

## **Patologias da Laringe com Análise Acústica Vocal**

**André Amorim Gonçalves**

Dissertação apresentada à  
**Escola Superior de Tecnologia e Gestão**  
**Instituto Politécnico de Bragança**  
para a obtenção do grau de Mestre em  
**Tecnologia Biomédica**

Este trabalho foi efetuado sob orientação de  
Professor Doutor João Paulo Teixeira

Bragança, Dezembro de 2015



# AGRADECIMENTOS

---

Ao terminar esta tese resta-me registrar os meus sinceros agradecimentos às individualidades que de várias formas contribuíram para que este se tornasse numa realidade.

Ao orientador o Professor Doutor João Paulo Teixeira, por toda a dedicação, compreensão e amizade, pelos desafios cada vez mais complexos que foi colocando na realização deste trabalho, e pelo estímulo e exigência crescente que foi impondo à medida que caminhava para a sua conclusão.

À minha família, namorada e colegas pelo apoio e ânimo que me deram ao longo deste trabalho.



Este trabalho tem como objetivo produzir um algoritmo robusto capaz de medir corretamente os parâmetros da voz, Jitter, Shimmer e HNR, em vozes patológicas para posterior identificação destas patologias com recurso a sistemas inteligentes. Foi ainda realizada uma análise comparativa de cada um destes parâmetros para vozes masculinas e femininas, ou entre vozes de controlo e vozes patológicas.

O algoritmo foi desenvolvido em linguagem do MatLab, e, basicamente consiste na deteção dos picos ciclo a ciclo, variando consoante o período glotal dos sinais de voz. Esta identificação pode ser relativamente simples em vozes sintetizadas, é um pouco mais complexa em vozes reais e pode ser bastante mais difícil em vozes patológicas em que os ciclos glotais podem ser muito irregulares, e diferentes de paciente para paciente.

O Jitter consiste na variação dos ciclos glotais e pode ser medido de formas diferentes como os parâmetros Jitt, Jitta, Rap e Ppq5. O Shimmer consiste na variação da amplitude dos sucessivos ciclos glotais e pode ser medido pelos parâmetros Shim, ShdB, Apq3 e Apq5. O HNR é a relação sinal ruído.

Utilizaram-se os valores obtidos pelo programa Praat para estes parâmetros como termo de comparação com os resultados do algoritmo desenvolvido. O algoritmo foi testado com sinais sintetizados, com valores bem conhecidos para os parâmetros, sinais de voz normal (Controlo) e sinais de voz patológicos provenientes da base de dados Saarbrucken Voice Database (SVD).

Na comparação realizada utilizando o sinal sintetizado o algoritmo produziu um erro inferior a 5  $\mu$ s para o parâmetro Jitta e inferiores a 0.1% para o Shim. Quando comparado com vozes reais (sinais de vozes de controlo e patológicas), as diferenças de valores entre o Praat e o algoritmo foram pouco significativas.

Foi também realizada uma comparação estatística do comportamento dos parâmetros do Jitter e do Shimmer em sinais de voz de controlo e sinais provenientes das patologias Laringite, Disfonia Hiperfuncional, Disfonia Espasmódica, Pólipos das Cordas Vocais e Envelhecimento das Cordas Vocais. Destas só as três últimas apresentaram distinção estatisticamente significativa dos parâmetros em relação ao grupo de sinais de voz de controlo.

Palavras-chave: Frequência Fundamental, Jitter, Shimmer, Harmonic to Noise Ratio, Patologias da Laringe



# ABSTRACT

---

This work aims to produce a robust algorithm that can correctly measure the voice parameters, Jitter, Shimmer and HNR in pathological voices for later identification of these pathologies using intelligent systems. It was also performed a comparative analysis of each of these parameters for male and female voices, or between control and pathological voices.

The algorithm was developed under Matlab language, and basically consists of the detection of the cycle to cycle peaks, varying depending on the period of glottal voice signals. This identification can be relatively simple in synthesized voice, it is somewhat more complex in real voices and can be quite difficult in pathological voices where the glottal cycles may be very irregular and different from patient to patient.

Jitter is the variation of the glottal cycles and can be measured in different ways like Jitt, Jitta, Rap and Ppq5 parameters. The shimmer is the amplitude variation of the successive glottal cycles and can be measured by Shim ShdB, Apq3 and Apq5 parameters. The HNR is the Harmonic to Noise Ratio.

The values obtained by the Praat program for these parameters were used as a comparison with the results of the developed algorithm. The algorithm has been tested with synthesized signals with well-known values, with normal voice signals (Control) and pathological voice signals from Saarbrücken Voice Database (SVD).

Using the synthesized signal the algorithm produced an error of less than 5  $\mu$ s for Jitta parameter and less than 0.1% for Shim. When compared with real voices (voice signals control and pathological), differences in values between the Praat and the algorithm were negligible.

It has also produced a statistical comparison between control signals and signals from Laryngitis, Hyperfunctional Dysphonia, Spasmodic Dysphonia, Vocal Cord Polyps and vocal Cords Ageing pathologies. Of these only the last three have statistically distinct parameters in relation to the group of voice control signals.

Keywords: Fundamental Frequency, Jitter, Shimmer, Harmonic to Noise Ratio, Pathologies of the Larynx



Agradecimentos.....	iii
Resumo.....	v
Abstract .....	vii
Índice.....	ix
Índice de Figuras .....	xii
Índice de Tabelas.....	xiv
Abreviaturas e Símbolos .....	xvii
Capítulo I.....	1
1. Introdução .....	1
1.1. Estado da arte .....	5
1.2. Funcionamento e anatomia do aparelho fonador.....	8
Capítulo II .....	11
2. Base de dados dos sinais de fala .....	11
2.1. Sinais patológicos.....	12
2.1.1. Laringite.....	12
2.1.2. Disfonia Hiperfuncional .....	13
2.1.3. Disfonia Espasmódica .....	13
2.1.4. Pólipo das Cordas Vocais .....	13
2.1.5. Envelhecimento das Cordas Vocais.....	13
2.2. Sinal sintetizado .....	14
Capítulo III.....	17
3. Determinação dos Parâmetros de fala .....	17
3.1. Introdução.....	17
3.1.1. Jitter .....	17
3.1.2. Shimmer.....	19

3.1.3. HNR.....	20
3.2. Variação de Jitter e Shimmer .....	23
3.2.1. Variação do Jitter .....	23
3.2.1.1. Perturbação de Jitter tipo 1 .....	23
3.2.1.2. Perturbação de Jitter tipo 2 .....	24
3.2.2. Variação de Shimmer.....	26
3.2.2.1. Perturbação de Shimmer do tipo 1.....	26
3.2.2.2. Perturbação de Shimmer do tipo 2.....	27
3.2.3. Variação da frequência fundamental .....	28
3.3. Algoritmo de Identificação dos Períodos Glotais.....	29
3.3.1. Evolução do algoritmo.....	31
3.3.2. Algoritmo Final .....	42
Capítulo IV.....	45
4. Análise da Precisão da Medida do Algoritmo.....	45
4.1. Precisão na medição dos parâmetros em sinais Sintetizados .....	45
4.1.1. Análise ao parâmetro Jitter .....	45
4.1.2. Análise ao parâmetro Shimmer .....	47
4.1.3. Análise das Medidas com variação da Frequência Fundamental .....	49
4.1.4. Resumo da Precisão das Medidas Com Sinais Sintetizados.....	51
4.2. Comparação das Medidas do Algoritmo aplicado a sinais de voz .....	53
4.2.1. Comparação Algoritmo/Praat para sinais de vozes de controlo .....	54
4.2.1.1. Jitter .....	54
4.2.1.2. Shimmer.....	57
4.2.1.3. HNR.....	60
4.2.2. Comparação Algoritmo/Praat para sinais de voz patológicos .....	61
4.2.2.1. Jitter .....	62
4.2.2.2. Shimmer.....	65

4.2.2.3. HNR .....	68
4.2.3. Resumo da Comparação das Medidas Com Sinais Reais.....	69
Capítulo V .....	71
5. Comparação dos Sinais do Grupo de Controlo Com Sinais dos Grupos Patológicos .....	71
Capítulo VI.....	83
6. Conclusão e Trabalhos Futuros .....	83
6.1. Conclusão .....	83
6.2. Trabalhos Futuros .....	85
Bibliografia .....	87

# ÍNDICE DE FIGURAS

---

Figura 1 – Representação das principais estruturas que contribuem para o mecanismo de fala.....	8
Figura 2 - Representação das Pregas Vocais.....	9
Figura 3 - Processo de abertura e fecho das cordas vocais correspondente a um período glotal (Dajer,2010).....	10
Figura 4 – Laringite na parte posterior.....	12
Figura 5 - Modelo genérico para a produção de fala. (alterado de Teixeira, 1995).....	14
Figura 6- Ilustração de três impulsos glotais após a aplicação da equação 1 à sequência de impulsos unitários.....	15
Figura 7 - – Ilustração de três impulsos glotais do sinal sintetizado.....	16
Figura 8- Representação dos parâmetros Jitter e Shimmer para um sinal de fala.....	17
Figura 9 - Perturbação de Jitter tipo 1 com variação de um em um período glotal. ....	24
Figura 10 - Perturbação de Jitter tipo 2 com variação de três em três períodos glotais.....	25
Figura 11 - Perturbação de Shimmer do tipo 1 com variação da amplitude de um em um período glotal.....	26
Figura 12 - Perturbação de Shimmer do tipo 2 com variação da amplitude de três em três períodos glotais.....	27
Figura 13 - Sinal de voz controlo da vogal /a/ tom normal, onde os períodos glotais estão bem definidos. ....	29
Figura 14 - Sinal de voz com a patologia Laringite da vogal /a/ tom normal, onde os períodos glotais não estão bem definidos.....	30
Figura 15 – Sinal de voz controlo da vogal /u/ tom baixo, em que a determinação é feita de forma errada.....	32
Figura 16 - Sinal de voz controlo da vogal /a/ tom alto, ilustrativa de que a determinação seria melhor efetuada através dos picos negativos. ....	33
Figura 17 - Sinal de voz controlo da vogal /u/ tom baixo, ilustrativa da dificuldade de determinação dos parâmetros para este tipo de sinais. ....	34
Figura 18 - Sinal de voz controlo da vogal /u/ tom alto, a – pico usado de critério de escolha na zona central do sinal. b – Zona do sinal onde a determinação é efetuada de forma errada. c – Zoom de uma secção de b.....	35

Figura 19 - Fluxograma do algoritmo de identificação dos períodos glotais.....	44
Figura 20 - Gráfico ilustrativo dos vários grupos para o parâmetro Jitta.....	73
Figura 21 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Jitt.....	73
Figura 22 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Rap. ....	74
Figura 23 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Ppq5. ....	74
Figura 24 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Shim. ....	75
Figura 25 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Shdb. ....	75
Figura 26 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Apq3.....	76
Figura 27 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Apq5.....	76
Figura 28 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro HNR. ....	77
Figura 29 – Comparação entre sinais do grupo controlo com sinais de grupos patológicos de Jitta para o tom baixo.....	78

# ÍNDICE DE TABELAS

---

Tabela 1 - Valores medidos para o parâmetro HNR (em dB) com variação da janela para a vogal /a/.....	37
Tabela 2 - Valores medidos para o parâmetro HNR (em dB) com variação da janela para a vogal /i/.....	38
Tabela 3 - Valores medidos para o parâmetro HNR (em dB) com variação da janela para a vogal /u/.....	39
Tabela 4 - Valores medidos para o parâmetro HNR (em dB) com variação da janela resumido conforme a vogal.....	40
Tabela 5 - Valores medidos para o parâmetro HNR (em dB) com variação da janela resumido conforme o tom.....	41
Tabela 6 - Valores de Jitter medidos para o sinal sintetizado com Jitter nulo.....	45
Tabela 7 - Valores de Jitter medidos para o sinal sintetizado com perturbação de Jitter tipo 1.....	46
Tabela 8 - Valores de Jitter para o sinal sintetizado com perturbação de Jitter tipo 2.....	46
Tabela 9 - Valores de medidos de Shimmer para o sinal sintetizado sem variação da amplitude.....	47
Tabela 10 - Valores de medidos de Shimmer para o sinal sintetizado com perturbação de Shimmer do tipo 1.....	48
Tabela 11 - Valores de medidos de Shimmer para o sinal sintetizado com perturbação de Shimmer do tipo 2.....	48
Tabela 12 - Valores de Shimmer para o sinal sintetizado com diferentes $F_0$ e sem variação da amplitude.....	49
Tabela 13 - Valores de Shimmer para o sinal sintetizado com diferentes valores de $F_0$ e com perturbação de Shimmer do tipo 1.....	50
Tabela 14 - Valores de Shimmer para o sinal sintetizado com diferentes valores de $F_0$ e com perturbação de Shimmer do tipo 2.....	50
Tabela 15 – Resumo dos erros para as medidas de Jitter e Shimmer com o algoritmo e com o Praat com o sinal sintetizado.....	51
Tabela 16 - Média e desvio padrão dos parâmetros de Jitter ordenados por vogal e tom para os sinais de controle.....	55

Tabela 17 - Média e desvio padrão dos parâmetros de Jitter ordenados por vogal para os sinais de controlo.....	56
Tabela 18 - Média e desvio padrão dos parâmetros de Jitter ordenados por tom para os sinais de voz controlo. ....	57
Tabela 19 - Média e desvio padrão dos parâmetros de Shimmer ordenados por vogal e tom para os sinais de controlo.....	58
Tabela 20 - Média e desvio padrão dos parâmetros de Shimmer ordenados por vogal para os sinais de controlo. ....	59
Tabela 21 - : Média e desvio padrão dos parâmetros de Shimmer ordenados por tom para os sinais de controlo. ....	59
Tabela 22 - Média e desvio padrão do parâmetro HNR (dB) para os sinais de controlo.....	60
Tabela 23 - Média e desvio padrão do parâmetro HNR (dB) ordenado por vogal para os sinais de controlo.....	60
Tabela 24 - Média e desvio padrão do parâmetro HRN (dB) ordenado por tom para os sinais de controlo.....	61
Tabela 25 - Média e desvio padrão das medidas de Jitter ordenados por vogal e tom para os sinais da patologia laringite. ....	63
Tabela 26 - Média e desvio padrão das medidas de Jitter ordenados por vogal para os sinais da patologia laringite. ....	64
Tabela 27 - Média e desvio padrão das medidas de Jitter ordenados por tom para os sinais da patologia laringite. ....	65
Tabela 28 - Média e desvio padrão das medidas de Shimmer ordenados por vogal e tom para os sinais da patologia laringite.....	66
Tabela 29 - Média e desvio padrão das medidas de Shimmer ordenados por vogal para os sinais da patologia laringite. ....	67
Tabela 30 - Média e desvio padrão das medidas de Shimmer ordenados por tom para os sinais da patologia laringite. ....	68
Tabela 31 - Média e desvio padrão do parâmetro HNR para os sinais da patologia laringite. ....	69
Tabela 32 - Média e desvio padrão do parâmetro HNR ordenado por vogal para os sinais da patologia laringite. ....	69
Tabela 33 - Média e desvio padrão do parâmetro HRN ordenado por tom para os sinais da patologia laringite. ....	69

Tabela 34 – Diferenças entre médias dos valores medidos pelo algoritmo e pelo programa Praat.....	69
Tabela 35 – Resumo das idades e média de idades dos grupos utilizados no estudo. ....	72
Tabela 36 – Resumo da variação dos parâmetros dos grupos patológicos relativamente ao grupo de controlo.....	77
Tabela 37 – Variação dos parâmetros dos sinais patológicos comparativamente ao grupo de controlo para a vogal /a/. ....	80
Tabela 38 - Variação dos parâmetros dos sinais patológicos comparativamente ao controlo para a vogal /i/. ....	81
Tabela 39 - Variação dos parâmetros dos sinais patológicos comparativamente ao controlo para a vogal /u/. ....	82

# ABREVIATURAS E SÍMBOLOS

---

## Lista de abreviaturas

AMD	Amplitude Média Deslizante
C	Controlo
$F_a$	Frequência de amostragem
$F_0$	Frequência fundamental
$G(z)$	Função de transferência do impulso glotal
HNR	Harmonic to Noise Ratio
P1	Patologia Laringite
P2	Patologia Disfonia Hiperfuncional
P3	Patologia Disfonia Espasmódica
P4	Patologia Pólipo das Cordas Vocais
P5	Patologia Envelhecimento das Cordas Vocais
SVD	Saarbrücken Voice Database



# CAPÍTULO I

---

## 1. INTRODUÇÃO

O sistema mais elaborado da comunicação humana é a voz falada, pois esta tem a capacidade de transmitir informação a um nível que nenhum outro sistema é ainda capaz (Guimarães 2007).

A voz é um som produzido pelo fluxo de ar expelido pelos pulmões e pela ação do diafragma, passando pelas pregas vocais, sendo moldado quando influenciado pelas propriedades de reflexão e configuração do trato vocal (Lopes, 2008).

As perturbações na voz podem manifestar-se de várias formas, como alterações na qualidade vocal, frequência e intensidade da voz, assim como mudanças no funcionamento laríngeo, respiratório e/ou do trato vocal (Almeida 2010).

Existe uma grande variedade de técnicas utilizadas para avaliar a voz tais como a avaliação áudio-percetiva, avaliação perceptiva, endoscopia, estroboscopia, eletroglotografia e análise acústica.

A avaliação áudio-percetual consiste na avaliação de uma amostra vocal pelo profissional. Geralmente são utilizadas provas que implicam a sustentação de vogais, fala encadeada, entre outras, ficando registadas em áudio e/ou vídeo para posterior análise. Existem alguns problemas inerentes à utilização desta técnica tais como a baixa confiabilidade intra e inter-avaliadores e a não existência de uma escala de avaliação perceptual (Freitas, 2010).

A avaliação perceptiva consiste na apreciação das características sonoras da voz do falante, sendo tipicamente avaliada por terapeutas da fala ou médicos otorrinolaringologistas. Esta apreciação é geralmente resultado da fonação sustentada de uma vogal, em relação a referências perceptivas, adquiridas pelo especialista durante a sua formação ou exercício profissional de vozes consideradas normais. Esta avaliação é muito utilizada em ambiente clínico. Este tipo de avaliação apresenta uma grande limitação nas possíveis divergências na classificação entre avaliadores e até do mesmo avaliador em momentos diferentes.

A endoscopia é dividida em duas técnicas: a endoscopia rígida e a endoscopia flexível.

## I - Introdução

Na endoscopia rígida é utilizado um endoscópio de luz fria para proceder à visualização da laringe por via oral e tem como vantagem a possibilidade de se conseguir obter imagens amplas, estáveis e nítidas. Esta é uma técnica bastante invasiva, pelo que é necessário a aplicação de anestesia para inibir o reflexo de vômito.

Na endoscopia flexível é utilizado um fibroscópio flexível de via nasal, possibilitando a observação das fossas nasais, das cavidades faríngeas e da laringe. Com esta técnica é possível a fonação (voz sustentada, fala e canto), sendo possível o acoplamento a um computador permitindo a visualização num monitor, impressão e arquivo. Tem como desvantagens o facto de esta técnica ser bastante invasiva e a imagem aparecer normalmente escura e distorcida.

Na estroboscopia é utilizado o fenómeno fisiológico “persistência da visão”. Não é possível através da visão distinguir imagens individuais se estas forem apresentadas a velocidade superior a 5 imagens por segundo. Nesta técnica são usados “flashes” de luz com a mesma frequência de vibração das pregas vocais, sendo observada uma imagem clara e nítida. Este método utiliza a endoscopia, sendo desta forma um método invasivo.

A electroglotografia (EGG) baseia-se no facto do tecido humano ser condutor de electricidade, sendo assim uma técnica não invasiva. É utilizado um circuito elétrico, não prejudicial para a saúde humana, sendo possível analisar modificações da transmissão da corrente elétrica que resultam da mobilidade de estruturas como a mobilidade das pregas vocais. Apesar de esta técnica não interferir no processo de fala e não ser invasiva, só permite avaliar o sinal laríngeo, não sendo possível obter informação relativa ao trato vocal (Lopes, 2008).

A área da análise acústica tem-se desenvolvido nos últimos tempos no ponto de vista científico e no número de profissionais envolvidos no estudo (Guimarães, 2007). Sendo uma técnica não-invasiva a análise acústica é fundamentada no processamento digital do sinal de voz, um conjunto de ferramentas eficientes quando o objetivo é ajudar na identificação de patologias da voz ou rastreio de doenças vocais e de voz (Godino-Llorente & Gómez-Vilda, 2006).

O trabalho desenvolvido nesta área é maioritariamente baseado no uso de parâmetros acústicos como perturbação da amplitude e frequência assim como parâmetros de ruído (Arias-Londoño & Godino-Llorente, 2011).

Neste trabalho são utilizados os principais parâmetros acústicos usados na deteção de patologias, sendo estes o Jitter (Jitta, Jitt, Rap e ppq5), Shimmer (Shim, Shdb, Apq3

## I - Introdução

e Apq5) e o *Harmonic-to-Noise Ratio* (HNR) ou relação das componentes harmónicas com o ruído no sinal de fala.

Para a identificação destes parâmetros foi desenvolvido um algoritmo em Matlab utilizando como referência o trabalho desenvolvido em Oliveira & Lopes (2012). Antes de ser aplicado a sinais de fala o algoritmo foi testado utilizando sinais de voz sintetizados utilizando como referência o trabalho desenvolvido em Teixeira & Fernandes (2013), permitindo utilizar sinais de fala sintetizados com os parâmetros pré-definidos. Os resultados obtidos pelo algoritmo foram posteriormente comparados com dados obtidos analiticamente e com dados obtidos através do programa Praat. Este programa foi escolhido para ser usado como termo de comparação devido à quantidade de estudos realizados no âmbito da análise de sinais de fala em que é utilizado, por ser de distribuição gratuita e também foram consideradas as declarações dos autores em Boesma & Heuven (2001), afirmando ser o programa mais completo disponível para pesquisas fonéticas, utilizando os melhores algoritmos.

O algoritmo original sofreu várias alterações com o objetivo de proceder a uma análise correta dos sinais de voz provenientes da base de dados SVD, sendo comprovada a sua exatidão através da comparação dos valores obtidos pelo algoritmo com o programa Praat para sinais de voz de controlo e patológicos.

Posteriormente procedeu-se a uma análise estatística comparativa dos valores dos parâmetros obtidos pelo algoritmo entre vozes patológicas e de controlo.

No primeiro capítulo é feita uma introdução ao tema, onde são abordados alguns métodos alternativos à análise acústica e uma pequena descrição dos conteúdos da tese, seguindo-se uma revisão da literatura e alguns conceitos relativos ao funcionamento e anatomia do aparelho fonador.

No capítulo 2 descreve-se a base de dados utilizada nesta tese, é feita uma breve descrição das patologias utilizadas e do sinal sintetizado.

No capítulo 3 faz-se uma descrição da forma como foram determinados os parâmetros de fala. Inicialmente são descritos os parâmetros Jitter, Shimmer e HNR e respetivas equações. De seguida são apresentadas as variações produzidas aos parâmetros Jitter, Shimmer e  $F_0$ . Por fim é descrito o processo de desenvolvimento do algoritmo e apresentado o algoritmo final.

No capítulo 4 é feita uma apreciação das medidas do algoritmo, inicialmente através da comparação dos resultados deste com os resultados do programa Praat e cálculos analíticos após aplicadas as variações nos parâmetros descritas no capítulo 3 e

## I - Introdução

posteriormente para sinais de voz. Nesta apreciação de sinais de voz é feita uma comparação entre o algoritmo e o Praat usando sinais de voz de controlo e patológicos com o objetivo de validar o algoritmo.

O capítulo 5 compara os sinais do grupo de controlo com os sinais dos grupos patológicos usando os resultados medidos pelo algoritmo.

No capítulo 6 são apresentadas as conclusões finais e trabalhos futuros.

