, “Morphological structure of the vocal cord as a vibrator and its variations,” *Folia Phoniatria et Logopaedica*, vol. 26, no. 2, pp. 89–94, 1974.

-
- [23] I. Guimarães, “A ciência e a arte da voz humana,” *Alcoitão, Escola Superior de Saúde de Alcoitão*, 2007.
- [24] H. R. Vikram, G. Dhaliwal, S. Saint, and C. B. Simpson, “A recurrent problem,” *New England Journal of Medicine*, vol. 364, no. 22, pp. 2148–2154, 2011.
- [25] I. R. Titze, *Vocal Fold Physiology: Frontiers in Basic Science*, vol. 7. Singular Publishing Group, 1993.
- [26] J. Van den Berg, “Myoelastic-aerodynamic theory of voice production,” *Journal of speech and hearing research*, vol. 1, no. 3, pp. 227–244, 1958.
- [27] J. P. Teixeira, *Modelização paramétrica de sinais para aplicação em sistemas de conversão texto-fala*. PhD thesis, FEUP, 1995.
- [28] A. Aronson, “Clinical voice disorders: an interdisciplinary approach. new york: Brain c,” 1990.
- [29] A. Aronson and D. Bless *Clinical Voice Disorders. 4th ed. New York: Thieme Med Pub*, 2009.
- [30] J. P. Teixeira and P. O. Fernandes, “Acoustic analysis of vocal dysphonia,” *Procedia Computer Science*, vol. 64, pp. 466–473, 2015.
- [31] K. Verdolini, C. A. Rosen, and R. C. Branski, *Classification manual for voice disorders-I*. Psychology Press, 2006.
- [32] M. E. Freeman and M. E. Fawcus, *Voice disorders and their management*. Whurr Publishers, 2000.
- [33] Y. J. Shin and K. H. Hong, “Cepstral analysis of voice in patients with thyroidectomy,” *Clinical and experimental otorhinolaryngology*, vol. 9, no. 2, p. 157, 2016.
- [34] N. P. Solomon, S. N. Awan, L. B. Helou, and A. Stojadinovic, “Acoustic analyses of thyroidectomy-related changes in vowel phonation,” *Journal of Voice*, vol. 26, no. 6, pp. 711–720, 2012.
- [35] N. P. Mcivor, D. J. Flint, J. Gillibrand, and R. P. Morton, “Thyroid surgery and voice-related outcomes,” *Australian and New Zealand Journal of Surgery*, vol. 70, no. 3, pp. 179–183, 2000.
- [36] G. W. Randolph, *Surgery of the thyroid and parathyroid glands: expert consult premium edition-enhanced online features and print*. Elsevier Health Sciences, 2012.

- [37] S. Akyildiz, F. Ogut, M. Akyildiz, and E. Z. Engin, “A multivariate analysis of objective voice changes after thyroidectomy without laryngeal nerve injury,” *Archives of Otolaryngology–Head & Neck Surgery*, vol. 134, no. 6, pp. 596–602, 2008.
- [38] P. W. Flint, B. H. Haughey, K. T. Robbins, J. R. Thomas, J. K. Niparko, V. J. Lund, and M. M. Lesperance, *Cummings otolaryngology-head and neck surgery e-book*. Elsevier Health Sciences, 2014.
- [39] G. B. Gour and V. Udayashankara, “A new diagnostic approach: Voice analysis of thyroid patients after thyroid surgery without laryngeal nerve injury,” 2013.
- [40] J. Stemple, L. Glaze, and B. Klaben, *Clinical voice pathology: Theory and management*. San Diego, CA: Plural Publishing Inc, 2010.
- [41] B. H. Jacobson, A. Johnson, C. Grywalski, A. Silbergleit, G. Jacobson, M. S. Benninger, and C. W. Newman, “The voice handicap index (vhi) development and validation,” *American Journal of Speech-Language Pathology*, vol. 6, no. 3, pp. 66–70, 1997.
- [42] N. D. Hogikyan and G. Sethuraman, “Validation of an instrument to measure voice-related quality of life (V-RQOL),” *Journal of Voice*, vol. 13, no. 4, pp. 557–569, 1999.
- [43] S. Fex, “Perceptual evaluation,” *Journal of Voice*, vol. 6, no. 2, pp. 155–158, 1992.
- [44] M. Hirano, “Clinical examination of voice. 1981,” *Viena: Springer-Verlag*.
- [45] G. B. Kempster, B. R. Gerratt, K. V. Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, “Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol,” *American Journal of Speech-Language Pathology*, 2009.
- [46] R. D. Kent, *The MIT Encyclopedia of Communication Disorders*. MIT Press, 2004.
- [47] M. Rothenberg, “A new inverse-filtering technique for deriving the glottal air flow waveform during voicing,” *The Journal of the Acoustical Society of America*, vol. 53, no. 6, pp. 1632–1645, 1973.
- [48] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt, “Toward improved ecological validity in the acoustic measurement of overall voice qua-