### 1.1. ESTADO DA ARTE

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

Em Bielałowicz et al, (1996), os autores utilizaram os processos “Cspeech”, “Computerized Speech Laboratory”, “SoundScope” e análise analítica para comparar os valores dos parâmetros  $F_0$ , Jitter, Shimmer e relação sinal ruído (SNR) em sinais de voz com disfonia. Estes concluíram que existe uma concordância nos programas comerciais apesar de não ser perfeita para as medições da frequência fundamental. Os autores apresentaram resultados inconclusivos relativamente às medidas de perturbação devido ao facto de os diversos programas utilizem algoritmos diferentes e fornecerem os dados em unidades diferentes produzindo muitas vezes valores em que os limites ciclo a ciclo não podem ser determinados com precisão.

Em Parraga (2002), é feita a classificação automática de vozes patológicas ou normais através da aplicação da Transformada Wavelet Packet e do algoritmo Best Basis. Neste, o autor demonstra que é possível classificar a voz usando este método, obtendo um erro de classificação de 23.07% para falsos positivos e de 14.58% para falsos negativos.

Em Wertzner et al (2005), é feito um estudo dos parâmetros Jitter, Shimmer, intensidade e  $F_0$  em sinais de voz de crianças comparando vozes com patologia e vozes de controlo. O método “Computer Speech Lab” foi utilizado para gravar os sinais de voz para as vogais /a/, /e/ e /i/ em português brasileiro, assim como para a análise dos parâmetros. Após a análise dos resultados os autores chegaram à conclusão que, para o parâmetro  $F_0$  a vogal /e/ apresenta valores mais baixos em relação às outras vogais para o grupo de vozes patológicas, enquanto que no grupo de pacientes de controlo todas as vogais apresentam valores de  $F_0$  semelhantes. Outra diferença encontrada entre os dois grupos foi a média da intensidade das vogais /a/, /e/ e /i/, que apresentou valores mais baixos para o grupo patológico. Quanto aos parâmetros Jitter e Shimmer, os autores não encontraram diferenças estatisticamente significativas entre os dois grupos.

## I - Introdução

Em Zwetsch et al (2006), é feita uma avaliação dos distúrbios da voz através da análise cepstral utilizando o programa MatLab. Os autores apresentam uma tabela com o perfil cepstrográfico das várias situações, e afirmam ter os parâmetros agrupados numa outra tabela, com os quais “é possível diagnosticar de forma coerente as alterações, tomando como base os achados cepstrais de diferentes vocalizações”. Apesar de não existir mais informação em relação a essa tabela os autores afirmam ter diagnosticado corretamente as patologias com uma taxa de acerto de 80%.

Em Lopes (2008), foi desenvolvido uma aplicação de análise e diagnóstico da voz denominada de SEEGNAL VoiceStudio. Esta ferramenta faz análise aos parâmetros  $F_0$ , Jitter, Shimmer e HNR. A verificação da aplicação foi efetuada através da medição dos parâmetros para vozes sintéticas e posteriormente comparação dos resultados com os programas Dr Speech e Praat. Para todos os parâmetros a aplicação desenvolvida apresentou resultados consistentes com os outros programas.

Em Silva (2010), é feito um estudo à deterioração da voz resultante do processo de envelhecimento. Neste estudo os pacientes foram separados consoante as idades e dispostos em grupos denominados de jovens, adultos e idosos. A obtenção dos dados foi efetuada através do programa Praat para os parâmetros  $F_0$ , Jitter, Shimmer e HNR e do programa Aparat para 14 parâmetros. Após a extração dos dados o autor implementou uma rede neuronal de modo a identificar automaticamente os casos de vozes envelhecidas.

Em Almeida (2010), é produzido um sistema de classificação de patologias da laringe aplicando às características dos sinais coeficientes de Predição Linear, Coeficientes Cepstrais de Frequência Mel e coeficientes obtidos através da Transformada Wavelet Packet, sendo a classificação efetuada através de Máquinas de Vetor de Suporte que otimiza as margens de separação entre as respetivas classes criando um hiperplano. De acordo com o autor este sistema apresenta uma taxa de acerto de 98.46% na classificação entre vozes normais e patológicas e de 98.75% na classificação da patologia.

Em Teixeira et al (2011), é implementado um algoritmo com o objetivo de determinar a  $F_0$  de um sinal de voz através do método do Cespstro e pelo método da autocorrelação, sendo posteriormente avaliados os valores através da comparação com os resultados obtidos pelo programa Praat. É também produzido um algoritmo de identificação dos parâmetros Jitter e Shimmer através do programa MatLab para um sinal.

## I - Introdução

Em Oliveira & Lopes (2012), são implementados algoritmos em MatLab com o objetivo de identificar os parâmetros Jitter, Shimmer e HNR para dois sinais da vogal /a/ de géneros diferentes. O programa Praat é utilizado de forma a verificar os valores obtidos pelo algoritmo que obteve resultados satisfatórios para todos os parâmetros nos sinais de voz analisados.

Em Teixeira & Gonçalves (2014), é feita uma análise à qualidade de medição dos parâmetros Jitter e Shimmer de um algoritmo produzido durante a elaboração deste trabalho em sinais sintetizados com variações controladas. Os resultados foram posteriormente comparados com o programa Praat, sendo os resultados do algoritmo em alguns casos mais precisos.

## 1.2. FUNCIONAMENTO E ANATOMIA DO APARELHO FONADOR

Considerada a principal ferramenta de comunicação, a voz humana é uma emissão acústica voluntária, tendo inúmeros fatores envolvidos na sua produção. É necessária a existência de sincronia entre os diferentes sistemas e órgãos (Figura 1), em que cada um deles tem diferentes funções, para que haja a produção de um sinal de voz (Dajer, 2010).

No decorrer do processo de expiração existe uma maior variação de pressão do que no processo de inspiração, sendo produzidas ondas sonoras, que moldadas pela laringe e pelas cavidades superiores orais e nasais, associando características à voz.

No processo de expiração o fluxo de ar segue em trajeto invertido ao do ar no processo de inspiração, a propulsão de ar pelos pulmões, por via dos brônquios, penetra na traqueia, atravessa a laringe, faringe e chega à cavidade nasal e/ou oral (Teixeira, 1995; Cunha & Cintra, 2010).

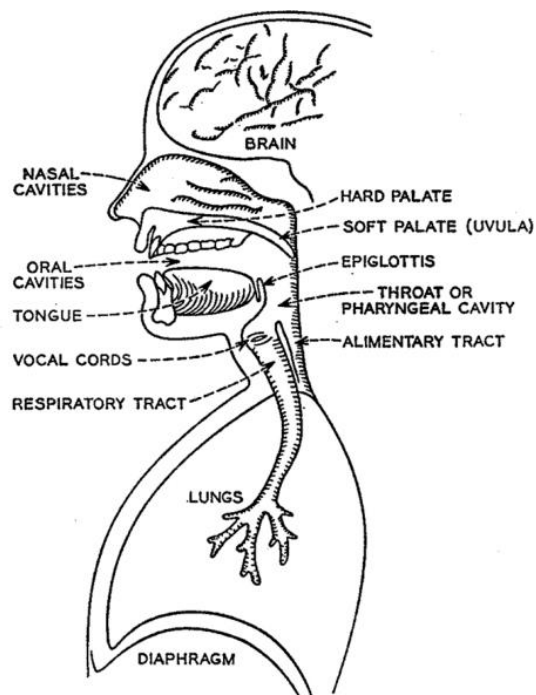


Figura 1 – Representação das principais estruturas que contribuem para o mecanismo de fala.

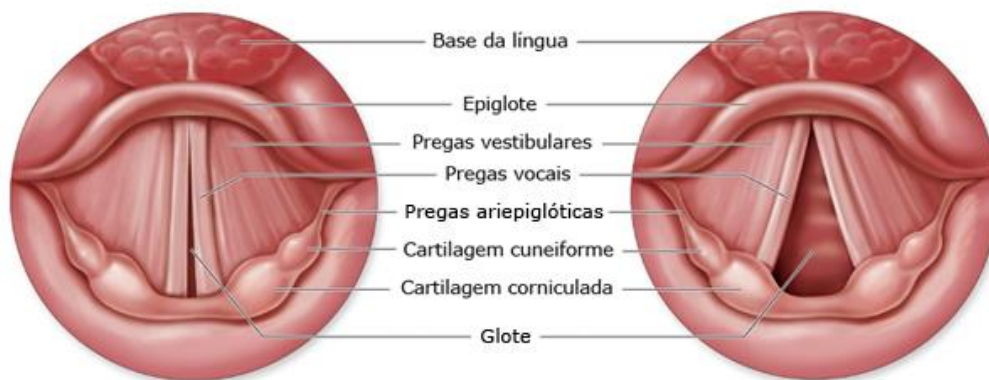
(Fonte: <http://imgarcade.com/1/speech-production/>)

Depois de conduzido para fora dos pulmões, o ar atravessa a traqueia e chega a laringe (Almeida, 2010). A laringe é “um sistema músculo-cartilágneo e neuromuscular com funções complexas que envolvem a proteção das vias aéreas inferiores, a

## I - Introdução

respiração, a deglutição e a comunicação oral”. Coordenada com a faringe e o esófago, na laringe é definido o evento que irá ocorrer. Este será respiração caso seja apenas passagem de ar, deglutição caso seja transferência de alimentos ou comunicação oral (voz) (Guimarães, 2007).

No interior da laringe situam-se as pregas vocais (Figura 2), estas estruturas que são constituídas por ligamentos e músculos, são normalmente conhecidas por cordas vocais. Na zona superior situam-se as cordas ventriculares, também chamadas de “falsas cordas vocais” por não terem função de fonação e na parte inferior situam-se as cordas vocais. No espaço entre as cordas ventriculares e as cordas vocais encontra-se o ventrículo. O espaço compreendido entre as cordas vocais é chamado de glote (Teixeira, 1995).



**Figura 2 - Representação das Pregas Vocais.**

(Fonte: [http://www.lookfordiagnosis.com/mesh\\_info.php?term=Glote&lang=3](http://www.lookfordiagnosis.com/mesh_info.php?term=Glote&lang=3))

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

Num ciclo vibratório (Figura 3), para o modo de fonação (também designado de voz modal), as pregas vocais estão próximas e começam a afastar-se (de baixo para cima e de dentro para fora) até ao ponto de abertura máxima. Este afastamento das cordas vocais ocorre porque a pressão do ar subglótico vence a sua força de resistência. Consoante o ar sobe, a pressão subglótica diminui e a elasticidade das pregas vocais provoca o movimento de adução das pregas vocais e encerramento da glote de baixo para cima. O final da fase de encerramento (total ou parcial) dá-se quando a pressão subglótica força novamente a resistência das pregas vocais, iniciando-se assim um novo

## I - Introdução

ciclo vibratório. O número de vezes por segundo que cada ciclo glótico é repetido corresponde à frequência fundamental (Guimarães, 2007).

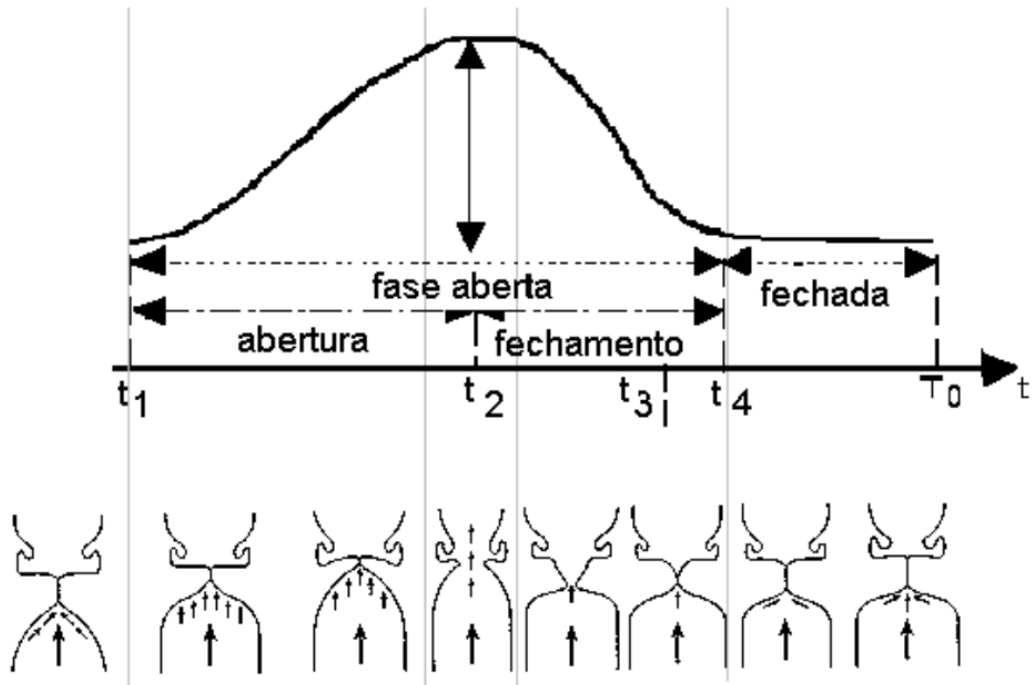


Figura 3 - Processo de abertura e fecho das cordas vocais correspondente a um período glotal (Dajer, 2010).

Após a passagem do ar pela laringe segue-se a faringe, que é um meio cilindro de músculo e fásia, fixado acima à base do crânio e abaixo às margens do esôfago. A faringe tem três partes distintas, a parte laríngea da faringe, a parte oral e a parte nasal da faringe. Às partes posteriores à cavidade oral e à laringe, correspondem a parte oral e parte laríngea da faringe respectivamente. A parte nasal da faringe encontra-se na zona da faringe posterior às cavidades nasais. Entre a parte oral da faringe e a nasal existe o véu palatino que separa estas duas partes (Teixeira, 1995; Drake et al, 2005).

A faringe e as cavidades nasais e oral formam o conjunto das cavidades supraglóticas, que têm um papel fundamental na fonação de diferentes sons. Consoante é alterada a forma e as dimensões das cavidades supraglóticas quando sujeitas a uma fonte sonora pode-se produzir uma grande variedade de sons. Nas cavidades supraglóticas existe a formação de uma espécie de ressoadores que favorecem a passagem de algumas frequências, e a atenuação de outras em função das suas formas e dimensões. As frequências que são favorecidas pelas cavidades supraglóticas dá-se o nome de frequências formantes e ao conjunto das formas tomadas pelas cavidades supraglóticas trata-se de trato vocal (Teixeira, 1995).

## CAPÍTULO II

---

### 2. BASE DE DADOS DOS SINAIS DE FALA

Para a realização deste estudo foi utilizada a base de dados Saarbrücken Voice Database (SVD). Foi utilizada esta base de dados pois apresenta sinais de voz de locução de uma vogal de forma sustentada no mesmo tom por um período de tempo de alguns segundos, que é condição essencial para a análise pretendida neste trabalho e por ser de acesso gratuito.

A SDV possui um repositório de sinais de voz pertencentes a mais de 2000 indivíduos. Para cada indivíduo é disponibilizado um ficheiro de som isolado referente a gravação de cada vogais /a/, /i/ e /u/ com tons baixo, normal, alto, variando entre tons e a gravação da frase "Guten Morgen, wie geht es Ihnen?" ("Bom dia, como estás?") em alemão (Pützer & Barry). Os ficheiros de som foram registados com uma frequência de amostragem de 50 kHz. A esta frequência corresponde um período de amostragem de 20  $\mu$ s.

Para os testes realizados com o objetivo de validar o algoritmo foram utilizados sinais provenientes de 30 indivíduos saudáveis, os quais foram admitidos como controlo, e 30 indivíduos com a patologia Laringite.

Para os testes realizados com o objetivo comparar os sinais do grupo controlo com os sinais dos grupos patológicos foram utilizados 22 sinais de voz controlo, e 22 sinais de voz para cada uma das patologias, sendo estas Laringite, Disfonia Hiperfuncional, Disfonia Espasmódica, Pólipo das Cordas Vocais e Envelhecimento das Cordas Vocais. As patologias associadas aos pacientes em estudo estão descritas abaixo.

Para cada paciente foi usado um sinal de voz das vogais /a/, /i/ e /u/ com os tons baixo, medio/normal e alto, fazendo um total de 9 sinais por paciente.

## 2.1. SINAIS PATOLÓGICOS

De forma a tornar este estudo com alguma relevância estatística, só foram utilizados sinais de patologias que apresentassem mais de 20 casos registados e que estivessem relacionados com um possível pré-rastreio clínico. Os casos patológicos em que a presença de patologias era nitidamente notório, é exemplo os casos em que os pacientes foram submetidos a laringotomia, foram excluídos deste trabalho.

De seguida apresenta-se uma descrição breve sobre as patologias dos pacientes usados neste estudo.

### 2.1.1. LARINGITE

As Laringites (Figura 4) são as disfunções que mais afetam a laringe, sendo caracterizadas como inflamações da laringe e áreas próximas. Estas podem ocorrer através de uma manifestação única de insulto viral, bacteriano ou químico, mas é mais comum partir de uma infeção generalizada do trato respiratório superior ou o resultado de exposição maciça ao fumo de tabaco. A laringe pode também ser afetada com muitas doenças infecciosas sistêmicas, como a tuberculose e a difteria. A maioria das infeções é autolimitada, contudo elas podem ser por vezes graves, especialmente na infância (Robbins et al, 2005).



Figura 4 – Laringite na parte posterior.

(Fonte: <http://www.clinicacoser.com/veja-fotos-de/laringoscopia/#11>)

### **2.1.2. DISFONIA HIPERFUNCIONAL**

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

### **2.1.3. DISFONIA ESPASMÓDICA**

A disfonia espasmódica é “um raro distúrbio da voz causado por uma distonia focal dos músculos da laringe”. É mais frequente no sexo feminino e tem início tipicamente por volta dos 30 anos. As causas que provocam esta disfonia ainda não são bem compreendidas (Coelho et al, 2010).

### **2.1.4. PÓLIPO DAS CORDAS VOCAIS**

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

### **2.1.5. ENVELHECIMENTO DAS CORDAS VOCAIS**

O envelhecimento das estruturas da laringe é um acontecimento que acompanha o envelhecimento das pessoas. As alterações fisiológicas e anatómicas do trato vocal decorridas aquando do processo de envelhecimento são geralmente subestimadas, apesar de ser essencial para a compreensão de algumas patologias comuns na população idosa.

As alterações mais proeminentes a nível anatómico e fisiológico são a calcificação e ossificação das cartilagens laríngeas, que causam diminuição da mobilidade laríngea e a diminuição da maior parte da musculatura das cordas vocais (Miranda et al, 2011).

## 2.2. SINAL SINTETIZADO

Na produção deste sinal foi utilizado o módulo acústico do sintetizador de fala didático (Teixeira & Fernandes, 2013). O sintetizador foi desenvolvido segundo o modelo de formantes de Klaat. Este modelo simula a função de transferência do trato vocal por uma conexão de circuitos ressoadores em série ou em paralelo, com as respectivas frequências e larguras de banda usadas como parâmetros de entrada no sintetizador (Teixeira & Fernandes, 2013; Klatt, 1987).

Este sinal usa o modelo de terminais análogos como demonstrado na Figura 5.

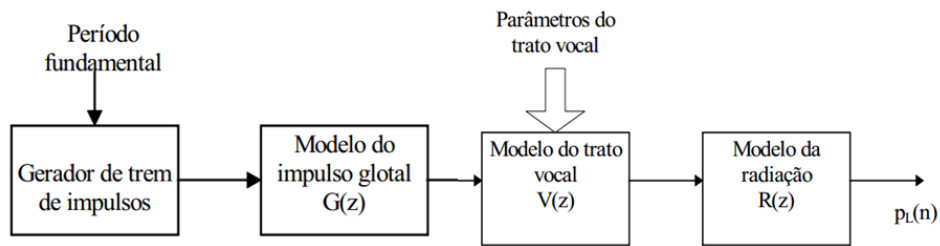


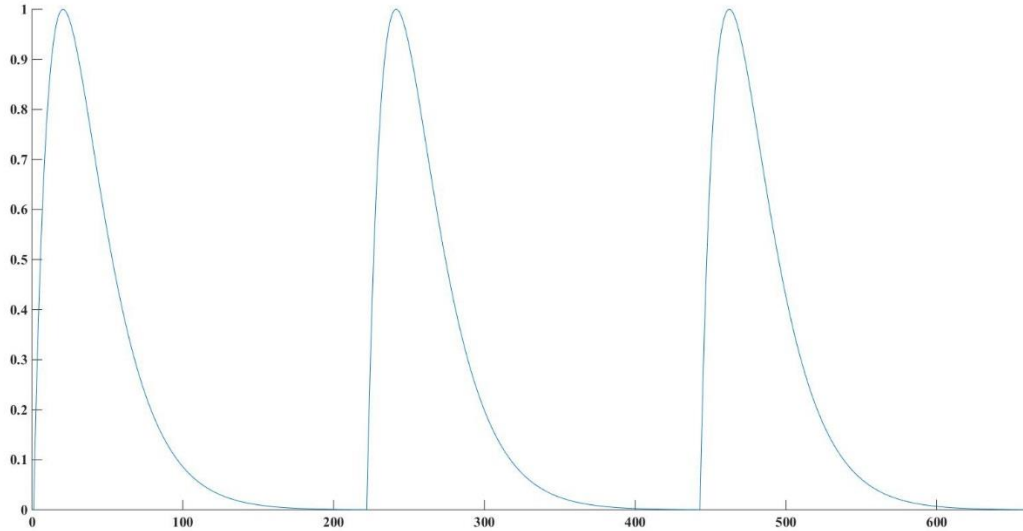
Figura 5 - Modelo genérico para a produção de fala. (alterado de Teixeira, 1995)

Esta é a aproximação mais usada para representar sinais de fala, em que a saída tem as propriedades desejadas da fala quando é feito um controlo dos parâmetros relacionados com o processo de produção de fala (Teixeira, 1995).

A sequência de impulsos glotais (Figura 6) é produzida pela aplicação da equação 1 à sequência de impulsos unitários com um intervalo de tempo correspondente ao período glotal.

$$G(z) = \frac{-a \ln(a) z^{-1}}{(1 - az^{-1})^2} \quad (1)$$

## II – Base de dados dos sinais de fala



**Figura 6- Ilustração de três impulsos glotais após a aplicação da equação 1 à sequência de impulsos unitários.**

Os impulsos glotais foram gerados com o parâmetro  $a=0.9$  (Teixeira & Fernandes, 2013, Fernandes, 2012), de forma a que cada impulso glotal chegue a um valor de amplitude quase nulo não interferindo com o impulso seguinte. Isto permite uma maior flexibilidade para variar a duração dos impulsos glotais sem perda de qualidade da fala sintetizada.

Para o trato vocal foi utilizada a equação 2, em que  $V(z)$  é a função de transferência,  $Z_k$  está diretamente relacionado com a largura de banda das formantes e  $F_k$  é o valor da frequência do formante  $k$  (Fernandes, 2012).

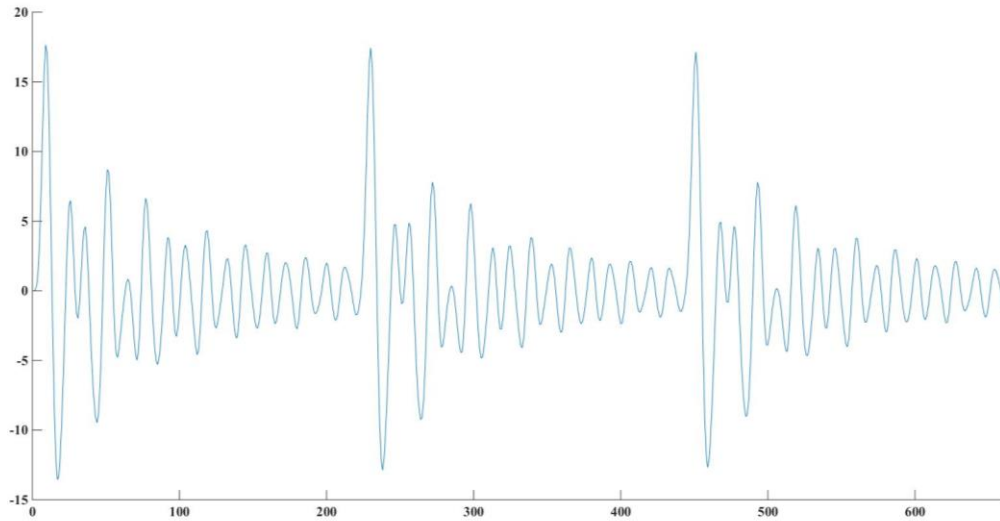
$$V_k(z) = \frac{(1 - 2|z_k| \cos(2\pi F_k T) + |z_k|^2)}{(1 - 2|z_k| \cos(2\pi F_k T) z^{-1} + |z_k|^2 z^{-2})} \quad (2)$$

A radiação dos lábios (equação 3), é utilizada quando se pretende um modelo para a pressão de ar nos lábios (Teixeira & Fernandes, 2013, Fernandes, 2012).

$$R(z) = R_0(1 - z^{-1}) \quad (3)$$

## II – Base de dados dos sinais de fala

Para o sinal sintetizado foi utilizada uma frequência de amostragem de 22050 Hz com uma frequência fundamental ( $F_0$ ) de aproximadamente 100 Hz, o que corresponde a períodos glotais de aproximadamente 10 ms como ilustrado na Figura 7. Neste estudo utilizaram-se os sinais sintetizados com formantes e larguras de banda correspondentes à vogal /a/ (Teixeira & Fernandes, 2013, Fernandes, 2012).



**Figura 7 - – Ilustração de três impulsos glotais do sinal sintetizado.**

## CAPÍTULO III

### 3. DETERMINAÇÃO DOS PARÂMETROS DE FALA

#### 3.1. INTRODUÇÃO

Nesta secção é feita a descrição dos parâmetros utilizados para o desenvolvimento deste trabalho. Utilizaram-se os parâmetros mais empregados em análise vocal, que são o Jitter, Shimmer e HNR

Na Fig. 8 ilustra-se o conceito de Jitter e de Shimmer como sendo o Jitter uma medida da variação da duração dos períodos glotais e o Shimmer como uma variação da amplitude dos sucessivos períodos glotais

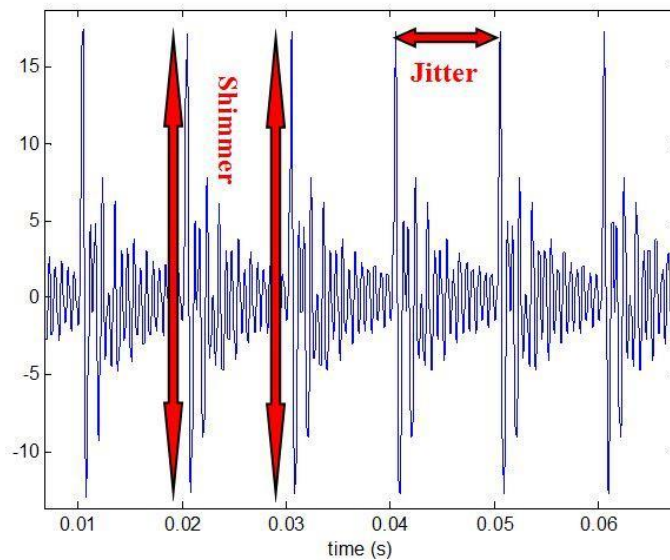


Figura 8- Representação dos parâmetros Jitter e Shimmer para um sinal de fala.

#### 3.1.1. JITTER

O Jitter é uma medida de termo curto, isto é, apresenta variabilidade não voluntária na  $F_0$  ciclo a ciclo. O Jitter sofre alterações principalmente quando há falta de controlo de vibração das cordas vocais (Teixeira et al, 2011).

As vozes dos pacientes com patologia geralmente têm valor de Jitter maiores.

### III – Determinação Dos Parâmetros de Fala

**Jitter (local):** média absoluta da diferença entre períodos consecutivos, dividido pelo período médio, em percentagem e denominado de Jitt.

$$Jitt = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (4)$$

Onde  $T_i$  é a duração do período glotal  $i$ , e  $N$  é o número de períodos glotais.

**Jitter (local, absoluto):** média absoluta da diferença entre dois períodos consecutivos, em segundos ou  $\mu s$  e denominado de Jitta.

$$Jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \quad (5)$$

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

$$rap = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| T_i - \left( \frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (6)$$

**Jitter (ppq5):** o quociente de perturbação num intervalo de cinco períodos (five-point Period Perturbation Quotient) é a média absoluta entre um período e a média desse com os seus quatro vizinhos, dividido pelo período médio, em percentagem.

$$ppq5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} \left| T_i - \left( \frac{1}{5} \sum_{n=i-2}^{i+2} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (7)$$

### 3.1.2. SHIMMER

O Shimmer (Figura 8), mede as variações da amplitude do sinal com base em cada ciclo fonatório. O Shimmer é inversamente proporcional à intensidade vocal média, ou seja, quanto maior a intensidade menor o valor de Shimmer (Guimarães, 2007). É esperado que os pacientes com voz patológica tenham maior valor de Shimmer.

O Shimmer aumenta com a redução da resistência glotal e com as lesões na massa muscular das cordas vocais e está correlacionado com voz sussurrada e fala suspirada.

**Shimmer (local):** média absoluta da diferença das amplitudes de dois períodos consecutivos, dividido pela amplitude média, em percentagem.

$$Shim = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (8)$$

Onde  $A_i$  é a amplitude do período glotal  $i$  e  $N$  o número de períodos.

**Shimmer (local, dB):** média absoluta do logaritmo de base 10 do quociente de amplitudes de dois períodos consecutivos, multiplicado por 20, e numa escala em decibel (dB)

$$ShdB = \frac{1}{N-1} \sum_{i=1}^{N-1} |20 * \log(A_{i+1}/A_i)| \quad (9)$$

**Shimmer (apq3):** o quociente de perturbação da amplitude de três pontos (three-points Amplitude Perturbation Quocient) é a média absoluta da diferença entre a amplitude de um período e a média das amplitudes dos dois vizinhos, dividido pela amplitude média, em percentagem.

$$apq3 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| A_i - \left( \frac{1}{3} \sum_{n=i-1}^{i+1} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (10)$$

**Shimmer (apq5):** o quociente de perturbação da amplitude de cinco pontos (five-points Amplitude Perturbation Quocient) é a média absoluta da diferença entre a amplitude de um período e a média das amplitudes dos quatro vizinhos, dividido pela amplitude média, em percentagem.

$$apq5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} \left| A_i - \left( \frac{1}{5} \sum_{n=i-2}^{i+2} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (11)$$

### 3.1.3. HNR

Por definição o HNR (Harmonic to Noise Ratio) é a medida que quantifica a relação entre a componente harmónica (componente periódica) e a componente de ruído (componente aperiódica) de um segmento de voz vozeada, expresso em dB (Guimarães, 2007). Esta unidade de medida relaciona a energia concebida através do sinal de voz por intermédio dos impulsos glotais, e a energia da fração de ruído glótico após serem filtrados pelo trato vocal. Este ruído resulta da turbulência gerada pela passagem do fluxo de ar através da glote durante o processo de fonação, ocorrendo quando por exemplo existe um fechamento inadequado das cordas vocais (Silva, 2010).

Traduzindo um sinal vozeado com estrutura harmónica no domínio das frequências para termos matemáticos, este pode ser representado pela equação 12.

$$X(w) = H(w) + N(w) \quad (12)$$

Onde  $X(w)$  corresponde ao sinal de voz no domínio das frequências,  $H(w)$  à componente harmónica e  $N(w)$  à componente de ruído.

Como o HNR por definição é uma medida logarítmica da relação das energias associadas à componente harmónica e de ruído, compreende-se que a integração da potência espectral ao longo da gama audível de frequências seja possível como demonstrado na equação 13 (Lopes, 2008).

$$HNR = 10 \times \log_{10} \frac{\int_w |H(w)|^2}{\int_w |N(w)|^2} \quad (13)$$

### III – Determinação Dos Parâmetros de Fala

Para este estudo o parâmetro HNR foi implementado com base nos estudos publicados em Boersma, 1993. Neste o autor utiliza um procedimento baseado nas propriedades da função autocorrelação, para obter a separação de componentes atrás descritas. A autocorrelação consiste na correlação de um sinal com ele mesmo. Considerando um sinal de voz  $x(t)$ , a função autocorrelação  $r_x(\tau)$  é representada pela equação 14.

$$r_x(\tau) \equiv \int x(t)x(t + \tau) dt \quad (14)$$

Esta função tem um máximo global para  $\tau=0$ . Se existirem valores globais máximos fora de 0, o sinal é periódico e existe um desfasamento  $T_0$ , chamado período, de modo a que todos esses máximos sejam colocados no desfasamento  $nT_0$ , para cada inteiro  $n$ , com  $r_x(nT_0)=r_x(0)$ . A frequência fundamental  $F_0$  deste sinal periódico é definida pela equação 15.

$$F_0 = \frac{1}{T_0} \quad (15)$$

Se não houver máximos globais fora de 0, ainda pode haver máximos locais. Se o maior deles for no desfasamento  $\tau_{\max}$ , e se a sua altura  $r_x(\tau_{\max})$  for suficiente, o sinal é designado como tendo uma parte periódica, e a sua força harmónica  $R_0$  é um número entre 0 e 1, igual ao máximo local  $r'_x(\tau_{\max})$  da autocorrelação normalizada (equação 16).

$$r'_x(\tau) \equiv \frac{r_x(\tau)}{r_x(0)} \quad (16)$$

A autocorrelação do sinal total é igual à soma das autocorrelações das suas componentes harmónica e de ruído como demonstrado na equação 17.

$$r_x(0) = r_H(0) + r_N(0) \quad (17)$$

### III – Determinação Dos Parâmetros de Fala

Se o ruído for branco (não é possível a correlação com o mesmo), encontra-se o máximo local em  $\tau_{max}=T_0$  com a altura  $r_x(\tau_{max})=r_H(T_0)=r_H(0)$  (Boersma, 1993).

Ou seja, a função autocorrelação de um sinal de voz sustentada exibe máximos locais para valores múltiplos de  $\tau$ , múltiplos inteiros do período fundamental. Assim sendo, para determinar o HNR basta calcular a função autocorrelação do sinal de voz, identificar o primeiro máximo local que será correspondente à componente harmónica. O valor da componente de ruído determina-se usando a equação 17. O HNR calcula-se através da equação 18 (Lopes, 2008).

$$HNR (em dB) = 10 * \log_{10} \frac{r'_x(\tau_{max})}{1 - r'_x(\tau_{max})} \quad (18)$$

### **3.2. VARIAÇÃO DE JITTER E SHIMMER**

Através do sinal sintetizado foi possível obter sinais sem perturbação, ou seja, sem qualquer tipo de variação entre períodos e consequentemente com Jitter e Shimmer iguais a zero. Estes sinais são gerados com períodos glotais sucessivos com exatamente a mesma duração.

Com o objetivo de testar os valores obtidos pelo algoritmo foram produzidos sinais com diferentes tipos de perturbação através do sinal sintetizado. Para cada um destes parâmetros foram criados dois tipos de perturbação conforme está descrito abaixo.

#### **3.2.1. VARIAÇÃO DO JITTER**

O Jitter define-se como sendo a variação de um ciclo de vibração das pregas vocais para o ciclo seguinte (Lopes, 2008). Assim sendo, para proceder a uma implementação de perturbação no sinal sintetizado e influenciando somente o parâmetro Jitter, foram alterados os períodos do sinal sintetizado de duas formas diferentes. A primeira foi designada de perturbação de Jitter tipo 1 e consiste num trem de impulsos com um período diferente a cada dois períodos consecutivos e a segunda designada de perturbação de Jitter tipo 2 consiste num trem de impulsos com um período diferente a cada três períodos consecutivos. Cada um deste trem de impulsos é submetido ao modelo de impulsos glotais apresentado na Eq. 1, para produzir os impulsos glotais e gerar o sinal sintetizado como descrito na secção 2.2.

Para a produção destes sinais a frequência de amostragem ( $F_a$ ) utilizada foi de 22050Hz, com frequência fundamental ( $F_0$ ) de 100Hz, a que correspondem períodos glotais de 221 amostras ( $F_a/F_0$ ).

##### **3.2.1.1. PERTURBAÇÃO DE JITTER TIPO 1**

Este tipo de perturbação corresponde a um par de períodos glotais diferentes sucessivos como demonstrado na Figura 9.

Para este tipo de perturbação tentou-se produzir uma variação de 5% do Jitt, o que corresponde a uma diferença de 11 amostras entre períodos sucessivos. Portanto definiu-se que  $T_0'=210$  e  $T_0''=221$  amostras.

Aplicado à equação 5 fica:

$$Jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| = \frac{|(221 - 210)|}{22050} = 0.49 \times 10^{-3}$$

Na operação anterior a diferença das amostras é dividida pela frequência de amostragem (22050 Hz), com o intuito de passar o número de amostras para tempo.

A média de cada período glotal corresponde a 215.5 amostras que aplicado à equação 19 obtém-se um valor de período fundamental ( $T_0$ ) de 9.8 ms.

$$T_0 = \frac{N_0}{F_0} \quad (19)$$

Empregando este valor de Jitta à equação 4 obtém-se um valor de Jitt de 5.09% como demonstrado em seguida:

$$Jitt = \frac{jitta}{T_0} * 100 = 5.09\%$$

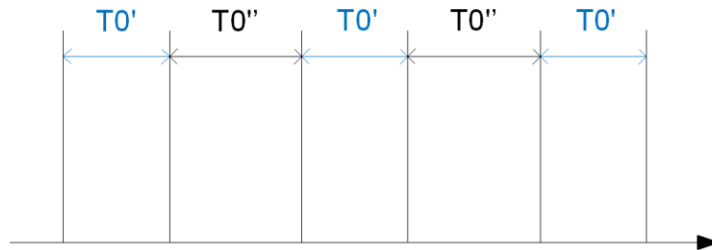


Figura 9 - Perturbação de Jitter tipo 1 com variação de um em um período glotal.

### 3.2.1.2. PERTURBAÇÃO DE JITTER TIPO 2

De forma a testar o comportamento do parâmetro Jitter para variações irregulares procedeu-se também a uma variação de três em três períodos glotais, em vez de um em um período glotal, utilizando a mesma variação utilizada na perturbação de Jitter de tipo 1 como demonstrado na Figura 10.

### III – Determinação Dos Parâmetros de Fala

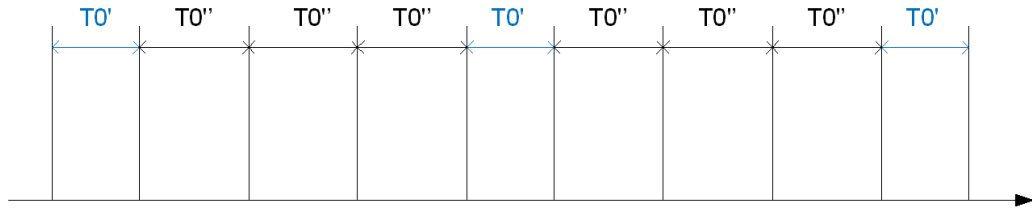


Figura 10 - Perturbação de Jitter tipo 2 com variação de três em três períodos glotalis.

Aplicado este tipo de perturbação à equação 5 fica:

Sendo:

$$|T_i - T_{i-1}| = \frac{|(221 - 210)| + |(221 - 221)| + |(221 - 221)| + |(210 - 221)|}{4} = 5.5$$

$$Jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| = \frac{5.5}{22050} = 2.49 \times 10^{-4}$$

Em que 22050 é a frequência de amostragem.

Para esta situação a média de cada período glotal é de 218.25 amostras, resultando um  $T_0=9.9$  ms quando este valor é aplicado à equação 13

Empregando o valor de Jitta à equação 4 com o valor de  $T_0$  resultante da equação 13 obtém-se um valor de Jitt de 2.52% como explicado em seguida:

$$Jitt = \frac{Jitta}{T_0} * 100 = 2.52\%$$

Como demonstrado com este tipo de variação foi produzida uma variação de 2.52% de Jitt.

### 3.2.2. VARIAÇÃO DE SHIMMER

Sendo o Shimmer a medida que quantifica as alterações mínimas da amplitude do sinal (Guimarães, 2007), as alterações ao sinal sintetizado foram feitas de forma a produzir alterações na amplitude, mas mantendo o período glotal original.

Para implementar a perturbação no Shimmer procedeu-se à variação da amplitude do sinal de duas formas diferentes, uma foi designada de perturbação de Shimmer do tipo 1 que corresponde a variar a amplitude de um em um período glotal e à outra perturbação de Shimmer do tipo 2 que corresponde a variar a amplitude de três em três períodos glotais.

#### 3.2.2.1. PERTURBAÇÃO DE SHIMMER DO TIPO 1

Para esta perturbação foram produzidas alterações ao sinal sintetizado com o objetivo de produzir uma variação de Shim de 25% por período glotal como demonstrado na Figura 11. As amplitudes A0' e A0'' correspondem a impulsos de 1 e 1.25 respetivamente.

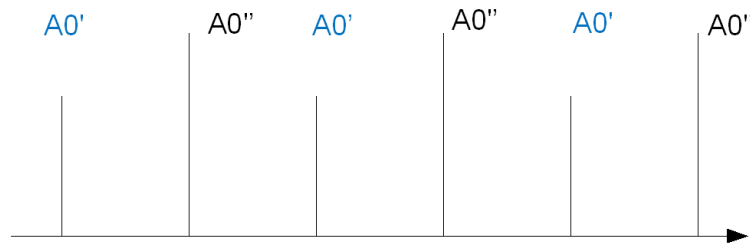


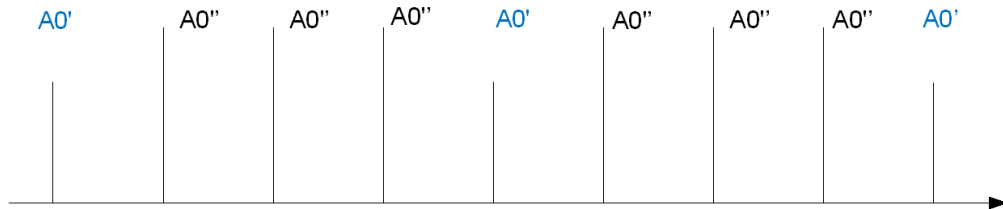
Figura 11 - Perturbação de Shimmer do tipo 1 com variação da amplitude de um em um período glotal.

Este valor de Shim surgiu após a aplicação dos valores correspondentes aos impulsos deste tipo de perturbação à equação 8 como demonstrado em seguida:

$$Shim = \frac{1}{N-1} \frac{\sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 = \frac{|1 - 1.25|}{\frac{|1 + 1.25|}{2}} * 100 = 22.22\%$$

**3.2.2.2. PERTURBAÇÃO DE SHIMMER DO TIPO 2**

Foi também produzida uma variação da amplitude, em que esta varia de três em três períodos glotais, de forma a testar o comportamento do parâmetro shimmer como demonstrado na Figura 12. As amplitudes A0' e A0'' correspondem a impulsos de 1 e 1.25 respectivamente.



**Figura 12 - Perturbação de Shimmer do tipo 2 com variação da amplitude de três em três períodos glotais.**

Tal como anteriormente, foi determinado a variação de Shim correspondente a este tipo de perturbação aplicando a equação 8 como está demonstrado em seguida:

Sendo:

$$|A_i - A_{i+1}| = \frac{|(1 - 1.25)| + |(1.25 - 1.25)| + |(1.25 - 1.25)| + |(1.25 - 1)|}{4}$$

$$= 0.125$$

$$Shim = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 = \frac{|1 - 1.25|}{\frac{1 + 3 * 1.25}{4}} * 100 = 10.53\%$$

### 3.2.3. VARIAÇÃO DA FREQUÊNCIA FUNDAMENTAL

Após ser testado o comportamento de Jitter e Shimmer com algumas variações, decidiu-se testar a influência do parâmetro frequência fundamental. Com esta finalidade foram produzidos sinais de fala sintetizada com  $F_0$  igual a 75Hz, 100Hz e 190Hz.

A  $F_0$  pode influenciar o Shimmer no sinal sintetizado porque frequências fundamentais altas resultam em períodos glotalis mais curtos, e como o modelo de formantes é um filtro com resposta impulsional infinita (IIR) o comprimento da resposta do impulso é superior ao período glotal. Consequentemente, a influência da amplitude do período anterior no período seguinte é maior em períodos glotalis mais curtos, ou em frequências fundamentais superiores. Depois de um certo valor de período glotal esta influência não é significativa.

Não é esperado qualquer tipo de alteração no parâmetro Jitter com a variação da  $F_0$ .

Os resultados destas variações são apresentados e discutidos na secção 4.1.3.

### 3.3. ALGORITMO DE IDENTIFICAÇÃO DOS PERÍODOS GLOTAIS

Nesta secção apresenta-se o algoritmo que identifica o instante inicial de cada ciclo glotal do sinal de fala. Este instante é considerado como sendo o valor de maior magnitude de cada ciclo glotal. Uma vez identificado o instante de tempo e a amplitude de cada início de ciclo glotal, as equações 4 a 11 permitem fazer a determinação das 4 medidas de Jitter e de Shimmer.

A identificação do início dos períodos glotalis num sinal de fala sintetizado ou em alguns sinais de fala saudável não oferece dificuldades de maior, pois como se pode ver na Figura 13 os períodos glotalis são, normalmente, bem definidos.

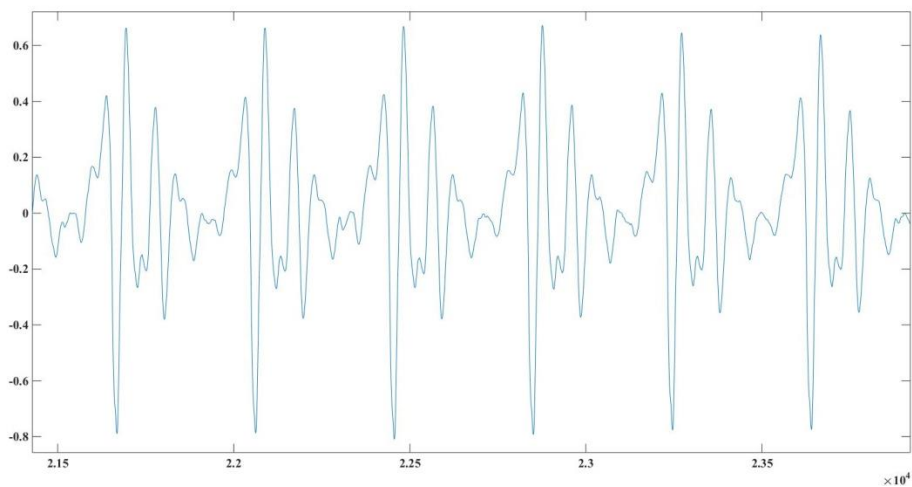
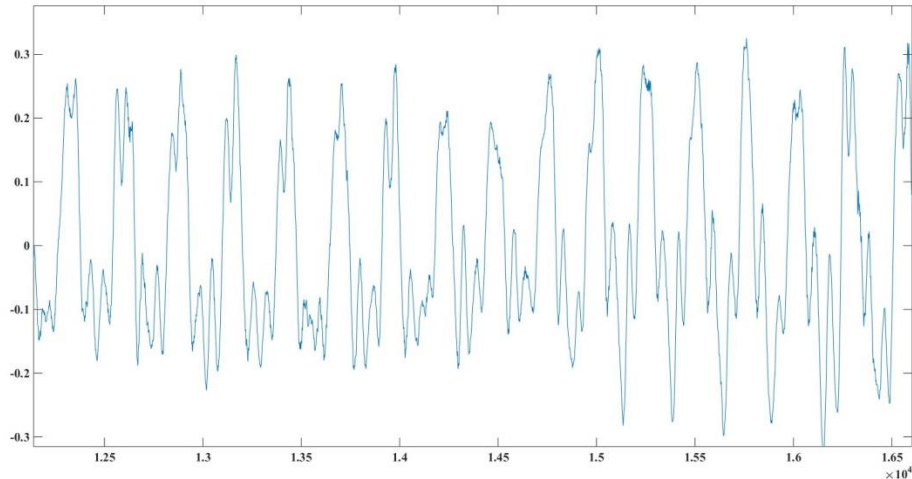


Figura 13 - Sinal de voz controlado da vogal /a/ tom normal, onde os períodos glotalis estão bem definidos.