- lity: combining continuous speech and sustained vowels,” *Journal of Voice*, vol. 24, no. 5, pp. 540–555, 2010.
- [49] J. C. Morente, J. A. Torres, M. C. Jiménez, D. P. Maroto, V. P. Rodríguez, E. M. Gomariz, E. C. Baños, and A. J. Ramos, “Estudio objetivo de la voz en población normal y en la disfonía por nódulos y pólipos vocales,” *Acta Otorrinolaringológica Española*, vol. 52, no. 6, pp. 476–482, 2001.
- [50] J. W. Martens, H. Versnel, and P. H. Dejonckere, “The effect of visible speech in the perceptual rating of pathological voices,” *Archives of Otolaryngology–Head & Neck Surgery*, vol. 133, no. 2, pp. 178–185, 2007.
- [51] A. P. Mendes, L. J. Ferreira, and E. Castro, “Softwares e hardwares de análise acústica da voz e da fala,” *Distúrbios da Comunicação*, vol. 24, no. 3, 2012.
- [52] C. P. Carson, D. R.-S. Ingrisano, and K. D. Eggleston, “The effect of noise on computer-aided measures of voice: a comparison of cspeechsp and the multi-dimensional voice program software using the csl 4300b module and multi-speech for windows,” *Journal of Voice*, vol. 17, no. 1, pp. 12–20, 2003.
- [53] Y. Maryn, N. Roy, M. De Bodt, P. Van Cauwenberge, and P. Corthals, “Acoustic measurement of overall voice quality: a meta-analysis,” *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2619–2634, 2009.
- [54] P. Boersma and D. Weenink, “Praat: Doing phonetics by computer (version 6.1.04) [Software de computador] adquirido a 28 setembro de 2019 de <http://www.praat.org/>,” *Amsterdam: Institute of Phonetic Sciences*, 2019.
- [55] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glott. Int.*, vol. 5, no. 9, pp. 341–345, 2001.
- [56] P. Van Lieshout, “Praat – short tutorial – an introduction,” 2017.
- [57] E. Seifert, “Stress and distress in non-organic voice disorder,” *Swiss medical weekly*, vol. 135, no. 2728, 2005.
- [58] M. H. M. Mateus, I. Falé, and M. J. Freitas, *Fonética e fonologia do português*. Universidade Aberta, 2016.
- [59] D. L. Sinagra, M. R. Montesinos, V. A. Tacchi, J. C. Moreno, J. E. Falco, N. A. Mezzadri, D. L. Debonis, and H. P. Curutchet, “Voice changes after thyroidectomy without recurrent laryngeal nerve injury,” *Journal of the American College of Surgeons*, vol. 199, no. 4, pp. 556–560, 2004.

- [60] J. Fernandes, L. Silva, F. Teixeira, V. Guedes, J. Santos, and J. P. Teixeira, "Parameters for vocal acoustic analysis-cured database," *Procedia Computer Science*, vol. 164, pp. 654–661, 2019.
- [61] M. Brockmann, M. J. Drinnan, C. Storck, and P. N. Carding, "Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task," *Journal of Voice*, vol. 25, no. 1, pp. 44–53, 2011.
- [62] R. E. Deal and F. W. Emanuel, "Some waveform and spectral features of vowel roughness," *Journal of Speech and Hearing Research*, vol. 21, no. 2, pp. 250–264, 1978.
- [63] E. Mendoza and G. Carballo, "Acoustic analysis of induced vocal stress by means of cognitive workload tasks," *Journal of Voice*, vol. 12, no. 3, pp. 263–273, 1998.
- [64] E. Yumoto, W. J. Gould, and T. Baer, "Harmonics-to-noise ratio as an index of the degree of hoarseness," *The journal of the Acoustical Society of America*, vol. 71, no. 6, pp. 1544–1550, 1982.
- [65] J. Lopes, S. Freitas, R. Sousa, J. Matos, F. Abreu, and A. Ferreira, "A medida hnr: sua relevância na análise acústica da voz e sua estimação precisa," *Guarda: I Jornadas sobre Tecnologia e Saúde*, 2008.
- [66] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the Institute of Phonetic Sciences*, vol. 17, pp. 97–110, Amsterdam, 1993.
- [67] C. T. Ferrand, "Harmonics-to-noise ratio: an index of vocal aging," *Journal of Voice*, vol. 16, no. 4, pp. 480–487, 2002.
- [68] B. R. Gerratt, J. Kreiman, and M. Garellek, "Comparing measures of voice quality from sustained phonation and continuous speech," *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 5, pp. 994–1001, 2016.
- [69] L. M. Jesus, A. Barney, R. Santos, J. Caetano, J. Jorge, and P. S. Couto, "Universidade de Aveiro's Voice Evaluation Protocol," in *Tenth Annual Conference of the International Speech Communication Association*, 2009.
- [70] D. M. Recorder, "Operation manual," *DEF*, vol. 8, 2016.
- [71] "Gibraltar Desertcart." <https://gibraltar.desertcart.com/products/64153987-zoom-h4n>. Acedido: 2020-10-05.

- [72] “Billauer.” <http://www.billauer.co.il/peakdet.html>. Acedido em: 2019-11-26.
- [73] “Mathworks.” <https://www.mathworks.com/matlabcentral/fileexchange/18378-manual-audio-annotation1>. Acedido em: 2020-02-17.
- [74] D. D. Mehta and R. E. Hillman, “Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods,” *Current opinion in otolaryngology & head and neck surgery*, vol. 16, no. 3, p. 211, 2008.