Contudo em vozes patológicas ou em alguns sinais de vozes saudáveis essa tarefa pode ser bastante mais complexa pelo facto dos ciclos não serem tão bem definidos, como se ilustra na Figura 14, ou porque a maior magnitude de um mesmo sinal surge ora na fase positiva da onda ora na fase negativa, levando a uma errada interpretação do início dos ciclos glotalis que torna erradas as medidas de Jitter.

### III – Determinação Dos Parâmetros de Fala



**Figura 14 - Sinal de voz com a patologia Laringite da vogal /a/ tom normal, onde os períodos glotais não estão bem definidos.**

Antes de iniciar a explicação do algoritmo convém ter presente que os sinais de fala analisados para determinação do Jiter e Shimmer correspondem à locução de uma forma sustentada de uma vogal por um período de tempo de alguns segundos. Assim, espera-se que não existam grandes variações de frequência fundamental ao longo de um mesmo sinal e que os períodos glotais tenham uma duração que corresponde ao inverso da frequência fundamental. Contudo, em vozes patológicas, e dependendo da patologia, nem sempre estão presentes os ciclos glotais, e este pressuposto pode não se verificar.

Na secção 3.3.1 apresenta-se a descrição e justificação de diversos detalhes que levaram à solução final apresentada na secção 3.3.2

### 3.3.1. EVOLUÇÃO DO ALGORITMO

Este algoritmo surge através da melhoria de um algoritmo previamente desenvolvido em Oliveira & Lopes, 2012. O trabalho referido consiste na análise de dois sinais de voz (um masculino e um feminino) da vogal \a\ com fala sustentada. Para a determinação dos parâmetros Jitter, Shimmer, os autores utilizaram uma função amplitude média deslizante com janela fixa aplicada ao sinal, à qual se seguiam uma serie de funções, que tinham como objetivo identificar os picos através da comparação do valor da amostra com a amostra anterior. Para o parâmetro HNR os autores utilizaram o método da autocorrelação (Oliveira & Lopes, 2012).

A função amplitude média deslizante (AMD) mencionada acima consiste em fazer deslizar uma janela dentro da qual se calcula a média, ao longo de uma sequência de amostras do sinal de fala (Teixeira, 2013) e é governada pela equação 20.

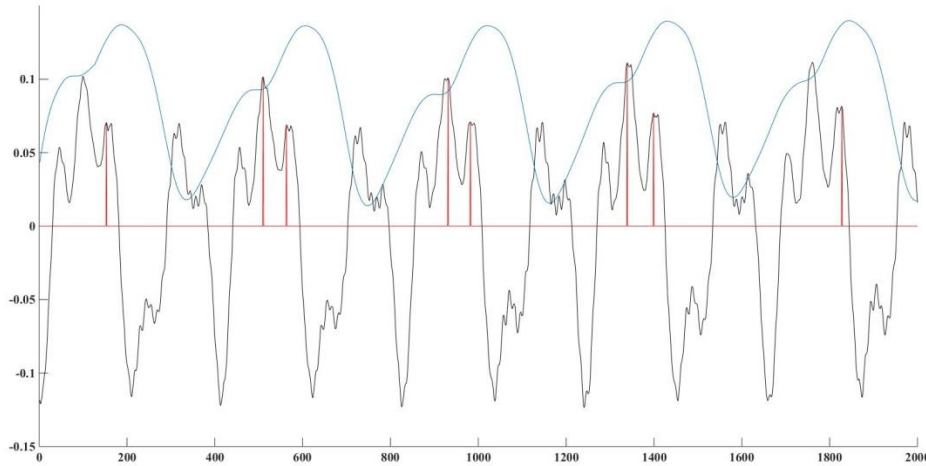
$$M(n) = \frac{1}{N + 1} \sum_{m=n-N/2}^{n+N/2} |x(m)| \quad (20)$$

Quando este algoritmo foi aplicado aos sinais da base de dados SVD, este não reproduziu bons resultados, evidenciando a necessidade de ser melhorado para responder a sinais reais diferentes dos usados no trabalho e a sinais patológico. A diferença mais evidente entre os sinais utilizados pelos autores e os sinais da base de dados SVD foi a frequência de amostragem que passou de 22.05 kHz para 50 KHz. Desta forma foram produzidas alterações no comprimento da janela da função média deslizante para se ajustar ao novo valor da frequência de amostragem mas com uma duração temporal da mesma ordem. Após várias tentativas chegou-se à conclusão que esta forma não era viável, pois devido à diversidade de sinais contidos na base de dados SVD, o algoritmo não era capaz de identificar com sucesso os picos em todos os períodos glotais. Assim sendo, decidiu-se fazer variar o valor da janela em função da frequência fundamental. Para tal desenvolveu-se um pequeno algoritmo, que determina a frequência fundamental através do método da autocorrelação na zona intermédia do sinal. Decidiu-se determinar somente na zona intermédia, pois os sinais a analisar consistiam numa produção de fala sustentada de vogais, não existindo desta forma variação significativa da frequência fundamental ao longo do sinal.

Após a determinação da frequência fundamental, tentou-se chegar a um comprimento de janela para a função média deslizante, em função do comprimento do

### III – Determinação Dos Parâmetros de Fala

período glotal. Para a maioria dos casos foi possível chegar com sucesso a um valor global, mas como pode ser observado na Figura 15, para alguns casos o algoritmo continuava a proceder a uma avaliação errada dos picos.



**Figura 15 – Sinal de voz controlado da vogal /u/ tom baixo, em que a determinação é feita de forma errada.**

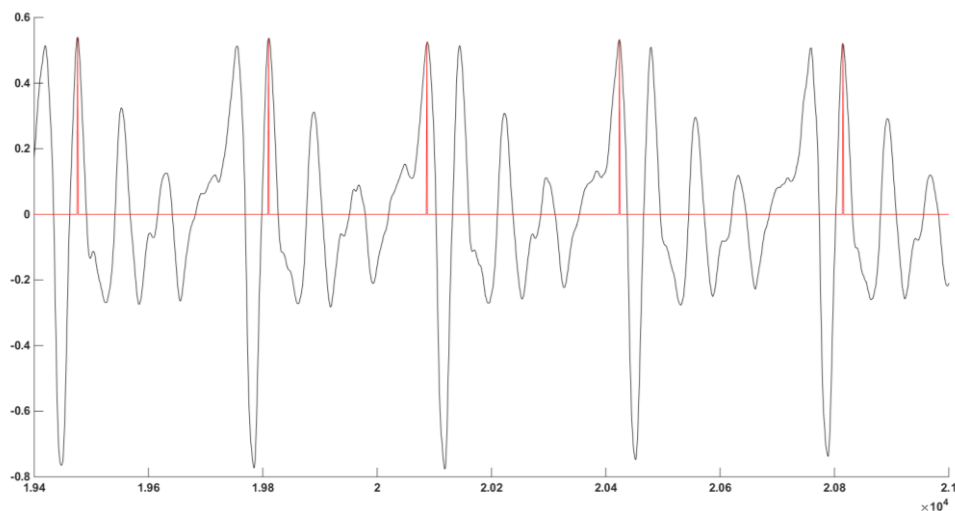
Nesta figura a linha preta representa o sinal de voz, a linha azul o sinal AMD e a linha vermelha os picos identificados. Como é observável, neste exemplo a determinação dos picos máximos ou é feita de forma errada ou procede à identificação de mais do que um pico por período glotal. Tal acontece pois a identificação dos picos era diretamente influenciada pela função da AMD. A condição de identificação era a seguinte: quando a função AMD passa de declive positivo para negativo procede à identificação de um pico nesse local. Esta identificação correspondia à utilização da função máximo na zona circundante a esta quebra de declive (30 amostras).

Como o objetivo era produzir um algoritmo capaz de avaliar corretamente qualquer tipo de voz, efetuaram-se várias alterações ao código de forma a encontrar uma solução funcional para vozes masculinas, femininas, saudáveis e patológicas. Após várias alternativas experimentadas chegou-se à solução final com a utilização da função máximo. Utilizando esta função, a determinação do valor da amplitude e a posição dos picos do sinal começaram a ser obtidas através da implementação de janelas sucessivas com comprimento igual a um período glotal. Após ser identificada a posição do primeiro pico do sinal através da função máximo, o pico seguinte encontra-se avançando-se um período glotal em relação ao primeiro pico, e para uma secção igual a 1/3 do PG antes e 1/3 do PG depois desse ponto é utilizada a função máximo com o intuito de encontrar o segundo pico. Este processo é repetido até ao final do sinal

### III – Determinação Dos Parâmetros de Fala

utilizando sempre como referência o pico anterior. Aquando este processo, os valores da amplitude e posição dos picos são guardados numa matriz de forma a proceder posteriormente ao cálculo dos parâmetros.

Através deste processo foi possível determinar com sucesso os picos para todos os sinais contidos na base de dados onde visualmente é possível identificar a existência de períodos glotais. Apesar da avaliação dos picos ser correta para todas as situações observadas experimentalmente, os valores das medidas dos parâmetros de Jitter e de Shimmer obtidos por este procedimento para alguns casos eram absurdos. Após análise de alguns gráficos correspondentes a estas situações, constou-se que para alguns sinais, a avaliação dos picos negativos seria mais precisa quando o sinal apresentava mais do que um pico de amplitude semelhante por ciclo na parte positiva e picos bem definidos na sua parte negativa como ilustrado na Figura 16.



**Figura 16 - Sinal de voz controlado da vogal /a/ tom alto, ilustrativa de que a determinação seria melhor efetuada através dos picos negativos.**

Desta forma procedeu-se à implementação de um critério de seleção dos sinais, de forma a diferenciar a determinação dos picos através dos picos positivos ou negativos. Este critério consistiu em avaliar o valor obtido pelas funções máximo e mínimo na zona intermédia do sinal, partindo do princípio que o valor mais alto representaria um pico isolado.

Com o uso deste critério, os sinais com picos bem definidos começaram a apresentar valores semelhantes aos obtidos pelo programa Praat porém, para sinais como o ilustrado na Figura 17 os resultados obtidos continuavam a ser desapropriados.

### III – Determinação Dos Parâmetros de Fala

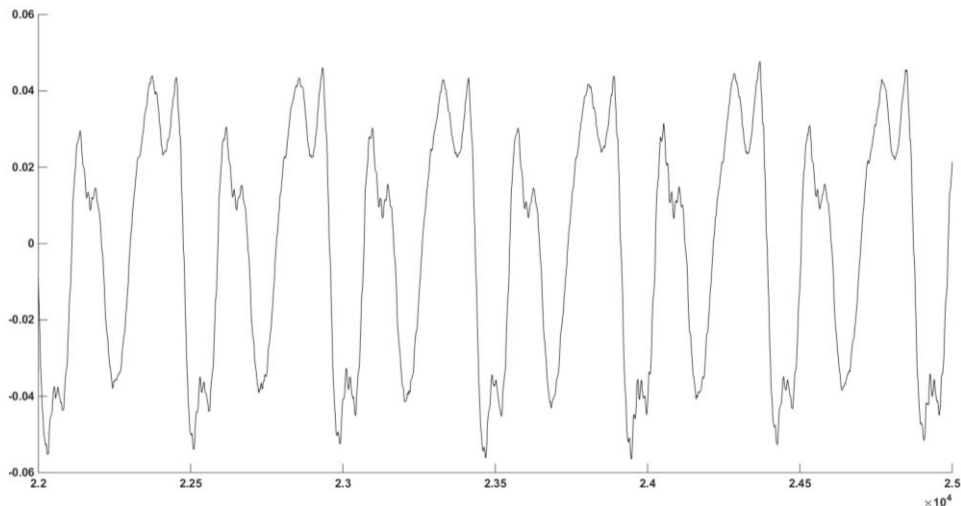


Figura 17 - Sinal de voz controlado da vogal /u/ tom baixo, ilustrativa da dificuldade de determinação dos parâmetros para este tipo de sinais.

Como a determinação direta no próprio sinal dos picos para situações como a apresentada na figura anterior eram apresentados valores errados, decidiu-se que para estes casos a melhor solução seria alisa-los através de uma função média deslizante. Após serem alisados foi aplicado o mesmo tipo de critério que o utilizado para identificar os sinais que apresentavam picos bem definidos, só que aplicado aos sinais alisados. Foi também criado um novo critério de forma a diferenciar os sinais com picos bem definidos dos que necessitavam de ser alisados. Este critério foi produzido utilizando a função do Matlab “*findpeaks*” com argumento “*MinPeakHeight*”, esta função associada ao argumento “*MinPeakHeight*” permite encontrar somente picos que tenham valor superior ao argumento pretendido.

Assim sendo, após algumas experiencias o critério de decisão se os picos são procurados no próprio sinal ou no sinal alisado, ficou da seguinte forma: para uma janela igual a um período glotal após a zona intermédia do sinal, era determinado o valor do máximo e do mínimo, assim como o número de picos superiores a 70% do valor máximo (picos positivos) e inferiores a 70% do valor mínimo (picos negativos). Se o número de picos positivos for 1 e o pico positivo for maior que o pico negativo em módulo ou se o número de picos positivos for igual a 1 e o número de picos negativos for superior a 1, são procurados os **picos positivos** do **próprio sinal** usando janelas sucessivas de comprimento igual a um período glotal até ao final do sinal. Para os casos em que o valor do pico negativo é superior ao valor do pico positivo e o número de picos negativos é igual a 1 ou o número de picos negativos é igual a 1 e o número de

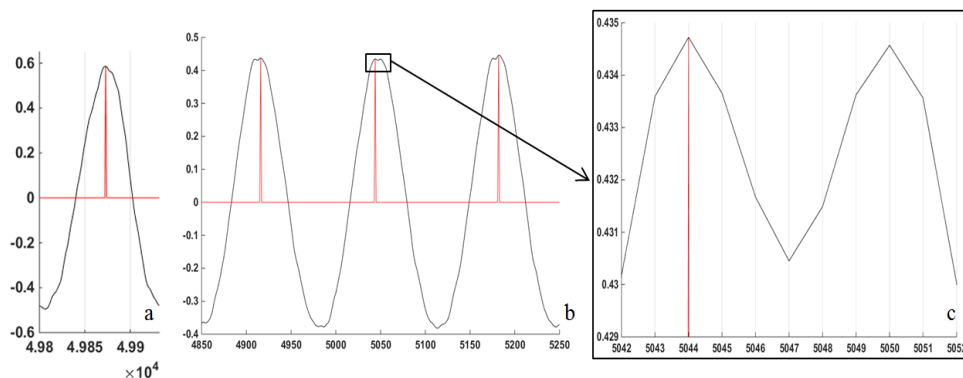
### III – Determinação Dos Parâmetros de Fala

picos positivos maior que 1, são procurados os **picos negativos do próprio sinal** usando janelas sucessivas de comprimento igual a 1 período glotal ao longo do sinal.

Para o caso de o número de picos positivos ou negativos ser superior a 1, o sinal é alisado através da função média deslizante e são utilizadas as mesmas condições descritas anteriormente, mas desta vez para o sinal alisado.

Da análise do algoritmo verificou-se que para uma pequena gama de sinais os valores não estavam corretos. Após análise pormenorizada constatou-se que o critério que determina a quantidade de picos positivos ou negativos do sinal após a zona intermédia era muito limitado.

Como pode ser observado na figura a baixo, na zona intermédia onde é determinado o número de picos (Figura 18a) existe um pico bem definido, mas para outras zonas do mesmo sinal (Figura 18b) já existe mais que um pico no mesmo período glotal como se pode ver no detalhe apresentado no zoom da figura 18c. Nestes sinais o pico do período glotal umas vezes é marcado como sendo o primeiro pico e noutros períodos glotais como sendo o segundo pico, consoante um é maior ou menor que o outro, dado que têm magnitudes muito próximas. No exemplo da figura os picos diferem de 6 amostras que à frequência de amostragem de 50 kHz corresponde a uma diferença de 6/50k ou seja 120  $\mu$ s. Estas diferenças associadas ao sinal todo resultam numa grande diferença significativa nas medidas do parâmetro de Jitter.



**Figura 18 - Sinal de voz controlo da vogal /u/ tom alto, a – pico usado de critério de escolha na zona central do sinal. b – Zona do sinal onde a determinação é efetuada de forma errada. c – Zoom de uma secção de b.**

Para evitar este tipo de erros o critério passou a determinar a quantidade de picos para 10 períodos glotais após a zona intermédia.

Este tipo de erro deixou de aparecer após o critério ter sido alterado para: determina a quantidade de picos para 10 períodos glotais após a zona intermédia. Como a

### III – Determinação Dos Parâmetros de Fala

quantidade de informação sobre o sinal passou a ser maior, esta forma de generalização passou a ser mais eficiente.

O cálculo do HNR foi efetuado utilizando como base as propriedades da função autocorrelação descritas em Boersma 1993. Para implementar o algoritmo de determinação deste parâmetro foi utilizada a função de MatLab “*xcorr*” com comprimento de janela igual a um PG. O HNR foi então determinado de acordo com as expressões apresentadas na secção 3.1.3, usando a autocorrelação normalizada para  $r'_x$ .

Com o intuito de obter valores de HNR próximos dos apresentados pelo Praat testou-se a influencia do comprimento da janela na obtenção dos valores de  $r'_x(\tau_{\max})$ . Utilizou-se uma janela com comprimento igual ao do sinal (HNRt) e de comprimentos iguais a 5 PG (HNR5), 10 PG (HNR10) e 50 PG (HNR50). Experimentou-se também aplicar janelas de hanning a comprimentos de 5 PG (HNRh5), 10 PG (HNRh10) e 50 PG (HNRh50).

Esta análise foi feita utilizando uma amostra de 6 pacientes controlo (3 masculinos e 3 femininos) provenientes da base de dados SVD com idades compreendidas entre os 45 e os 58 anos, tendo em média 51.5 anos e desvio padrão de 4.59.

Nas tabelas 1, 2 e 3 apresentam-se os dados referentes às médias e desvio padrão das vogais /a/, /i/ e /u/, para os tons baixo, normal e alto das medidas de HNR com os comprimentos de janela mencionados acima. Nestas tabelas é notório que os dados da média do algoritmo quando é utilizada uma janela HNRh10 são a maior parte das vezes os mais próximos do valor do Praat. Comparando as médias de todas as vogais e todos os tons para esta janela a que difere mais é a vogal /u/ tom alto em 4.14dB e a que se aproxima mais do valor médio do Praat é a vogal /a/ tom alto com uma diferença de 0.02dB. O desvio padrão é quase sempre inferior quando determinado pelo algoritmo através de HNRh10 com exceção da vogal /a/ tom baixo e vogal /i/ tom baixo.

Os restantes valores apresentam dados relativamente afastados em todas as vogais e tons relativamente aos valores do Praat com a exceção da janela HNR50 para a vogal /u/ em todos os tons. Esta apresenta diferenças médias em relação ao Praat de 0.8dB para o tom baixo, 2.42dB para o tom normal e 2.71dB para o tom alto, sendo todas inferiores às apresentadas pela janela HNRh10.

### III – Determinação Dos Parâmetros de Fala

Tabela 1 - Valores medidos para o parâmetro HNR (em dB) com variação da janela para a vogal /a/.

		Média		Desvio Padrão	
		Algoritmo	Praat	Algoritmo	Praat
<b>/a/ Baixo</b>	<b>HNRt</b>	8,45		3,31	
	<b>HNR5</b>	11,24		1,06	
	<b>HNR10</b>	15,19		3,49	
	<b>HNR20</b>	15,60		5,35	
	<b>HNR50</b>	13,41	20,77	6,09	3,08
	<b>HNRh5</b>	11,91		1,16	
	<b>HNRh10</b>	21,29		3,43	
	<b>HNRh20</b>	23,65		6,09	
	<b>HNRh50</b>	17,69		8,96	
	<b>/a/ Normal</b>	<b>HNRt</b>	9,83		4,41
<b>HNR5</b>		11,57		0,84	
<b>HNR10</b>		17,57		1,57	
<b>HNR20</b>		21,54		2,85	
<b>HNR50</b>		23,35	22,22	3,65	4,41
<b>HNRh5</b>		11,94		1,10	
<b>HNRh10</b>		23,00		2,10	
<b>HNRh20</b>		29,69		4,45	
<b>HNRh50</b>		29,25		6,53	
<b>/a/ Alto</b>		<b>HNRt</b>	13,78		2,27
	<b>HNR5</b>	11,78		0,36	
	<b>HNR10</b>	18,27		0,41	
	<b>HNR20</b>	22,66		1,55	
	<b>HNR50</b>	25,90	23,81	3,73	1,07
	<b>HNRh5</b>	11,96		0,65	
	<b>HNRh10</b>	23,93		0,59	
	<b>HNRh20</b>	31,69		1,87	
	<b>HNRh50</b>	33,14		4,97	

### III – Determinação Dos Parâmetros de Fala

Tabela 2 - Valores medidos para o parâmetro HNR (em dB) com variação da janela para a vogal /i/.

		Média		Desvio Padrão	
		Algoritmo	Praat	Algoritmo	Praat
<b>/i/ Baixo</b>	<b>HNRt</b>	9,24		4,23	
	<b>HNR5</b>	11,08		2,46	
	<b>HNR10</b>	15,58		5,76	
	<b>HNR20</b>	18,89		8,01	
	<b>HNR50</b>	24,70	22,07	5,29	4,46
	<b>HNRh5</b>	11,98		1,84	
	<b>HNRh10</b>	21,85		5,02	
	<b>HNRh20</b>	27,14		9,69	
	<b>HNRh50</b>	32,18		8,41	
<b>/i/ Normal</b>	<b>HNRt</b>	11,90		3,51	
	<b>HNR5</b>	12,03		0,27	
	<b>HNR10</b>	18,57		0,35	
	<b>HNR20</b>	23,73		1,20	
	<b>HNR50</b>	29,22	24,69	2,14	2,62
	<b>HNRh5</b>	12,33		0,44	
	<b>HNRh10</b>	24,32		0,57	
	<b>HNRh20</b>	32,91		1,93	
	<b>HNRh50</b>	38,21		3,38	
<b>/i/ Alto</b>	<b>HNRt</b>	13,67		4,27	
	<b>HNR5</b>	11,96		0,26	
	<b>HNR10</b>	18,54		0,43	
	<b>HNR20</b>	23,80		1,46	
	<b>HNR50</b>	29,37	25,73	2,84	4,76
	<b>HNRh5</b>	12,20		0,47	
	<b>HNRh10</b>	24,16		0,65	
	<b>HNRh20</b>	33,28		2,69	
	<b>HNRh50</b>	39,64		5,88	

### III – Determinação Dos Parâmetros de Fala

Tabela 3 - Valores medidos para o parâmetro HNR (em dB) com variação da janela para a vogal /u/.

		Média		Desvio Padrão	
		Algoritmo	Praat	Algoritmo	Praat
<b>/u/ Baixo</b>	<b>HNRt</b>	10,37		3,78	
	<b>HNR5</b>	11,64		0,29	
	<b>HNR10</b>	18,17		0,59	
	<b>HNR20</b>	23,34		1,20	
	<b>HNR50</b>	26,61	25,81	4,47	3,51
	<b>HNRh5</b>	11,59		0,42	
	<b>HNRh10</b>	23,71		0,69	
	<b>HNRh20</b>	32,99		1,81	
	<b>HNRh50</b>	36,24		6,18	
	<b>/u/ Normal</b>	<b>HNRt</b>	12,22		3,12
<b>HNR5</b>		11,91		0,10	
<b>HNR10</b>		18,75		0,11	
<b>HNR20</b>		24,74		0,44	
<b>HNR50</b>		30,99	28,52	1,35	2,07
<b>HNRh5</b>		11,98		0,22	
<b>HNRh10</b>		24,46		0,23	
<b>HNRh20</b>		35,14		0,58	
<b>HNRh50</b>		43,22		3,02	
<b>/u/ Alto</b>		<b>HNRt</b>	14,42		2,27
	<b>HNR5</b>	11,84		0,22	
	<b>HNR10</b>	18,78		0,22	
	<b>HNR20</b>	24,86		0,41	
	<b>HNR50</b>	31,23	28,52	1,69	2,90
	<b>HNRh5</b>	11,84		0,39	
	<b>HNRh10</b>	24,38		0,36	
	<b>HNRh20</b>	35,25		0,88	
	<b>HNRh50</b>	42,91		3,31	

Na tabela 4 é apresentada uma análise comparativa entre vogais. Neste tipo de comparação os valores de HNR utilizando uma janela HNR10 são novamente os mais próximos do Praat para as vogais /a/ e /i/ com diferenças de 0.48dB e 0.72dB respectivamente. A vogal /u/ apresenta novamente resultados mais próximos do valores do Praat para a janela HNR50 com uma diferença de 2dB enquanto que HNR10 apresenta uma diferença de 3.42dB.

### III – Determinação Dos Parâmetros de Fala

Tabela 4 - Valores medidos para o parâmetro HNR (em dB) com variação da janela resumido conforme a vogal.

		Média		Desvio Padrão	
		Algoritmo	Praat	Algoritmo	Praat
<i>/a/</i>	<b>HNRt</b>	10,69		3,98	
	<b>HNR5</b>	11,53		0,79	
	<b>HNR10</b>	17,01		2,49	
	<b>HNR20</b>	19,93		4,66	
	<b>HNR50</b>	20,88	22,26	7,05	3,24
	<b>HNRh5</b>	11,94		0,94	
	<b>HNRh10</b>	22,74		2,47	
	<b>HNRh20</b>	28,34		5,49	
	<b>HNRh50</b>	26,69		9,44	
	<i>/i/</i>	<b>HNRt</b>	11,60		4,21
<b>HNR5</b>		11,69		1,42	
<b>HNR10</b>		17,56		3,45	
<b>HNR20</b>		22,14		5,05	
<b>HNR50</b>		27,76	24,16	3,95	4,13
<b>HNRh5</b>		12,17		1,07	
<b>HNRh10</b>		23,44		2,99	
<b>HNRh20</b>		31,11		6,26	
<b>HNRh50</b>		36,68		6,52	
<i>/u/</i>		<b>HNRt</b>	12,33		3,39
	<b>HNR5</b>	11,80		0,23	
	<b>HNR10</b>	18,56		0,45	
	<b>HNR20</b>	24,31		1,02	
	<b>HNR50</b>	29,61	27,61	3,47	3,02
	<b>HNRh5</b>	11,80		0,37	
	<b>HNRh10</b>	24,19		0,56	
	<b>HNRh20</b>	34,46		1,56	
	<b>HNRh50</b>	40,79		5,31	

A tabela 5 apresenta a análise comparativa entre tons para o parâmetro HNR. Nesta os valores para HNRh10 são novamente mais próximos dos obtidos pelo Praat, sendo a diferença para o tom baixo de 0.59dB, para o tom normal de 1.21dB e para o tom alto de 1.86dB.

### III – Determinação Dos Parâmetros de Fala

Tabela 5 - Valores medidos para o parâmetro HNR (em dB) com variação da janela resumido conforme o tom.

		Média		Desvio Padrão	
		Algoritmo	Praat	Algoritmo	Praat
<b>Baixo</b>	<b>HNR</b>	9,35		3,65	
	<b>HNR5</b>	11,32		1,48	
	<b>HNR10</b>	16,31		3,91	
	<b>HNR20</b>	19,28		6,19	
	<b>HNR50</b>	21,57	22,88	7,90	4,70
	<b>HNRh5</b>	11,83		1,21	
	<b>HNRh10</b>	22,29		3,48	
	<b>HNRh20</b>	27,93		7,43	
	<b>HNRh50</b>	28,70		11,19	
	<b>Normal</b>	<b>HNRt</b>	11,31		3,66
<b>HNR5</b>		11,84		0,52	
<b>HNR10</b>		18,30		1,03	
<b>HNR20</b>		23,34		2,18	
<b>HNR50</b>		27,85	25,14	4,14	4,01
<b>HNRh5</b>		12,08		0,68	
<b>HNRh10</b>		23,93		1,37	
<b>HNRh20</b>		32,58		3,51	
<b>HNRh50</b>		36,89		7,35	
<b>Alto</b>		<b>HNRt</b>	13,96		2,92
	<b>HNR5</b>	11,86		0,28	
	<b>HNR10</b>	18,53		0,40	
	<b>HNR20</b>	23,77		1,49	
	<b>HNR50</b>	28,83	26,02	3,53	3,60
	<b>HNRh5</b>	12,00		0,51	
	<b>HNRh10</b>	24,16		0,55	
	<b>HNRh20</b>	33,40		2,37	
	<b>HNRh50</b>	38,56		6,17	

Após análise destas tabelas consta-se que para a maioria das situações os resultados mais consistentes com os do Praat são quando se utiliza a janela HNRh10, sendo a diferença mais significativa de 4.14dB. De forma geral pode-se considerar que para a determinação do HNR a janela HNRh10 é aquela que se aproxima mais dos pretendidos, sendo esta aplicada para todas as medições de HNR apresentadas neste trabalho.

#### 3.3.2. ALGORITMO FINAL

A determinação do Jitter e Shimmer inicia com a leitura do ficheiro de som a ser analisado. De seguida determina a frequência fundamental utilizando o método da autocorrelação. Esta função devolve o valor da frequência fundamental sendo o seu inverso o comprimento do período glotal (eq. 15).

O algoritmo inicia com uma análise aos 10 períodos glotais após a zona intermédia do sinal, por ser a zona em que o sinal tem normalmente melhores características de estacionaridade. A identificação de cada período glotal é realizada considerando o período glotal determinado anteriormente. Para o primeiro período vai guardar o valor da amplitude do pico máximo positivo e negativo e de seguida, para cada um dos 10 períodos glotais, determina o valor da amplitude máxima e mínima e compara com os picos adjacentes, enumerando aqueles que forem superiores a 70% do valor do máximo e aqueles que forem inferiores a 70% do mínimo, para os picos negativos.

Em seguida, o algoritmo testa a primeira condição: para o primeiro período glotal a seguir à zona intermédia do sinal, se o valor da amplitude máxima positiva for maior ou igual ao valor da amplitude negativa, em módulo, e para os 10 primeiros períodos glotais se o número de picos superiores a 70% do máximo for igual a 10 (ou seja, só foi encontrado um pico na janela definida para cada período glotal), ou se o número de picos superiores a 70% do máximo for igual a 10 e o número de picos negativos inferiores a 70% do mínimo forem maiores que 10, o algoritmo vai utilizar localizar o máximo para cada período glotal e guardar o seu valor e posição. No caso desta condição se verificar serão considerados os picos de amplitude positivos. A obtenção do valor e da posição dos picos faz-se selecionando uma janela igual a 1 PG no início do sinal onde identifica o máximo (pico). É criado um ponto de referência localizado a 1 PG à frente deste máximo. De seguida pesquisa-se o próximo impulso glotal numa janela de procura. Essa janela está definida a 1/3 PG antes e depois relativamente a esse ponto de referência. Após ser identificado o pico, o respetivo valor e posição é guardado numa matriz. Este processo é repetido sucessivamente utilizando sempre como referência o pico anterior até ao final do sinal.

Se a condição descrita acima falhar, segue-se outra muito semelhante em que são testados os picos negativos em vez dos máximos. A condição é a seguinte: Para o primeiro período glotal após a zona intermédia do sinal, se o valor do mínimo em módulo for maior que o valor do máximo, e para os 10 primeiros períodos glotais o número de picos inferiores a 70% do mínimo for igual a 10, ou se o número de picos

### III – Determinação Dos Parâmetros de Fala

inferiores a 70% do mínimo for igual a 10 e o número de picos superiores a 70% do máximo for maior que 10, o algoritmo vai localizar o mínimo para cada período glotal e guardar o seu valor e posição. Neste caso serão considerados os picos de amplitude negativos. A obtenção dos valores e posições dos picos é feita de forma similar à condição anterior, sendo nesta localizados os mínimos.

Se alguma das condições acima tiver sido verificada, então ficou definido que os picos são ou positivos ou negativos, consoante a condição verificada e os picos são analisados diretamente no sinal de fala. No caso de estas condições falharem é aplicada uma função de média deslizante e a procura dos picos é feita no sinal resultante desta. A média deslizante (AMD) tem um comprimento de  $1/3$  do PG e é governada pela equação 20.

Uma vez obtido o sinal AMD, as ações vão ser executadas neste sinal e não no sinal original. Assim sendo, para o período glotal localizado na zona intermédia do sinal, guarda-se o valor da amplitude dos picos positivo e negativo, e enumeram-se os picos com amplitude superior a 70% dos picos positivos e negativos.

Segue-se assim a última condição do algoritmo: se o valor do pico positivo for maior ou igual ao pico negativo, e o número de picos com valor superior a 70% do pico positivo for igual a 1, ou se o número de picos com valor superior a 70% do pico positivo for igual a 1 e o número de picos com valor inferior a 70% do pico negativo maior que 1, para cada período glotal do sinal, então serão usados os picos positivos do sinal alisado. A obtenção dos dados dos picos é efetuada da mesma forma que na primeira condição mas desta vez no sinal resultado da AMD.

Caso esta condição não seja satisfeita, então serão considerados os picos negativos do sinal AMD. O registo dos valores e posição dos picos é igual à segunda condição, mas utilizando o sinal AMD.

Para a implementação do HNR foi utilizada a equação 18 com uma janela de hanning de comprimento igual a 10 PG (HNRh10).

Na Figura abaixo é apresentado o fluxograma do algoritmo desenvolvido.

### III – Determinação Dos Parâmetros de Fala

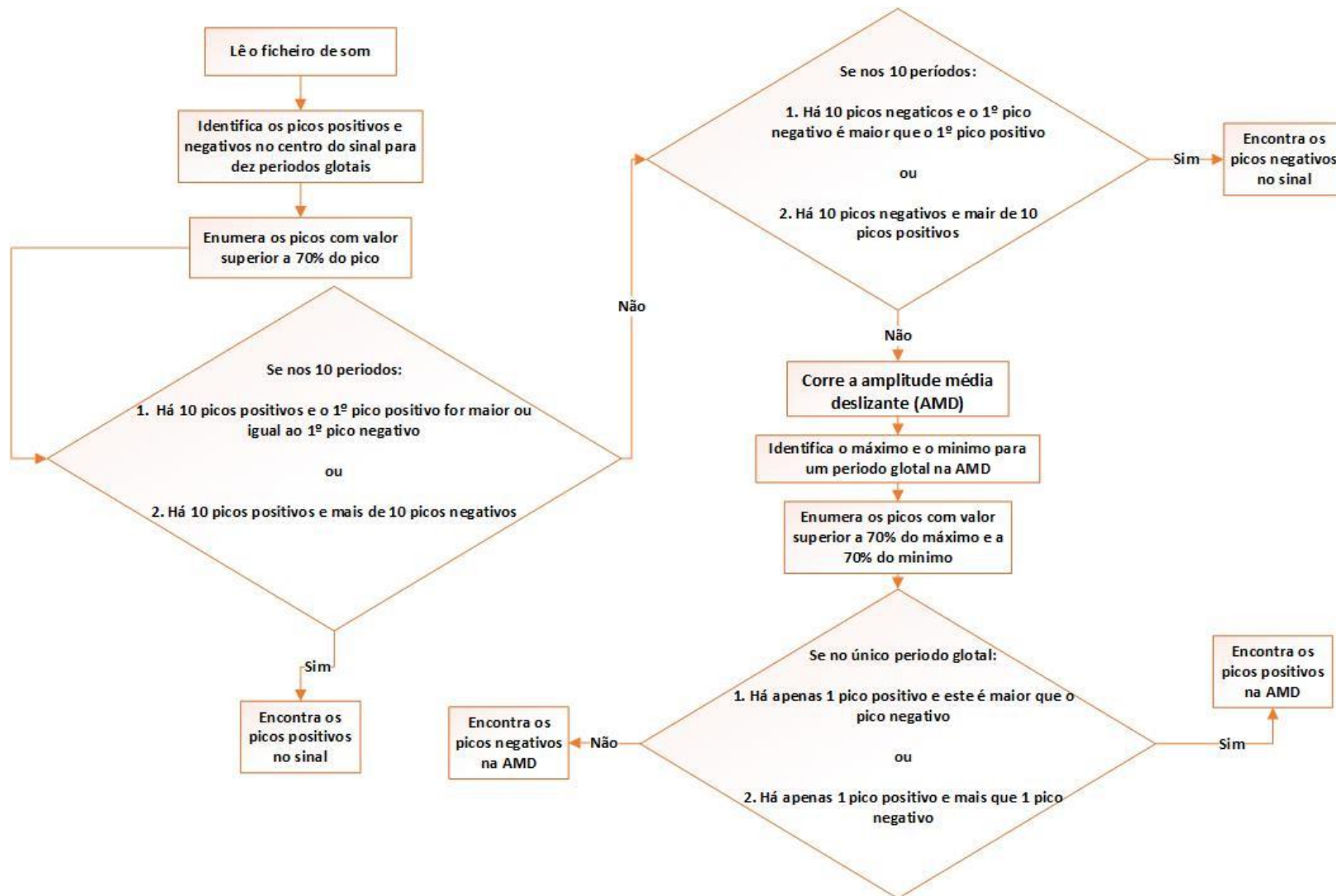


Figura 19 - Fluxograma do algoritmo de identificação dos períodos glotais.

## CAPÍTULO IV

### 4. ANÁLISE DA PRECISÃO DA MEDIDA DO ALGORITMO

#### 4.1. PRECISÃO NA MEDIÇÃO DOS PARÂMETROS EM SINAIS SINTETIZADOS

Nesta secção é feita uma análise aos resultados obtidos pelo algoritmo, comparando com resultados obtidos analiticamente e com o programa Praat (Boersma & Weenink) para sinais sintetizados com o modelo descrito na secção 2.2.

##### 4.1.1. ANÁLISE AO PARÂMETRO JITTER

Inicialmente procedeu-se à medição do Jitter do sinal sintetizado sem qualquer variação do período glotal, com uma frequência fundamental de 100Hz e frequência de amostragem de 22050Hz, a que corresponde um valor nulo para as medidas de Jitter. Na Tabela 6 estão os resultados obtidos pelo algoritmo desenvolvido e os resultados obtidos com o programa Praat. Os resultados obtidos pelos dois métodos para todas as medidas de Jitter são iguais a zero, tal como ocorre efetivamente no sinal.

Tabela 6 - Valores de Jitter medidos para o sinal sintetizado com Jitter nulo

Parâmetro	Praat	Algoritmo
<b>Jitta (µs)</b>	0	0
<b>Jitt (%)</b>	0.00	0.00
<b>rap (%)</b>	0.00	0.00
<b>ppq5 (%)</b>	0.00	0.00

A Tabela 7 apresenta a comparação dos resultados analíticos com os do algoritmo e do programa Praat para o sinal sintetizado com perturbação do tipo 1. Refira-se que a coluna com os valores analíticos apresenta os valores reais de Jitter para este sinal conforme determinado na secção 2. Como pode ser observado tanto o Praat como o algoritmo apresentam resultados muito próximos, mas o algoritmo consegue ser mais preciso. O algoritmo determinou o valor de Jitt, rap e ppq5 com um erro inferior a 0.04% enquanto o Praat determinou estes parâmetros com um erro um pouco inferior a

#### IV – Análise da Precisão da Medida do Algoritmo

0.07%. Para o parâmetro Jitta o Praat teve um erro de  $9\mu\text{s}$  enquanto o algoritmo teve um erro de  $0\mu\text{s}$ .

Considerando que o sinal foi sintetizado com uma frequência de amostragem de 22.050 kHz, o intervalo entre duas amostras corresponde a um período de amostragem de  $45\mu\text{s}$ . Assim, um erro inferior a meia amostra, isto é cerca de  $22.5\mu\text{s}$  não tem significado relevante porque nem a uma amostra de desvio médio nos períodos analisados corresponde.

Tabela 7 - Valores de Jitter medidos para o sinal sintetizado com perturbação de Jitter tipo 1.

<b>Parâmetro</b>	<b>Praat</b>	<b>Algoritmo</b>	<b>Analítico</b>
<b>Jitta (<math>\mu\text{s}</math>)</b>	490	499	499
<b>Jitt (%)</b>	5.02	5.10	5.09
<b>rap (%)</b>	3.35	3.40	3.40
<b>Ppq5 (%)</b>	2.01	2.00	2.04

Na Tabela 8 apresentam-se os resultados medidos no sinal sintetizado com perturbação de Jitter tipo 2. Como pode ser observado tanto o Praat como o algoritmo demonstraram resultados bastante precisos comparativamente aos resultados analíticos, mas neste caso o Praat apresenta resultados mais próximos do valor real que o algoritmo. Para o Jitta o Praat apresenta um erro de  $2\mu\text{s}$  e o algoritmo de  $5\mu\text{s}$ , ambos sem significado. Para os restantes parâmetros o Praat tem um erro inferior a 0.03% enquanto o algoritmo apresenta um erro de aproximadamente 0.05%.

Tabela 8 - Valores de Jitter para o sinal sintetizado com perturbação de Jitter tipo 2

<b>Parâmetro</b>	<b>Praat</b>	<b>Algoritmo</b>	<b>Analítico</b>
<b>Jitta (<math>\mu\text{s}</math>)</b>	247	244	249
<b>Jitt (%)</b>	2.49	2.47	2.52
<b>rap (%)</b>	1.66	1.66	1.68
<b>ppq5 (%)</b>	2.00	1.97	2.02

#### 4.1.2. ANÁLISE AO PARÂMETRO SHIMMER

Para o parâmetro Shimmer procedeu-se igualmente a uma avaliação do algoritmo e do programa Praat, utilizando também o sinal sintetizado com frequência de amostragem de 22050Hz e frequência fundamental de 100Hz, o que significa que os parâmetros de Shimmer teoricamente deverão ser zero. A Tabela 9 apresenta os valores para o algoritmo e para o programa Praat. Pode ser observado que o algoritmo apresenta medições de 0.00 para todos os parâmetros e o Praat apresenta 0.01% para o Shim e 0.00 para os outros parâmetros.

Tabela 9 - Valores de medidos de Shimmer para o sinal sintetizado sem variação da amplitude.

<b>Parâmetro</b>	<b>Praat</b>	<b>Algoritmo</b>
<b>Shim (%)</b>	0.01	0.00
<b>ShdB (dB)</b>	0.00	0.00
<b>Apq3 (%)</b>	0.00	0.00
<b>Apq5 (%)</b>	0.00	0.00

De seguida são apresentados os valores de Shimmer relativos ao sinal sintetizado com perturbação de Shimmer do tipo 1 com uma  $F_0=100\text{Hz}$ . Na Tabela 10 apresentam-se os resultados obtidos analiticamente, assim como os valores medidos pelo algoritmo e pelo Praat. Para este caso os valores medidos pelo algoritmo e pelo Praat são ligeiramente superiores aos determinados analiticamente. Isto pode ser explicado pelo facto de os valores analíticos terem sido determinados com base na amplitude dos impulsos, e como os outros métodos utilizaram o sinal de fala sintetizada para efetuar as medidas, como foi descrito na secção 3.2.3, a duração do período glotal pode alterar a amplitude dos períodos glotais do sinal sintetizado. Para todos os parâmetros as medições efetuadas pelo algoritmo e pelo Praat são muito próximas. A variação entre os dois métodos para Shdb é somente de 0.01dB e para os restantes parâmetros as diferenças são inferiores a 0.1%. Mesmo assim os valores medidos pelo algoritmo e pelo Praat estão muito próximos dos valores determinados analiticamente.

#### IV – Análise da Precisão da Medida do Algoritmo

**Tabela 10 - Valores de medidos de Shimmer para o sinal sintetizado com perturbação de Shimmer do tipo 1.**

<b>Parâmetro</b>	<b>Praat</b>	<b>Algoritmo</b>	<b>Analítico</b>
<b>Shim (%)</b>	22.67	22.57	22.22
<b>ShdB (dB)</b>	1.98	1.97	1.94
<b>Apq3 (%)</b>	15.11	15.05	14.82
<b>Apq5 (%)</b>	9.07	9.03	8.89

Os resultados relativos à determinação dos valores medidos no sinal de fala sintetizado com perturbação de Shimmer do tipo 2 com F0=100 HZ são apresentados na Tabela 11. Neste caso deve ser feita a mesma consideração que anteriormente no que diz respeito aos valores analíticos. Comparando as medições entre o algoritmo e o Praat, pode ser observado que para o parâmetro ShdB a diferença é de 0.00 dB e para os restantes parâmetros a diferença é inferior a 0.09%. Os valores obtidos analiticamente são novamente muito próximos dos valores obtidos pelo algoritmo e pelo Praat.

**Tabela 11 - Valores de medidos de Shimmer para o sinal sintetizado com perturbação de Shimmer do tipo 2**

<b>Parâmetro</b>	<b>Praat</b>	<b>Algoritmo</b>	<b>Analítico</b>
<b>Shim (%)</b>	10.92	10.99	10.53
<b>ShdB (dB)</b>	1.01	1.01	0.97
<b>Apq3 (%)</b>	7.22	7.16	7.02
<b>Apq5 (%)</b>	8.65	8.74	8.42

**4.1.3. ANÁLISE DAS MEDIDAS COM VARIAÇÃO DA FREQUÊNCIA FUNDAMENTAL**

Esta sequência de experiências consiste em fazer medições aos parâmetros de Shimmer usando o sinal de voz sintetizado com diferentes valores de  $F_0$ . Os resultados expostos correspondem somente aos valores obtidos pelo algoritmo desenvolvido. Para cada experiência são apresentadas situações diferentes considerando a amplitude dos impulsos glotais.

Como era de esperar, não se verificou qualquer tipo de alteração nos valores de Jitter perante as variações de valores de  $F_0$  pois, como é observável através da equação 19  $F_0$  e  $T_0$  são diretamente proporcionais.

A Tabela 12 apresenta os valores da medição dos parâmetros de Shimmer usando o sinal sintetizado sem qualquer variação da amplitude, isto significa que Shimmer deverá ser zero. Os valores de  $F_0$  utilizados foram de 75, 100 e 190 Hz. Os sinais de fala sintetizados com o valor de  $F_0=100\text{Hz}$  são os mesmos utilizados nas experiências anteriores. Analisando os resultados da Tabela 12, consta-se que para os três valores de  $F_0$  os resultados obtidos pelo algoritmo são iguais a 0.00 para todos os parâmetros exceto para  $F_0=190\text{Hz}$  em que  $\text{Shim}=0.01\%$ .

**Tabela 12 - Valores de Shimmer para o sinal sintetizado com diferentes  $F_0$  e sem variação da amplitude.**

<b>F0 (Hz)</b>	<b>75</b>	<b>100</b>	<b>190</b>
<b>Shim (%)</b>	0.00	0.00	0.01
<b>ShdB (dB)</b>	0.00	0.00	0.00
<b>Apq3 (%)</b>	0.00	0.00	0.00
<b>Apq5 (%)</b>	0.00	0.00	0.00

A Tabela 13 apresenta os valores obtidos pelo algoritmo utilizando o sinal de voz sintetizado com perturbação de Shimmer do tipo 1 variando a  $F_0$ . Para os três valores de  $F_0$ , os valores de shimmer são os apresentados na coluna de valores analíticos da tabela 5 e que se repete na última coluna da tabela 8.

As quatro medidas de Shimmer apresentam valores significativamente superiores para  $F_0=190\text{ Hz}$  e ligeiramente superiores para  $F_0=75\text{ Hz}$  comparativamente aos valores de referência analíticos. Para valores obtidos com  $F_0=75\text{ Hz}$  pode-se considerar que os resultados são aproximados pois a diferença é inferior a uma unidade para as medidas em percentagem. Quanto aos valores obtidos com  $F_0=190\text{ Hz}$ , estes podem ser justificados tendo em conta a consideração feita em 3.2.3.

#### IV – Análise da Precisão da Medida do Algoritmo

**Tabela 13 - Valores de Shimmer para o sinal sintetizado com diferentes valores de  $F_0$  e com perturbação de Shimmer do tipo 1.**

<b>F0 (Hz)</b>	<b>75</b>	<b>100</b>	<b>190</b>	<b>Analítico</b>
<b>Shim (%)</b>	23.54	22.57	30.67	22,22
<b>ShdB (dB)</b>	2.05	1.97	2.69	1,94
<b>Apq3 (%)</b>	15.69	15.05	20.45	14,82
<b>Apq5 (%)</b>	9.41	9.03	12.27	8,89

A Tabela 14 apresenta os valores obtidos pelo algoritmo utilizando o sinal de voz sintetizado com perturbação de Shimmer do tipo 2 para diferentes valores de  $F_0$ . Também nesta situação de variação da  $F_0$ , para os 75 Hz há um desvio inferior a uma unidade para as medidas apresentadas em porcentagem, e um desvio maior para a  $F_0$  de 190 Hz, justificada pela mesma razão apresentada anteriormente.

**Tabela 14 - Valores de Shimmer para o sinal sintetizado com diferentes valores de  $F_0$  e com perturbação de Shimmer do tipo 2.**

<b>F0 (Hz)</b>	<b>75</b>	<b>100</b>	<b>190</b>	<b>Analítico</b>
<b>Shim (%)</b>	11.09	10.99	14.62	10,53
<b>ShdB (dB)</b>	1.02	1.01	1.33	0,97
<b>Apq3 (%)</b>	7.44	7.16	9.74	7,02
<b>Apq5 (%)</b>	8.91	8.74	9.40	8,42

**4.1.4. RESUMO DA PRECISÃO DAS MEDIDAS COM SINAIS SINTETIZADOS**

Após análise do erro das medidas do Jitter e Shimmer com um sinal sintetizado, é possível constatar que para o parâmetro Jitter as medições foram muito precisas para todas as situações propostas. A tabela 15 apresenta um resumo dos erros para as medidas de Jitter e Shimmer com o algoritmo e com o Praat com o sinal sintetizado para as situações experimentadas (sem variação e variações tipos I e II).

**Tabela 15 – Resumo dos erros para as medidas de Jitter e Shimmer com o algoritmo e com o Praat com o sinal sintetizado.**

		<b>Sem variação</b>	<b>Tipo 1</b>	<b>Tipo 2</b>
<b>Jitter ALG.</b>	<b>Jitta(<math>\mu</math>s)</b>	0	0	5
	<b>Jitt,rap,ppq5(%)</b>	0	0,04	0,05
<b>Jitter PRAAT</b>	<b>Jitta (<math>\mu</math>s)</b>	0	9	2
	<b>Jitt,rap,qqp5(%)</b>	0	0,07	0,03
<b>Shimmer ALG.</b>	<b>ShdB (dB)</b>	0	0,03	0,04
	<b>Shim,APQ3,APQ5(%)</b>	0	0,35	0,46
<b>Shimmer PRAAT</b>	<b>SHdB (dB)</b>	0	0,04	0,04
	<b>Shim,APQ3,APQ5(%)</b>	0,01	0,45	0,39

O algoritmo produziu um erro inferior a 5  $\mu$ s para o parâmetro Jitta e inferior a 0.05% para os parâmetros Jitt, RAP e PPQ5. O programa Praat produziu um erro inferior a 9  $\mu$ s para Jitta e inferior a 0.07% para os parâmetros Jitt, rap e ppq5. Em comparação com o Praat o algoritmo mostrou um erro ligeiramente mais baixo, sendo que o erro quer do algoritmo quer do Praat são bastante inferiores ao erro de uma amostra para a frequência de amostragem usada, já que com uma  $F_a$  de 22.050 kHz o intervalo entre amostras é de 45  $\mu$ s.

No que diz respeito ao Shimmer, a última experiência mostrou que o sinal de voz sintetizado pode ter uma perturbação de Shimmer maior do que o produzido na amplitude do trem de pulsos glotais. Consequentemente os valores determinados analiticamente não podem ser considerados como valores precisos. É notório que para os parâmetros de Shimmer medidos pelo algoritmo e pelo Praat, existiu uma consistência de valores entre estes dois métodos. Estes apresentaram diferenças de perturbação inferiores a 0.01 dB para o parâmetro ShdB e inferiores a 0.1% para os parâmetros Shim, Apq3 e Apq5. As diferenças para os valores determinados analiticamente foram de 0.04 dB no ShdB quer para o algoritmo quer para o Praat e

#### IV – Análise da Precisão da Medida do Algoritmo

inferiores a 0.46% para as medidas de Shim, Apq3 e Apq5 pelo algoritmo e inferiores a 0.45% pelo Praat.

Em conclusão, todas as medidas de Jitter e de Shimmer efetuadas pelo algoritmo no sinal sintetizado podem ser consideradas de elevada precisão.

Estes tipos de perturbação foram publicados em Teixeira & Gonçalves (2014) utilizando um algoritmo com algumas melhorias em relação ao algoritmo apresentado por Oliveira & Lopes (2012). Os dados apresentados acima são provenientes do algoritmo descrito acima. A versão do algoritmo aqui apresentada já é uma versão muito desenvolvida e melhorada em relação ao algoritmo referido em função das dificuldades apresentadas por sinais reais quer de vozes controlo quer de vozes patológicas. No entanto durante todo o desenvolvimento do algoritmo garantiu-se que para os sinais de fala sintetizada os resultados não sofriam perda de precisão. Assim, o novo algoritmo aqui apresentado efetua as medidas exatamente com a mesma precisão das apresentadas nas secções anteriores. Assim se evidencia que não é suficiente fazer uma avaliação de um algoritmo somente através da análise de sinais sintetizados.

O HNR não foi testado porque não foi encontrada a forma de introduzir uma componente não harmónica no sinal sintetizado

## **4.2. COMPARAÇÃO DAS MEDIDAS DO ALGORITMO APLICADO A SINAIS DE VOZ**

Nesta secção é feita uma comparação dos resultados obtidos através do algoritmo desenvolvido com o programa Praat. O objetivo é comprovar que o algoritmo desenvolvido é capaz de determinar os vários parâmetros com uma precisão parecida com o programa Praat, atestando desta forma os valores obtidos e por consequência o algoritmo. É feita uma comparação dos valores obtidos pelos dois métodos usando sinais de controlo e sinais patológicos. Nesta análise os parâmetros mencionados acima são divididos em Jitter, Shimmer e HNR, existindo também uma distinção entre géneros. As vozes do grupo de controlo foram selecionadas para que a média de idades no grupo de controlo e grupos de patologias fosse muito semelhante.

Os valores exatos de Jitter e de Shimmer em sinais de voz naturais, tanto nas de controlo e mais ainda nas patológicas são desconhecidos e muito complicados de medir com exatidão, ainda que de forma manual. Assim, optou-se por uma análise comparativa das medidas efetuadas pelo algoritmo desenvolvido com as medidas de um software largamente difundido e credível, apesar de nunca haver a certeza de qual das medidas é a mais exata entre as do algoritmo e as do programa Praat. Para o caso de vozes patológicas a determinação das medidas pode ser uma tarefa muito difícil devido à anormalidade de alguns sinais. Nesses casos um valor muito elevado é suficiente para indicar que a voz é patológica não sendo relevante o valor exato de Jitter ou de Shimmer.

Como referido anteriormente, as vozes usadas nesta secção são retiradas da base de dados SVD que contém reproduções das vogais /a/, /i/ e /u/ nos tons baixo, normal e alto. Como a frequência de amostragem dos sinais da base de dados é de 50 kHz, o intervalo entre amostras é apenas de 20  $\mu$ s, pelo que um valor de Jitter inferior a 10  $\mu$ s é inferior a meio intervalo entre amostras.

#### **4.2.1. COMPARAÇÃO ALGORITMO/PRAAT PARA SINAIS DE VOZES DE CONTROLO**

A análise aos sinais de controlo foi feita utilizando uma amostra de 15 pessoas do género feminino e 15 do género masculino. O grupo feminino apresenta idades compreendidas entre os 39 e os 65 anos, tendo em média 50.53 anos e desvio padrão de 8.86. O grupo masculino apresenta idades compreendidas entre os 40 e os 65 anos, tendo em média 50.07 anos e desvio padrão de 7.99.

##### **4.2.1.1. JITTER**

Na Tabela 16 apresentam-se os dados relativos à média e desvio padrão das medidas de Jitter para as vogais /a/, /i/ e /u/, para os tons baixo, normal e alto.

Nesta tabela, é notório que os dados da média do algoritmo relativos a todos os parâmetros de Jitter para todos os tons e todas as vogais são ligeiramente superiores aos dados obtidos pelo Praat. Contudo, os valores de desvio padrão apresentados são muito semelhantes nos dois métodos exceto para a vogal /a/ tom alto, vogal /i/ tom normal e alto e vogal /u/ tom alto. No que diz respeito à comparação de géneros, os valores do algoritmo de Jitter para o sexo masculino são quase sempre superiores aos do sexo feminino exceto para as vogais /i/ e /u/ nos tons alto e normal, respetivamente. Os valores do Praat relativos ao parâmetro Jitter para o sexo masculino são sempre superiores aos valores apresentados pelo parâmetro Jitter para o sexo feminino. Entende-se como normal um valor absoluto de Jitter mais alto para as vozes masculinas por nestes casos a frequência fundamental é também em média mais baixo, logo têm em média um período fundamental mais longo, pelo que se compreende que as variações também tenham um valor maior, em valor absoluto. Já nos valores relativos de Jitter esta diferença não é justificada porque as variações são medidas relativamente ao período médio.

Para os restantes parâmetros os valores referentes ao sexo feminino são quase sempre superiores aos apresentados pelo sexo masculino para os dois métodos, excetuando-se os valores obtidos pelo Praat para a vogal /i/ no tom alto para os parâmetros Jitt e ppq5 e para a vogal /u/ tom alto para o parâmetro Jiit em que os valores de média dão iguais para os dois géneros. Ainda em relação ao Jitter, para a vogal /u/ tom baixo o parâmetro ppq5, contrariamente ao sucedido nos restantes valores, apresenta valor superior para vozes femininas.

IV – Análise da Precisão da Medida do Algoritmo

A diferença de valores apresentada pelo algoritmo e pelo Praat não é significativa pois, em nenhum dos casos a diferença entre os dois métodos chega a ser superior a uma amostra, sendo a maior diferença de apenas 11  $\mu$ s.

Tabela 16 - Média e desvio padrão dos parâmetros de Jitter ordenados por vogal e tom para os sinais de controlo.

Vogal/Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
<b>/a/ Baixo</b>	<b>Jitta (<math>\mu</math>s)</b>	27 $\pm$ 12	21 $\pm$ 11	37 $\pm$ 10	34 $\pm$ 10
	<b>Jitt (%)</b>	0,47 $\pm$ 0,20	0,37 $\pm$ 0,19	0,44 $\pm$ 0,12	0,39 $\pm$ 0,10
	<b>rap (%)</b>	0,27 $\pm$ 0,12	0,19 $\pm$ 0,12	0,22 $\pm$ 0,06	0,18 $\pm$ 0,04
	<b>ppq5 (%)</b>	0,27 $\pm$ 0,11	0,20 $\pm$ 0,11	0,25 $\pm$ 0,07	0,22 $\pm$ 0,06
<b>/a/ Normal</b>	<b>Jitta (<math>\mu</math>s)</b>	26 $\pm$ 11	18 $\pm$ 11	31 $\pm$ 10	24 $\pm$ 9
	<b>Jitt (%)</b>	0,52 $\pm$ 0,24	0,36 $\pm$ 0,23	0,42 $\pm$ 0,14	0,33 $\pm$ 0,12
	<b>rap (%)</b>	0,32 $\pm$ 0,15	0,21 $\pm$ 0,15	0,24 $\pm$ 0,09	0,17 $\pm$ 0,08
	<b>ppq5 (%)</b>	0,30 $\pm$ 0,11	0,20 $\pm$ 0,11	0,24 $\pm$ 0,07	0,19 $\pm$ 0,08
<b>/a/ Alto</b>	<b>Jitta (<math>\mu</math>s)</b>	19 $\pm$ 5	11 $\pm$ 4	19 $\pm$ 8	13 $\pm$ 7
	<b>Jitt (%)</b>	0,47 $\pm$ 0,12	0,27 $\pm$ 0,10	0,37 $\pm$ 0,20	0,23 $\pm$ 0,11
	<b>rap (%)</b>	0,29 $\pm$ 0,07	0,15 $\pm$ 0,06	0,22 $\pm$ 0,13	0,11 $\pm$ 0,06
	<b>ppq5 (%)</b>	0,29 $\pm$ 0,07	0,15 $\pm$ 0,06	0,22 $\pm$ 0,12	0,12 $\pm$ 0,05
<b>/i/ Baixo</b>	<b>Jitta (<math>\mu</math>s)</b>	24 $\pm$ 9	19 $\pm$ 11	29 $\pm$ 8	26 $\pm$ 6
	<b>Jitt (%)</b>	0,45 $\pm$ 0,17	0,35 $\pm$ 0,19	0,35 $\pm$ 0,11	0,31 $\pm$ 0,07
	<b>rap (%)</b>	0,27 $\pm$ 0,11	0,19 $\pm$ 0,12	0,16 $\pm$ 0,06	0,12 $\pm$ 0,04
	<b>ppq5 (%)</b>	0,25 $\pm$ 0,07	0,18 $\pm$ 0,09	0,20 $\pm$ 0,06	0,17 $\pm$ 0,04
<b>/i/ Normal</b>	<b>Jitta (<math>\mu</math>s)</b>	22 $\pm$ 11	13 $\pm$ 8	22 $\pm$ 5	19 $\pm$ 7
	<b>Jitt (%)</b>	0,48 $\pm$ 0,21	0,28 $\pm$ 0,15	0,32 $\pm$ 0,06	0,26 $\pm$ 0,08
	<b>rap (%)</b>	0,30 $\pm$ 0,13	0,16 $\pm$ 0,09	0,17 $\pm$ 0,04	0,12 $\pm$ 0,06
	<b>ppq5 (%)</b>	0,27 $\pm$ 0,10	0,16 $\pm$ 0,09	0,18 $\pm$ 0,03	0,15 $\pm$ 0,05
<b>/i/ Alto</b>	<b>Jitta (<math>\mu</math>s)</b>	18 $\pm$ 9	9 $\pm$ 5	16 $\pm$ 5	12 $\pm$ 6
	<b>Jitt (%)</b>	0,47 $\pm$ 0,21	0,22 $\pm$ 0,12	0,31 $\pm$ 0,07	0,22 $\pm$ 0,10
	<b>rap (%)</b>	0,30 $\pm$ 0,13	0,12 $\pm$ 0,07	0,19 $\pm$ 0,04	0,11 $\pm$ 0,07
	<b>ppq5 (%)</b>	0,28 $\pm$ 0,10	0,12 $\pm$ 0,07	0,19 $\pm$ 0,04	0,12 $\pm$ 0,06
<b>/u/ Baixo</b>	<b>Jitta (<math>\mu</math>s)</b>	24 $\pm$ 6	16 $\pm$ 6	30 $\pm$ 11	26 $\pm$ 12
	<b>Jitt (%)</b>	0,45 $\pm$ 0,09	0,30 $\pm$ 0,09	0,36 $\pm$ 0,12	0,31 $\pm$ 0,14
	<b>rap (%)</b>	0,26 $\pm$ 0,07	0,16 $\pm$ 0,06	0,18 $\pm$ 0,06	0,12 $\pm$ 0,07
	<b>ppq5 (%)</b>	0,26 $\pm$ 0,05	0,16 $\pm$ 0,05	0,20 $\pm$ 0,06	0,17 $\pm$ 0,08
<b>/u/ Normal</b>	<b>Jitta (<math>\mu</math>s)</b>	26 $\pm$ 10	16 $\pm$ 10	24 $\pm$ 4	19 $\pm$ 6
	<b>Jitt (%)</b>	0,55 $\pm$ 0,19	0,34 $\pm$ 0,19	0,35 $\pm$ 0,08	0,26 $\pm$ 0,08
	<b>rap (%)</b>	0,34 $\pm$ 0,12	0,20 $\pm$ 0,12	0,18 $\pm$ 0,05	0,11 $\pm$ 0,04
	<b>ppq5 (%)</b>	0,32 $\pm$ 0,10	0,19 $\pm$ 0,10	0,20 $\pm$ 0,04	0,14 $\pm$ 0,03
<b>/u/ Alto</b>	<b>Jitta (<math>\mu</math>s)</b>	20 $\pm$ 7	9 $\pm$ 4	22 $\pm$ 8	12 $\pm$ 6
	<b>Jitt (%)</b>	0,51 $\pm$ 0,20	0,23 $\pm$ 0,11	0,44 $\pm$ 0,18	0,23 $\pm$ 0,10
	<b>rap (%)</b>	0,32 $\pm$ 0,12	0,13 $\pm$ 0,07	0,26 $\pm$ 0,11	0,11 $\pm$ 0,07
	<b>ppq5 (%)</b>	0,30 $\pm$ 0,11	0,13 $\pm$ 0,05	0,25 $\pm$ 0,09	0,11 $\pm$ 0,05

#### IV – Análise da Precisão da Medida do Algoritmo

Como está demonstrado na Tabela 17, procedeu-se a uma análise comparativa entre vogais, fazendo a média para os três tons. Os dados do algoritmo são novamente superiores ao do Praat e o desvio padrão continua a ser muito aproximado ou mesmo igual para todos os parâmetros. Na comparação entre géneros, os dados do algoritmo referentes ao sexo masculino são superiores aos apresentados pelo sexo feminino para o parâmetro Jitta, exceto na vogal /i/, em que os valores são iguais. Quanto aos valores obtidos pelo Praat, estes são novamente superiores no sexo masculino para todas as vogais. Nos restantes parâmetros os dados apresentados pelo sexo feminino são sempre superiores aos apresentados pelo sexo masculino para os dois métodos (algoritmo e Praat). Neste caso a maior diferença de médias entre o algoritmo e o Praat é de 9  $\mu$ s.

Tabela 17 - Média e desvio padrão dos parâmetros de Jitter ordenados por vogal para os sinais de controlo.

Vogal	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
/a/	Jitta ( $\mu$ s)	24 $\pm$ 10	17 $\pm$ 10	29 $\pm$ 12	24 $\pm$ 12
	Jitt (%)	0,49 $\pm$ 0,19	0,33 $\pm$ 0,18	0,41 $\pm$ 0,16	0,32 $\pm$ 0,12
	rap (%)	0,30 $\pm$ 0,12	0,18 $\pm$ 0,12	0,22 $\pm$ 0,09	0,15 $\pm$ 0,07
	ppq5 (%)	0,29 $\pm$ 0,10	0,19 $\pm$ 0,10	0,24 $\pm$ 0,09	0,18 $\pm$ 0,08
/i/	Jitta ( $\mu$ s)	22 $\pm$ 10	14 $\pm$ 9	22 $\pm$ 8	19 $\pm$ 9
	Jitt (%)	0,47 $\pm$ 0,19	0,28 $\pm$ 0,16	0,33 $\pm$ 0,08	0,26 $\pm$ 0,09
	rap (%)	0,29 $\pm$ 0,13	0,15 $\pm$ 0,10	0,18 $\pm$ 0,05	0,12 $\pm$ 0,05
	ppq5 (%)	0,27 $\pm$ 0,09	0,15 $\pm$ 0,09	0,19 $\pm$ 0,04	0,14 $\pm$ 0,05
/u/	Jitta ( $\mu$ s)	23 $\pm$ 8	14 $\pm$ 8	25 $\pm$ 9	19 $\pm$ 10
	Jitt (%)	0,50 $\pm$ 0,17	0,29 $\pm$ 0,14	0,38 $\pm$ 0,14	0,27 $\pm$ 0,11
	rap (%)	0,31 $\pm$ 0,11	0,16 $\pm$ 0,09	0,21 $\pm$ 0,09	0,11 $\pm$ 0,06
	ppq5 (%)	0,29 $\pm$ 0,09	0,16 $\pm$ 0,07	0,22 $\pm$ 0,07	0,14 $\pm$ 0,06

Na Tabela 18 apresentam-se as medidas do parâmetro Jitter efetuadas pelo algoritmo e pelo Praat comparando os tons, fazendo agora a média entre vogais. Nesta, constata-se que os valores do algoritmo são novamente superiores aos valores do Praat para todos os parâmetros apesar de não existir nenhuma diferença considerada significativa. Os valores apresentados pelo algoritmo para o parâmetro Jitta no género masculino são superiores relativamente ao género feminino para os tons baixo e normal, e igual para o tom alto. Neste mesmo parâmetro o Praat apresentou valores superiores no género masculino para todos os tons. Para os restantes parâmetros o sexo feminino apresentou valores mais elevados nos dois métodos, exceto no tom baixo com os parâmetros Jitt,

#### IV – Análise da Precisão da Medida do Algoritmo

rap e ppq5 para os quais os valores obtidos pelo Praat apresentaram dois valores mais elevados e um valor igual para o sexo masculino. Também neste caso a maior diferença de médias é de apenas 9  $\mu$ s.

Tabela 18 - Média e desvio padrão dos parâmetros de Jitter ordenados por tom para os sinais de voz controle.

Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
Baixo	Jitta ( $\mu$ s)	25 $\pm$ 9	19 $\pm$ 10	32 $\pm$ 11	28 $\pm$ 10
	Jitt (%)	0,46 $\pm$ 0,16	0,34 $\pm$ 0,16	0,38 $\pm$ 0,12	0,34 $\pm$ 0,11
	rap (%)	0,27 $\pm$ 0,10	0,18 $\pm$ 0,10	0,19 $\pm$ 0,06	0,14 $\pm$ 0,06
	ppq5 (%)	0,26 $\pm$ 0,08	0,18 $\pm$ 0,09	0,22 $\pm$ 0,07	0,19 $\pm$ 0,07
Normal	Jitta ( $\mu$ s)	25 $\pm$ 11	16 $\pm$ 10	26 $\pm$ 8	21 $\pm$ 8
	Jitt (%)	0,52 $\pm$ 0,21	0,32 $\pm$ 0,19	0,36 $\pm$ 0,10	0,28 $\pm$ 0,10
	rap (%)	0,32 $\pm$ 0,13	0,19 $\pm$ 0,12	0,20 $\pm$ 0,07	0,13 $\pm$ 0,06
	ppq5 (%)	0,30 $\pm$ 0,10	0,18 $\pm$ 0,10	0,21 $\pm$ 0,06	0,16 $\pm$ 0,06
Alto	Jitta ( $\mu$ s)	19 $\pm$ 7	10 $\pm$ 5	19 $\pm$ 7	12 $\pm$ 6
	Jitt (%)	0,49 $\pm$ 0,18	0,24 $\pm$ 0,11	0,37 $\pm$ 0,16	0,23 $\pm$ 0,10
	rap (%)	0,31 $\pm$ 0,11	0,14 $\pm$ 0,07	0,22 $\pm$ 0,10	0,11 $\pm$ 0,06
	ppq5 (%)	0,29 $\pm$ 0,09	0,13 $\pm$ 0,06	0,22 $\pm$ 0,09	0,12 $\pm$ 0,05

#### 4.2.1.2. SHIMMER

Para o parâmetro Shimmer procedeu-se à mesma sequência de análise usada para o Jitter. Na Tabela 19 apresentam-se os dados relativos à média e desvio padrão das medidas de Shimmer para as vogais /a/, /i/ e /u/, para os tons baixo, normal e alto.

Apesar da consistência de valores entre os dois métodos, para a vogal /a/ os valores obtidos pelo Praat para todos os parâmetros são sempre superiores aos obtidos pelo algoritmo. A vogal /u/ tom alto apresenta valores de médias equivalentes para os dois métodos em quase todos os parâmetros. No que diz respeito às restantes vogais e tons os valores apresentados pelo algoritmo e pelo Praat são muito semelhantes não havendo nenhuma tendência. A maior diferença entre as medidas realizadas pelo algoritmo e pelo Praat para o parâmetro ShdB é de apenas 0.08 dB.

IV – Análise da Precisão da Medida do Algoritmo

Tabela 19 - Média e desvio padrão dos parâmetros de Shimmer ordenados por vogal e tom para os sinais de controlo.

Vogal/Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
/a/ Baixo	Shim (%)	2,14±0,92	2,80±1,60	2,96±1,37	3,44±1,24
	Shdb (dB)	0,19±0,08	0,25±0,14	0,26±0,12	0,30±0,11
	apq3 (%)	0,97±0,56	1,42±0,93	1,41±0,73	1,71±0,64
	apq5 (%)	1,31±0,66	1,70±1,08	1,85±0,77	2,08±0,73
/a/ Normal	Shim (%)	2,02±0,53	2,58±1,45	2,29±0,75	3,16±1,86
	Shdb (dB)	0,18±0,05	0,23±0,13	0,20±0,07	0,28±0,17
	apq3 (%)	0,99±0,28	1,35±0,81	1,12±0,47	1,61±0,93
	apq5 (%)	1,18±0,33	1,57±1,00	1,46±0,50	2,00±1,25
/a/ Alto	Shim (%)	2,03±0,62	2,11±0,92	1,68±0,88	1,92±0,85
	Shdb (dB)	0,18±0,05	0,19±0,08	0,15±0,10	0,17±0,09
	apq3 (%)	1,06±0,38	1,12±0,58	0,81±0,60	0,95±0,51
	apq5 (%)	1,19±0,38	1,23±0,55	0,95±0,40	1,07±0,48
/i/ Baixo	Shim (%)	1,92±0,67	1,86±0,73	2,20±0,87	2,21±0,94
	Shdb (dB)	0,17±0,06	0,16±0,07	0,19±0,08	0,19±0,08
	apq3 (%)	0,88±0,36	0,87±0,44	0,88±0,37	0,92±0,44
	apq5 (%)	1,11±0,44	1,05±0,38	1,48±0,67	1,43±0,61
/i/ Normal	Shim (%)	1,88±0,83	1,90±0,85	1,64±0,56	1,59±0,53
	Shdb (dB)	0,16±0,07	0,17±0,08	0,14±0,05	0,14±0,05
	apq3 (%)	0,84±0,38	0,89±0,44	0,68±0,26	0,67±0,25
	apq5 (%)	1,06±0,50	1,07±0,52	1,05±0,43	1,01±0,39
/i/ Alto	Shim (%)	1,95±2,05	1,85±1,69	1,15±0,50	1,24±0,50
	Shdb (dB)	0,17±0,17	0,16±0,15	0,10±0,04	0,11±0,04
	apq3 (%)	0,81±0,67	0,79±0,60	0,49±0,23	0,55±0,26
	apq5 (%)	1,06±1,15	1,02±0,97	0,68±0,38	0,73±0,36
/u/ Baixo	Shim (%)	2,21±0,99	2,16±0,90	2,39±0,97	2,35±1,10
	Shdb (dB)	0,19±0,09	0,19±0,08	0,21±0,09	0,21±0,10
	apq3 (%)	1,03±0,48	1,04±0,48	1,09±0,46	1,07±0,51
	apq5 (%)	1,34±0,65	1,29±0,58	1,60±0,74	1,50±0,76
/u/ Normal	Shim (%)	2,37±1,37	2,41±1,61	1,90±0,60	1,83±0,57
	Shdb (dB)	0,21±0,12	0,21±0,15	0,17±0,05	0,16±0,05
	apq3 (%)	1,14±0,72	1,18±0,82	0,85±0,28	0,84±0,31
	apq5 (%)	1,45±0,90	1,49±1,08	1,22±0,45	1,14±0,41
/u/ Alto	Shim (%)	1,77±1,09	1,74±1,05	1,49±0,60	1,46±0,64
	Shdb (dB)	0,16±0,10	0,15±0,09	0,13±0,05	0,13±0,06
	apq3 (%)	0,87±0,58	0,87±0,58	0,66±0,30	0,63±0,31
	apq5 (%)	1,02±0,69	1,02±0,67	0,88±0,39	0,86±0,43

Na Tabela 20 apresentam-se as medidas do parâmetro Shimmer relativas ao grupo de controlo, para as vogais, efetuando a média entre tons. Os valores medidos pelo

#### IV – Análise da Precisão da Medida do Algoritmo

algoritmo e pelo Praat são muito próximos para todos os parâmetros. A maior diferença para o ShdB é apenas de 0.04 dB denotando uma grande consistência dos resultados medidos pelos dois métodos.

Tabela 20 - Média e desvio padrão dos parâmetros de Shimmer ordenados por vogal para os sinais de controle.

Vogal	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
/a/	Shim (%)	2,06±0,70	2,50±1,35	2,31±1,14	2,84±1,51
	Shdb (dB)	0,18±0,06	0,22±0,12	0,21±0,11	0,25±0,14
	apq3 (%)	1,00±0,41	1,30±0,78	1,11±0,64	1,42±0,78
	apq5 (%)	1,22±0,47	1,50±0,91	1,42±0,68	1,72±0,98
/i/	Shim (%)	1,92±1,30	1,87±1,14	1,66±0,78	1,68±0,78
	Shdb (dB)	0,17±0,11	0,16±0,10	0,15±0,07	0,15±0,07
	apq3 (%)	0,84±0,48	0,85±0,49	0,68±0,33	0,71±0,36
	apq5 (%)	1,08±0,75	1,05±0,66	1,07±0,60	1,06±0,54
/u/	Shim (%)	2,12±1,16	2,11±1,23	1,92±0,82	1,88±0,87
	Shdb (dB)	0,19±0,10	0,19±0,11	0,17±0,07	0,17±0,08
	apq3 (%)	1,01±0,59	1,03±0,64	0,87±0,39	0,85±0,42
	apq5 (%)	1,27±0,76	1,27±0,82	1,23±0,61	1,16±0,61

Tal como para o Jitter foi também feita uma comparação entre tons para o parâmetro Shimmer que se apresenta na tabela 21.

Os dados entre os dois métodos são novamente muito próximos para todos os parâmetros apresentados, sendo a maior diferença de apenas 0.02 dB.

Tabela 21 - : Média e desvio padrão dos parâmetros de Shimmer ordenados por tom para os sinais de controle.

Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
Baixo	Shim (%)	2,09±0,86	2,28±1,18	2,51±1,12	2,67±1,21
	Shdb (dB)	0,18±0,08	0,20±0,11	0,22±0,10	0,23±0,11
	apq3 (%)	0,96±0,47	1,11±0,68	1,13±0,57	1,23±0,63
	apq5 (%)	1,25±0,59	1,35±0,77	1,64±0,73	1,67±0,75
Normal	Shim (%)	2,09±0,98	2,30±1,35	1,94±0,68	2,19±1,34
	Shdb (dB)	0,18±0,09	0,20±0,12	0,17±0,06	0,19±0,12
	apq3 (%)	0,99±0,50	1,14±0,72	0,89±0,39	1,04±0,71
	apq5 (%)	1,23±0,63	1,38±0,91	1,25±0,48	1,38±0,89
Alto	Shim (%)	1,92±1,36	1,90±1,24	1,44±0,70	1,54±0,72
	Shdb (dB)	0,17±0,12	0,17±0,11	0,13±0,07	0,14±0,07
	apq3 (%)	0,91±0,55	0,93±0,59	0,65±0,42	0,71±0,41
	apq5 (%)	1,09±0,79	1,09±0,74	0,84±0,40	0,89±0,44

**4.2.1.3. HNR**

Foi também feita uma análise comparativa ao HNR. Na Tabela 22 apresentam-se os valores da variação deste parâmetro com as vogais /a/, /i/ e /u/, nos tons baixo normal e alto para os sinais de voz controlo com os dois métodos utilizados. Analisando os valores da tabela pode ver-se que as diferenças são menores para a vogal /a/ e maiores para as vogais /i/ e /u/. A maior diferença medida entre o algoritmo e o Praat é de cerca de 7 dB e ocorre para a vogal /u/ no tom alto para as vozes masculinas.

Tabela 22 - Média e desvio padrão do parâmetro HNR (dB) para os sinais de controlo.

Vogal/Tom	Feminino		Masculino	
	Algoritmo	Praat	Algoritmo	Praat
/a/ Baixo	23,80±1,28	23,54±3,76	23,12±1,32	22,74±2,71
/a/ Normal	24,00±0,97	24,35±4,12	23,61±1,70	23,42±3,91
/a/ Alto	24,61±0,61	25,70±3,20	24,54±0,67	27,64±2,84
/i/ Baixo	24,48±1,19	26,43±3,57	24,30±1,42	25,58±3,88
/i/ Normal	24,20±1,05	25,95±4,95	24,65±0,63	26,96±2,64
/i/ Alto	24,39±1,09	28,23±6,08	24,61±0,49	28,60±4,68
/u/ Baixo	24,59±0,39	29,34±3,28	24,31±0,73	27,32±3,28
/u/ Normal	24,42±0,93	29,17±5,79	24,54±0,53	28,90±2,76
/u/ Alto	24,79±0,48	31,74±5,29	24,67±0,77	31,81±2,54

Relacionando o HNR com a vogal, como demonstrado na Tabela 23, apesar de não existir qualquer diferença significativa entre os dois métodos ou género, existe uma pequena ascensão de valores da vogal /a/ para a vogal /i/ e desta para a vogal /u/, sendo agora a maior diferença medida de apenas 6 dB entre os dois métodos

Tabela 23 - Média e desvio padrão do parâmetro HNR (dB) ordenado por vogal para os sinais de controlo.

Vogal	Feminino		Masculino	
	Algoritmo	Praat	Algoritmo	Praat
/a/	24,14±1,03	24,53±3,74	23,76±1,40	24,60±3,82
/i/	24,36±1,09	26,87±4,96	24,52±0,93	27,04±3,94
/u/	24,60±0,65	30,08±4,94	24,51±0,69	29,34±3,38

Na Tabela 24 apresenta-se a média das medidas do parâmetro HNR nos três tons. Pode notar-se uma ligeira tendência de valores maiores de HNR medidos pelo Praat nos

três tons tanto para vozes femininas como masculinas. A maior diferença neste caso é cerca de 5 dB.

Tabela 24 - Média e desvio padrão do parâmetro HRN (dB) ordenado por tom para os sinais de controlo.

Tom	Feminino		Masculino	
	Algoritmo	Praat	Algoritmo	Praat
Baixo	24,29±1,07	26,44±4,21	23,91±1,30	25,21±3,77
Normal	24,21±0,98	26,49±5,29	24,27±1,16	26,42±3,84
Alto	24,60±0,77	28,56±5,49	24,61±0,64	29,35±3,85

Em resumo, da comparação das medidas de Jitter entre o algoritmo e o Praat a média das medidas difere no pior dos casos em apenas 11  $\mu$ s, sendo este valor inferior a uma amostra. No caso das medidas de Shimmer a maior diferença é de apenas 0.08 dB, podendo-se considerar insignificante, esta diferença. Contudo, na medida do HNR a maior diferença é de cerca de 7 dB.

Desta forma pode-se considerar que os valores medidos pelo algoritmo foram consistentes com os medidos pelo Praat, confirmando o algoritmo desenvolvido como muito preciso também na análise de vozes saudáveis.

#### 4.2.2. COMPARAÇÃO ALGORITMO/PRAAT PARA SINAIS DE VOZ PATOLÓGICOS

Como o objetivo é demonstrar a viabilidade do algoritmo, nesta secção só irá ser feita a análise de sinais de fala com a patologia laringite. A análise a esta patologia foi executada utilizando uma amostra de 15 pacientes referentes ao sexo feminino e 15 do sexo masculino. Os pacientes do sexo feminino apresentam idades compreendidas entre os 39 e 73 anos, tendo em média 51.73 anos e desvio padrão de 10.53. Os pacientes do sexo masculino apresentam idades compreendidas entre os 37 e 61 anos, tendo em média 52.87 anos e desvio padrão de 7.62.

#### 4.2.2.1. JITTER

Na Tabela 25 apresentam-se os dados (média e desvio padrão) das vogais /a/, /i/ e /u/, para os tons baixo, normal e alto das medidas de Jitter para os sinais de voz patológicos utilizando os dois métodos.

Através da análise destes resultados é possível denotar que os valores (média e desvio padrão) obtidos pelo algoritmo e pelo Praat são muito próximos para ambos os géneros, em todas as vogais e tons. A diferença de valores entre os dois métodos é muito pequena, sendo que os valores obtidos pelo algoritmo são tendencialmente superiores aos obtidos pelo Praat. Refira-se que a maior diferença de Jitter entre as medidas efetuadas pelo algoritmo e pelo Praat é de apenas 12  $\mu$ s, novamente inferior ao intervalo entre amostras sucessivas.

Analisando os dados em pormenor, para o parâmetro Jitter os valores relativos ao sexo masculino são normalmente mais elevados tanto para o Praat como para o algoritmo, existindo para este último uma exceção na vogal /u/ tom alto em que o valor de Jitter é superior para o sexo feminino. Mais uma vez, esta diferença é compreendida pelo fato do período glotal masculino ser habitualmente mais longo. Para os restantes parâmetros os valores apresentados pelo algoritmo são maioritariamente superiores no sexo feminino.

Comparando os valores desta tabela com os dados obtidos para os sinais de voz controlo (Tabela 16), é evidente que para todos os parâmetros e para os dois métodos os valores apresentados pelos sinais patológicos são sempre superiores aos sinais controlo, excetuando-se os valores para o sexo feminino nas vogais /i/ e /u/ nos tons baixo e normal respetivamente, os quais contrariamente ao esperado são ligeiramente superiores nos sinais de controlo para os dados do algoritmo e superiores ou iguais para o Praat.

IV – Análise da Precisão da Medida do Algoritmo

Tabela 25 - Média e desvio padrão das medidas de Jitter ordenados por vogal e tom para os sinais da patologia laringite.

Vogal/Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
/a/ Baixo	Jitta ( $\mu$ s)	35 $\pm$ 13	25 $\pm$ 12	45 $\pm$ 20	41 $\pm$ 21
	Jitt (%)	0,62 $\pm$ 0,26	0,44 $\pm$ 0,17	0,53 $\pm$ 0,21	0,47 $\pm$ 0,21
	rap (%)	0,36 $\pm$ 0,17	0,24 $\pm$ 0,10	0,28 $\pm$ 0,14	0,23 $\pm$ 0,14
	ppq5 (%)	0,35 $\pm$ 0,13	0,25 $\pm$ 0,09	0,30 $\pm$ 0,11	0,27 $\pm$ 0,12
/a/ Normal	Jitta ( $\mu$ s)	31 $\pm$ 10	24 $\pm$ 12	42 $\pm$ 29	40 $\pm$ 32
	Jitt (%)	0,61 $\pm$ 0,19	0,46 $\pm$ 0,18	0,52 $\pm$ 0,32	0,48 $\pm$ 0,35
	rap (%)	0,37 $\pm$ 0,12	0,27 $\pm$ 0,12	0,29 $\pm$ 0,21	0,25 $\pm$ 0,24
	ppq5 (%)	0,34 $\pm$ 0,08	0,26 $\pm$ 0,10	0,28 $\pm$ 0,14	0,26 $\pm$ 0,16
/a/ Alto	Jitta ( $\mu$ s)	31 $\pm$ 15	21 $\pm$ 14	34 $\pm$ 21	27 $\pm$ 23
	Jitt (%)	0,77 $\pm$ 0,39	0,46 $\pm$ 0,25	0,50 $\pm$ 0,25	0,39 $\pm$ 0,29
	rap (%)	0,48 $\pm$ 0,25	0,27 $\pm$ 0,16	0,29 $\pm$ 0,15	0,21 $\pm$ 0,19
	ppq5 (%)	0,43 $\pm$ 0,19	0,25 $\pm$ 0,13	0,29 $\pm$ 0,12	0,21 $\pm$ 0,14
/i/ Baixo	Jitta ( $\mu$ s)	25 $\pm$ 11	19 $\pm$ 9	38 $\pm$ 23	35 $\pm$ 21
	Jitt (%)	0,47 $\pm$ 0,22	0,35 $\pm$ 0,17	0,42 $\pm$ 0,17	0,37 $\pm$ 0,17
	rap (%)	0,28 $\pm$ 0,13	0,19 $\pm$ 0,09	0,20 $\pm$ 0,10	0,16 $\pm$ 0,10
	ppq5 (%)	0,27 $\pm$ 0,11	0,20 $\pm$ 0,09	0,24 $\pm$ 0,10	0,21 $\pm$ 0,11
/i/ Normal	Jitta ( $\mu$ s)	24 $\pm$ 12	15 $\pm$ 9	31 $\pm$ 17	26 $\pm$ 14
	Jitt (%)	0,52 $\pm$ 0,21	0,32 $\pm$ 0,16	0,39 $\pm$ 0,16	0,32 $\pm$ 0,13
	rap (%)	0,32 $\pm$ 0,14	0,18 $\pm$ 0,11	0,20 $\pm$ 0,09	0,15 $\pm$ 0,07
	ppq5 (%)	0,29 $\pm$ 0,08	0,23 $\pm$ 0,22	0,23 $\pm$ 0,09	0,19 $\pm$ 0,09
/i/ Alto	Jitta ( $\mu$ s)	19 $\pm$ 10	13 $\pm$ 11	29 $\pm$ 23	25 $\pm$ 23
	Jitt (%)	0,47 $\pm$ 0,19	0,31 $\pm$ 0,22	0,45 $\pm$ 0,30	0,37 $\pm$ 0,30
	rap (%)	0,30 $\pm$ 0,12	0,18 $\pm$ 0,14	0,26 $\pm$ 0,19	0,20 $\pm$ 0,19
	ppq5 (%)	0,28 $\pm$ 0,09	0,17 $\pm$ 0,12	0,24 $\pm$ 0,12	0,19 $\pm$ 0,13
/u/ Baixo	Jitta ( $\mu$ s)	28 $\pm$ 12	24 $\pm$ 15	39 $\pm$ 19	33 $\pm$ 18
	Jitt (%)	0,50 $\pm$ 0,19	0,42 $\pm$ 0,22	0,45 $\pm$ 0,14	0,37 $\pm$ 0,14
	rap (%)	0,30 $\pm$ 0,12	0,22 $\pm$ 0,12	0,22 $\pm$ 0,08	0,16 $\pm$ 0,07
	ppq5 (%)	0,29 $\pm$ 0,11	0,22 $\pm$ 0,11	0,26 $\pm$ 0,08	0,21 $\pm$ 0,09
/u/ Normal	Jitta ( $\mu$ s)	24 $\pm$ 8	16 $\pm$ 7	43 $\pm$ 42	31 $\pm$ 34
	Jitt (%)	0,50 $\pm$ 0,15	0,33 $\pm$ 0,12	0,51 $\pm$ 0,35	0,37 $\pm$ 0,29
	rap (%)	0,30 $\pm$ 0,09	0,18 $\pm$ 0,07	0,28 $\pm$ 0,22	0,18 $\pm$ 0,19
	ppq5 (%)	0,30 $\pm$ 0,08	0,19 $\pm$ 0,07	0,29 $\pm$ 0,19	0,21 $\pm$ 0,17
/u/ Alto	Jitta ( $\mu$ s)	27 $\pm$ 13	15 $\pm$ 12	23 $\pm$ 10	17 $\pm$ 9
	Jitt (%)	0,63 $\pm$ 0,18	0,33 $\pm$ 0,13	0,37 $\pm$ 0,13	0,26 $\pm$ 0,08
	rap (%)	0,39 $\pm$ 0,11	0,18 $\pm$ 0,07	0,20 $\pm$ 0,08	0,11 $\pm$ 0,04
	ppq5 (%)	0,38 $\pm$ 0,11	0,19 $\pm$ 0,07	0,22 $\pm$ 0,06	0,14 $\pm$ 0,04

Na Tabela 26 encontrar-se a análise das medidas de Jitter por vogal, realizando a média entre tons, para os sinais de voz com a patologia laringite. Neste caso a maior

#### IV – Análise da Precisão da Medida do Algoritmo

diferença para o Jitta entre o algoritmo e o Praat é apenas de 9  $\mu$ s, ainda menor que na situação anterior.

Comparando os dois géneros, tal como aconteceu para este tipo de comparação para os sinais de voz de controlo, para o parâmetro Jitta os valores referentes ao sexo masculino são sempre mais elevados relativamente aos do sexo feminino e para os restantes parâmetros os valores referentes ao sexo feminino são sempre superiores para ambos os métodos.

Comparando os sinais de controlo com os patológicos, para ambos os métodos os valores referentes aos sinais de voz patológicos são sempre mais elevados, exceto para a vogal /a/, que para o parâmetro ppq5 referente ao sexo feminino os valores relativos aos sinais de voz patológicos são inferiores aos de controlo.

**Tabela 26 - Média e desvio padrão das medidas de Jitter ordenados por vogal para os sinais da patologia laringite.**

Vogal	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
/a/	Jitta ( $\mu$ s)	32 $\pm$ 13	23 $\pm$ 12	40 $\pm$ 24	36 $\pm$ 26
	Jitt (%)	0,67 $\pm$ 0,29	0,45 $\pm$ 0,20	0,52 $\pm$ 0,26	0,45 $\pm$ 0,29
	rap (%)	0,40 $\pm$ 0,19	0,26 $\pm$ 0,13	0,28 $\pm$ 0,16	0,23 $\pm$ 0,19
	ppq5 (%)	0,37 $\pm$ 0,15	0,25 $\pm$ 0,11	0,29 $\pm$ 0,12	0,25 $\pm$ 0,14
/i/	Jitta ( $\mu$ s)	23 $\pm$ 11	16 $\pm$ 10	33 $\pm$ 21	28 $\pm$ 20
	Jitt (%)	0,49 $\pm$ 0,20	0,33 $\pm$ 0,18	0,42 $\pm$ 0,22	0,35 $\pm$ 0,21
	rap (%)	0,30 $\pm$ 0,13	0,18 $\pm$ 0,11	0,22 $\pm$ 0,14	0,17 $\pm$ 0,13
	ppq5 (%)	0,28 $\pm$ 0,09	0,20 $\pm$ 0,15	0,24 $\pm$ 0,10	0,20 $\pm$ 0,11
/u/	Jitta ( $\mu$ s)	26 $\pm$ 11	18 $\pm$ 12	35 $\pm$ 28	27 $\pm$ 23
	Jitt (%)	0,55 $\pm$ 0,18	0,36 $\pm$ 0,17	0,44 $\pm$ 0,23	0,33 $\pm$ 0,19
	rap (%)	0,33 $\pm$ 0,11	0,19 $\pm$ 0,09	0,24 $\pm$ 0,14	0,15 $\pm$ 0,12
	ppq5 (%)	0,32 $\pm$ 0,11	0,20 $\pm$ 0,09	0,26 $\pm$ 0,12	0,19 $\pm$ 0,11

Na Tabela 27 são apresentados os parâmetros de Jitter referentes aos sinais de voz com a patologia laringite selecionados segundo o tom, após efetuar a média entre vogais. Os valores medidos do algoritmo voltam a ser tendencialmente mais elevados que os do Praat para todos os parâmetros, apesar de não haver nenhuma diferença significativa. A maior diferença para o Jitta é agora de 10  $\mu$ s, continuando abaixo do período de amostragem, sendo inclusivamente uma diferença inferior ao tempo correspondente a metade da diferença de uma amostra.

#### IV – Análise da Precisão da Medida do Algoritmo

Para o parâmetro Jitta os valores do sexo masculino voltam a ser superiores aos apresentados pelo sexo feminino para ambos os métodos. Quanto aos valores dos restantes parâmetros, estes voltam a ser mais elevados para o sexo feminino nos dados apresentados pelo algoritmo

Os valores apresentados nesta tabela são sempre superiores para todos os parâmetros em ambos os métodos quando relacionados com os valores medidos no grupo de controlo.

Tabela 27 - Média e desvio padrão das medidas de Jitter ordenados por tom para os sinais da patologia laringite.

Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
Baixo	Jitta ( $\mu$ s)	29 $\pm$ 12	23 $\pm$ 12	41 $\pm$ 20	37 $\pm$ 20
	Jitt (%)	0,53 $\pm$ 0,23	0,40 $\pm$ 0,19	0,46 $\pm$ 0,18	0,40 $\pm$ 0,18
	rap (%)	0,31 $\pm$ 0,14	0,22 $\pm$ 0,10	0,23 $\pm$ 0,11	0,19 $\pm$ 0,11
	ppq5 (%)	0,30 $\pm$ 0,12	0,22 $\pm$ 0,10	0,27 $\pm$ 0,10	0,23 $\pm$ 0,11
Normal	Jitta ( $\mu$ s)	26 $\pm$ 10	18 $\pm$ 10	39 $\pm$ 31	32 $\pm$ 28
	Jitt (%)	0,54 $\pm$ 0,19	0,37 $\pm$ 0,17	0,47 $\pm$ 0,29	0,39 $\pm$ 0,27
	rap (%)	0,33 $\pm$ 0,12	0,21 $\pm$ 0,11	0,26 $\pm$ 0,18	0,19 $\pm$ 0,18
	ppq5 (%)	0,31 $\pm$ 0,08	0,22 $\pm$ 0,14	0,27 $\pm$ 0,15	0,22 $\pm$ 0,14
Alto	Jitta ( $\mu$ s)	26 $\pm$ 14	16 $\pm$ 13	29 $\pm$ 19	23 $\pm$ 19
	Jitt (%)	0,63 $\pm$ 0,29	0,37 $\pm$ 0,21	0,44 $\pm$ 0,24	0,34 $\pm$ 0,25
	rap (%)	0,39 $\pm$ 0,18	0,21 $\pm$ 0,14	0,25 $\pm$ 0,15	0,17 $\pm$ 0,16
	ppq5 (%)	0,36 $\pm$ 0,15	0,20 $\pm$ 0,11	0,25 $\pm$ 0,11	0,18 $\pm$ 0,12

#### 4.2.2.2. SHIMMER

Os valores de shimmer ordenados por vogal e ordem referentes aos sinais de voz com a patologia laringite estão dispostos na Tabela 28.

Comparando os dois métodos, para todos os tons da vogal /a/ o Praat apresentou sempre valores superiores em relação ao algoritmo para todas as medidas, para todos os tons da vogal /i/ os valores apresentados pelos dois métodos são muito aproximados, para a vogal /u/ não existe nenhuma uniformidade que permita tirar alguma conclusão. Neste caso a maior diferença para o parâmetro ShdB é de apenas 0.13 dB.

Os valores exibidos pelos dois géneros são demasiado próximos, com a exceção da vogal /u/ no tom alto em que os dados do sexo feminino são bastante superiores aos do sexo masculino para ambos os métodos.

IV – Análise da Precisão da Medida do Algoritmo

Não é evidente qualquer tipo de característica neste parâmetro quando são comparados os valores dos sinais controle com os sinais desta patologia que permita tirar alguma conclusão.

Tabela 28 - Média e desvio padrão das medidas de Shimmer ordenados por vogal e tom para os sinais da patologia laringite.

Vogal/Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
/a/ Baixo	Shim (%)	2,55±1,11	2,61±0,98	2,49±1,36	3,72±2,35
	Shdb (dB)	0,23±0,11	0,24±0,10	0,22±0,12	0,33±0,22
	apq3 (%)	1,27±0,70	1,36±0,56	1,20±0,73	1,95±1,26
	apq5 (%)	1,49±0,58	1,53±0,56	1,58±0,89	2,34±1,67
/a/ Normal	Shim (%)	2,06±0,71	2,91±1,18	2,49±1,59	3,92±2,31
	Shdb (dB)	0,18±0,06	0,26±0,11	0,22±0,14	0,35±0,22
	apq3 (%)	1,08±0,43	1,61±0,74	1,25±0,98	1,98±1,20
	apq5 (%)	1,20±0,43	1,70±0,65	1,50±0,87	2,37±1,54
/a/ Alto	Shim (%)	2,77±1,47	3,19±1,85	2,16±1,43	2,84±2,46
	Shdb (dB)	0,24±0,13	0,29±0,17	0,19±0,13	0,25±0,23
	apq3 (%)	1,48±0,94	1,77±1,15	1,11±0,87	1,50±1,35
	apq5 (%)	1,59±0,84	1,84±1,14	1,25±0,77	1,74±1,66
/i/ Baixo	Shim (%)	1,64±0,70	1,65±0,68	2,42±1,79	2,23±1,41
	Shdb (dB)	0,15±0,06	0,15±0,06	0,21±0,16	0,20±0,12
	apq3 (%)	0,76±0,43	0,77±0,39	0,99±0,71	0,93±0,51
	apq5 (%)	0,96±0,37	0,96±0,37	1,57±1,20	1,40±0,93
/i/ Normal	Shim (%)	1,72±1,15	1,74±1,18	1,62±0,97	1,61±0,87
	Shdb (dB)	0,15±0,10	0,15±0,10	0,14±0,08	0,14±0,08
	apq3 (%)	0,85±0,71	0,89±0,73	0,77±0,61	0,77±0,56
	apq5 (%)	0,98±0,68	0,99±0,68	0,98±0,52	0,98±0,48
/i/ Alto	Shim (%)	1,43±0,92	1,46±0,96	1,51±0,89	1,51±0,91
	Shdb (dB)	0,13±0,08	0,13±0,08	0,13±0,08	0,13±0,08
	apq3 (%)	0,72±0,55	0,73±0,58	0,69±0,50	0,71±0,51
	apq5 (%)	0,81±0,57	0,83±0,60	0,90±0,56	0,89±0,55
/u/ Baixo	Shim (%)	2,48±1,34	2,51±1,21	2,62±1,67	2,20±1,11
	Shdb (dB)	0,22±0,12	0,22±0,11	0,23±0,15	0,19±0,10
	apq3 (%)	1,18±0,74	1,22±0,69	1,20±0,94	1,01±0,58
	apq5 (%)	1,55±0,97	1,55±0,90	1,66±1,02	1,38±0,75
/u/ Normal	Shim (%)	2,50±1,57	2,30±0,86	2,60±1,79	2,21±1,19
	Shdb (dB)	0,23±0,19	0,20±0,08	0,23±0,16	0,19±0,11
	apq3 (%)	1,26±0,88	1,16±0,46	1,35±1,02	1,13±0,69
	apq5 (%)	1,48±0,69	1,43±0,56	1,70±1,25	1,45±0,86
/u/ Alto	Shim (%)	2,01±1,05	2,13±1,15	1,67±1,17	1,52±0,79
	Shdb (dB)	0,18±0,09	0,19±0,10	0,15±0,10	0,13±0,07
	apq3 (%)	1,00±0,53	1,07±0,60	0,81±0,66	0,72±0,42
	apq5 (%)	1,24±0,67	1,28±0,70	1,08±0,83	0,96±0,52

#### IV – Análise da Precisão da Medida do Algoritmo

Os dados do parâmetro Shimmer relativos aos pacientes com a patologia laringite aos quais se procedeu a uma análise por vogal estão representados na Tabela 29.

Comparando os dois métodos, para a vogal /a/ o Praat apresenta tendencialmente valores mais elevados em relação ao algoritmo, apresentando valores para todos os parâmetros mais elevados para o sexo masculino, ao contrário do que acontece com o algoritmo que apresenta dados mais elevados para o sexo feminino. Para a vogal /i/ os dois métodos apresentam valores bastante semelhantes, existindo um pequeno aumento de valores do sexo masculino em comparação com o sexo feminino. Quanto à vogal /u/, para o sexo feminino os valores para os dois métodos são muito aproximados, já para o sexo masculino existe uma pequena discrepância de valores entre os dois métodos. Neste caso a maior diferença de ShdB é de 0.10 dB.

Comparando o mesmo tipo de tabelas entre sinais controlo e patológicos, não é observável qualquer tipo de característica distintiva entre estes.

**Tabela 29 - Média e desvio padrão das medidas de Shimmer ordenados por vogal para os sinais da patologia laringite.**

Vogal	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
/a/	Shim (%)	2,46±1,15	2,90±1,38	2,38±1,44	3,49±2,37
	Shdb (dB)	0,22±0,11	0,26±0,13	0,21±0,13	0,31±0,22
	apq3 (%)	1,28±0,72	1,58±0,85	1,19±0,85	1,81±1,26
	apq5 (%)	1,43±0,64	1,69±0,81	1,44±0,84	2,15±1,61
/i/	Shim (%)	1,60±0,93	1,62±0,95	1,85±1,32	1,79±1,11
	Shdb (dB)	0,14±0,08	0,14±0,08	0,16±0,12	0,16±0,10
	apq3 (%)	0,78±0,56	0,80±0,58	0,82±0,61	0,80±0,53
	apq5 (%)	0,92±0,55	0,93±0,56	1,15±0,86	1,09±0,70
/u/	Shim (%)	2,33±1,32	2,31±1,08	2,30±1,60	1,98±1,07
	Shdb (dB)	0,21±0,14	0,21±0,10	0,20±0,14	0,17±0,09
	apq3 (%)	1,15±0,72	1,15±0,58	1,12±0,90	0,95±0,59
	apq5 (%)	1,42±0,78	1,42±0,73	1,48±1,06	1,26±0,74

Na Tabela 30 representam-se os dados de Shimmer relativos aos sinais de voz patológicos por tons. Os dados obtidos pelos dois métodos são muito semelhantes para todos os parâmetros apresentados. A maior diferença de ShdB é agora apenas de 0.03 dB.

#### IV – Análise da Precisão da Medida do Algoritmo

Os valores apresentados pelos dois métodos são superiores para o sexo masculino quando comparado com o feminino, com a exceção do tom alto no qual acontece o oposto.

Não é visível novamente uma diferença entre sinais de controlo e patológicos devido ao facto de os valores apresentados por estes serem muito aproximados.

**Tabela 30 - Média e desvio padrão das medidas de Shimmer ordenados por tom para os sinais da patologia laringite.**

Tom	Parâmetro	Feminino		Masculino	
		Algoritmo	Praat	Algoritmo	Praat
Baixo	Shim (%)	2,22±1,14	2,26±1,07	2,51±1,58	2,72±1,82
	Shdb (dB)	0,20±0,10	0,20±0,10	0,22±0,14	0,24±0,16
	apq3 (%)	1,07±0,67	1,12±0,61	1,13±0,79	1,30±0,96
	apq5 (%)	1,34±0,72	1,35±0,69	1,60±1,02	1,71±1,24
Normal	Shim (%)	2,09±1,21	2,32±1,16	2,24±1,52	2,58±1,83
	Shdb (dB)	0,19±0,13	0,21±0,10	0,20±0,13	0,23±0,17
	apq3 (%)	1,07±0,70	1,22±0,71	1,12±0,91	1,29±0,98
	apq5 (%)	1,22±0,63	1,37±0,69	1,39±0,96	1,60±1,18
Alto	Shim (%)	2,07±1,27	2,26±1,53	1,78±1,19	1,96±1,67
	Shdb (dB)	0,18±0,11	0,20±0,14	0,16±0,11	0,17±0,15
	apq3 (%)	1,06±0,75	1,19±0,91	0,87±0,70	0,98±0,92
	apq5 (%)	1,21±0,76	1,32±0,93	1,08±0,73	1,20±1,10

#### 4.2.2.3. HNR

Nas Tabelas 31, 32 e 33 apresentam-se os valores medidos relativos ao parâmetro HNR para os sinais de voz com a patologia laringite. Na tabela 31, onde se apresentam as medidas por vogal e tom a maior diferença entre as medidas do algoritmo e do Praat é de cerca de 6 dB. Na tabela 32, onde se apresentam as medidas de HNR por vogal, a maior diferença é agora de aproximadamente 4 dB. Na tabela 33, apresentam-se os valores medidos de HNR agrupados por tom, a maior diferença é de cerca de 4 dB também. Para esta medida o algoritmo parece não apresentar uma gama de medidas tão extensa como o Praat.

Na comparação entre os grupos de controlo e o grupo com a patologia laringite para este parâmetro não são observáveis diferenças sem uma análise estatística mais profunda.

#### IV – Análise da Precisão da Medida do Algoritmo

Tabela 31 - Média e desvio padrão do parâmetro HNR para os sinais da patologia laringite.

Vogal/Tom	Feminino		Masculino	
	Algoritmo	Praat	Algoritmo	Praat
/a/ Baixo	23,87±0,94	23,34±3,09	22,35±2,10	22,39±4,61
/a/ Normal	23,88±0,95	23,40±3,24	22,82±2,13	22,96±4,57
/a/ Alto	23,80±1,29	22,89±4,08	23,68±1,53	25,56±4,59
/i/ Baixo	24,63±0,87	26,97±4,60	23,32±2,04	24,31±5,45
/i/ Normal	24,23±1,01	27,14±5,30	23,96±1,30	26,43±3,90
/i/ Alto	24,48±0,64	28,64±5,71	24,22±0,65	27,48±4,024
/u/ Baixo	24,12±1,05	27,84±4,29	23,85±1,02	26,70±4,78
/u/ Normal	24,27±0,74	28,37±2,89	24,05±1,95	27,37±4,28
/u/ Alto	25,24±1,40	29,38±4,67	24,70±0,58	30,92±3,82

Tabela 32 - Média e desvio padrão do parâmetro HNR ordenado por vogal para os sinais da patologia laringite.

Vogal	Feminino		Masculino	
	Algoritmo	Praat	Algoritmo	Praat
/a/	23,85±1,05	23,21±3,42	22,95±1,98	23,64±4,70
/i/	24,45±0,85	27,58±5,16	23,83±1,46	26,08±4,67
/u/	24,55±1,18	28,53±3,98	24,20±1,34	28,33±4,61

Tabela 33 - Média e desvio padrão do parâmetro HRN ordenado por tom para os sinais da patologia laringite.

Tom	Feminino		Masculino	
	Algoritmo	Praat	Algoritmo	Praat
l	24,21±0,99	26,05±4,42	23,18±1,86	24,47±5,16
n	24,13±0,90	26,30±4,41	23,61±1,88	25,59±4,58
h	24,51±1,28	26,97±5,58	24,20±1,08	27,98±4,70

#### 4.2.3. RESUMO DA COMPARAÇÃO DAS MEDIDAS COM SINAIS REAIS

Na tabela 34, apresenta-se um resumo das maiores diferenças entre as médias dos valores medidos pelo algoritmo e pelo programa Praat para os parâmetros Jitta, ShdB e HNR. Registe-se que estas diferenças maiores foram registadas nas médias dos sinais agrupados por vogais e tons.

Tabela 34 – Diferenças entre médias dos valores medidos pelo algoritmo e pelo programa Praat.

	Grupo de controlo	Grupo patologia da laringe
Jitta	11 $\mu$ s	12 $\mu$ s
ShdB	0.08 dB	0.13 dB
HNR	7 dB	6 dB

#### IV – Análise da Precisão da Medida do Algoritmo

Uma análise às maiores diferenças apresentadas na tabela 34 permite verificar que o comportamento do algoritmo em relação ao Praat é muito semelhante para as vozes de controlo e para as vozes patológicas. Deve ainda concluir-se que para a medida do Jitta as maiores diferenças são de 11 e 12  $\mu$ s, que é um intervalo de tempo inferior ao tempo entre duas amostras. Isto significa que a diferença nas medidas do algoritmo e do Praat não chegam a ser de uma amostra do sinal discreto. Em relação ao parâmetro ShdB as diferenças são de 0.08 e 0.13 dB, que se podem considerar como uma diferença de amplitudes insignificante. Já em relação ao HNR as diferenças são mais significativas e situam-se na ordem dos 6 a 7 dB.

De uma forma geral, nesta análise dos dois métodos para sinais de voz com a patologia laringite, o algoritmo voltou a apresentar valores ligeiramente mais elevados que o Praat, sendo esses repetidamente inferiores a um período de amostragem quando avaliados pelo parâmetro Jitta. Após estas análises pode-se concluir que o algoritmo apresenta valores consistentes com os medidos pelo programa Praat, com boa precisão, considerando a avaliação feita com os sinais sintetizados, e que estes fatores validam a sua utilização nos processos de avaliação dos parâmetros mencionados acima.

## CAPÍTULO V

---

### 5. COMPARAÇÃO DOS SINAIS DO GRUPO DE CONTROLO COM SINAIS DOS GRUPOS PATOLÓGICOS

Pelas razões mencionadas acima, as medidas efetuadas pelo algoritmo são considerados como corretas. Assim, procedendo-se de seguida à comparação dos 9 parâmetros entre os pacientes controlo e cinco patologias usando as medidas realizadas com o algoritmo desenvolvido. As patologias utilizadas foram a Laringite, Disfonia Hiperfuncional, Disfonia Espasmódica, Pólipo das Cordas Vocais e Envelhecimento das Cordas Vocais. Na secção 2 faz-se uma breve descrição de cada uma destas patologias.

Para esta análise utilizaram-se diagramas de caixa (*box plot*) pois estes possibilitam uma fácil visualização da semelhança entre parâmetros e grupos de vozes. Estes diagramas apresentam para cada parâmetro (no eixo horizontal), uma linha dentro da caixa com o valor da mediana, depois uma caixa acima da mediana com o quartil superior e outra caixa abaixo da mediana com o quartil inferior. Depois são traçadas duas linhas, uma em cima e outra em baixo que correspondem ao valor do quartil respetivo mais ou menos 1,5 vezes a diferenças entre o quartil superior e o inferior [linha superior= $Q_s+1,5(Q_s-Q_i)$ ; linha inferior= $Q_i-1,5(Q_s-Q_i)$ ]. Se estes valores forem superiores aos medidos as linhas são traçadas no valor correspondente ao menor valor medido para a linha inferior e no maior valor medido para a linha superior. Fora destas duas linhas são considerado valores atípicos ou *outliers*.

Das patologias mencionadas acima, a que apresentava menor número de pacientes era o Envelhecimento das Cordas Vocais e como a utilização destes diagramas só permite a comparação de matrizes com dimensões iguais foram utilizados 22 sinais de voz para cada grupo de análise. Desta forma o grupo de controlo (C) ficou composto por 11 indivíduos do sexo masculino e 11 do sexo feminino com média de idades de 52.68 anos, a grupo patologia Laringite (P1) por 11 pacientes masculinos e 11 femininos e com média de idades de 50.45, o grupo com a patologia Disfonia Hiperfuncional (P2) por 7 pacientes masculinos e 15 femininos e com média de idades de 50.14 anos, o grupo com a patologia Disfonia Espasmódica (P3) por 11 masculinos e 11 femininos e com média de idades de 55.59 anos, o grupo com a patologia Pólipo das Cordas Vocais

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

(P4) por 14 masculinos e 8 femininos e com média de idades de 50.27 anos e o grupo Envelhecimento das Cordas Vocais (P5) com 7 masculinos e 15 femininos e com média de idades de 75.32 anos. A tabela 35 apresenta um resumo da composição destes grupos.

Tabela 35 – Resumo das idades e média de idades dos grupos utilizados no estudo.

<b>Grupo</b>	<b>Masculinos</b>	<b>Femininos</b>	<b>Média Idades</b>
<b>Controlo C</b>	11	11	52,68
<b>Laringite (P1)</b>	11	11	50,45
<b>Disfonia Hiperfuncional(P2)</b>	7	15	50,14
<b>Disfonia Espasmódica(P3)</b>	11	11	55,59
<b>Pólipo das Cordas Vocais (P4)</b>	14	8	50,27
<b>Envelhecimento das Cordas Vocais (P5)</b>	7	15	75,32

Procurou-se equilibrar a idade média de cada grupo de forma aos resultados não serem afetados pela idade. No caso da patologia Envelhecimento da Cordas Vocais, a idade média é necessariamente maior que nos restantes grupos pelo facto desta patologia estar fortemente relacionada com o envelhecimento dos pacientes e não ser possível reunir da base de dados um grupo com esta patologia com uma média de idades mais baixa. A seleção das patologias obedeceu à disponibilidade de um grupo significativamente elevado na base de dados SVD.

O objetivo desta comparação é encontrar diferenças nos valores dos parâmetros entre os grupos de controlo e os grupos das patologias em estudo. As diferenças são consideradas estatisticamente significativas quando a gama de valores de um grupo (caixa) está diferenciada de outro grupo, isto é as caixas não partilham a mesma linha horizontal.

Inicialmente esta comparação foi efetuada sem fazer distinção de tom ou vogal, como demonstrado em baixo. Para cada uma destas caixas foram considerados os 22 elementos do grupo vezes as nove vocalizações para as 3 vogais e 3 tons, num total de  $22 \times 9 = 198$  medidas.

Para o parâmetro Jitta ilustrado na Figura 20, só existe diferença significativa para a patologia Pólipo nas Cordas Vocais (P4) relativamente ao grupo de controlo (C). Pois neste caso os valores das caixas não se tocam. As restantes patologias apesar de apresentarem valores tendencialmente superiores aos do grupo de controlo, como as caixas se tocam, não se podem considerar grupos estatisticamente distintos.

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

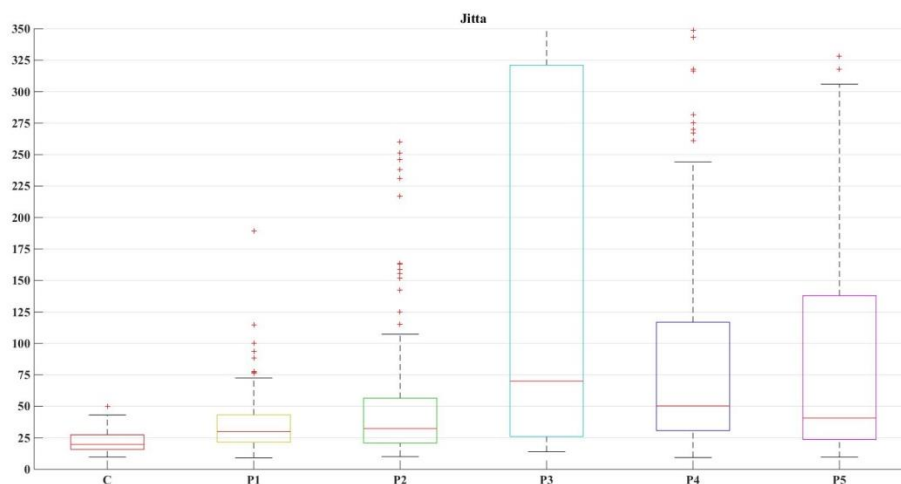


Figura 20 - Gráfico ilustrativo dos vários grupos para o parâmetro Jitta.

O parâmetro Jitt, ilustrado abaixo na Fig. 21, apresenta valores superiores para as patologias P3, P4 e P5 relativamente aos apresentados pelo grupo de controlo. Para as patologias P1 e P2 os valores são tendencialmente mais elevados mas não se podem considerar como grupos diferentes.

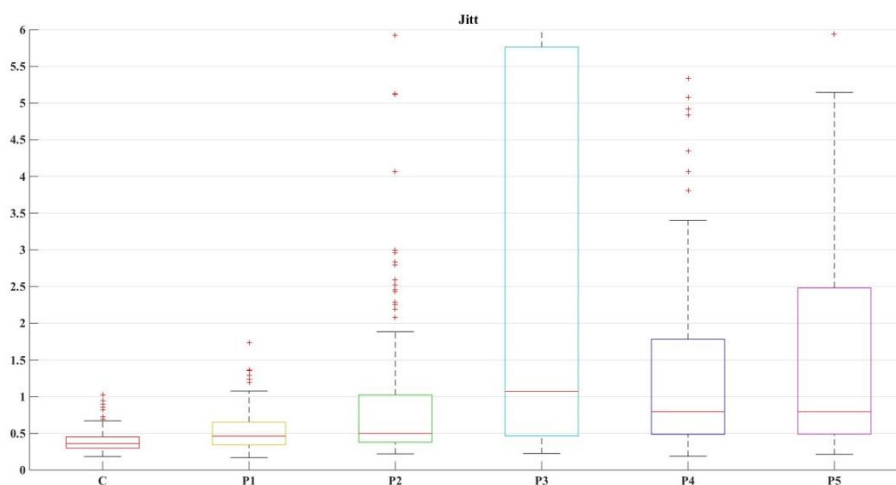


Figura 21 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Jitt.

Analisando o gráfico do parâmetro Rap (Figura 22), existem diferenças do grupo de controlo para as patologias P3 e P4.

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

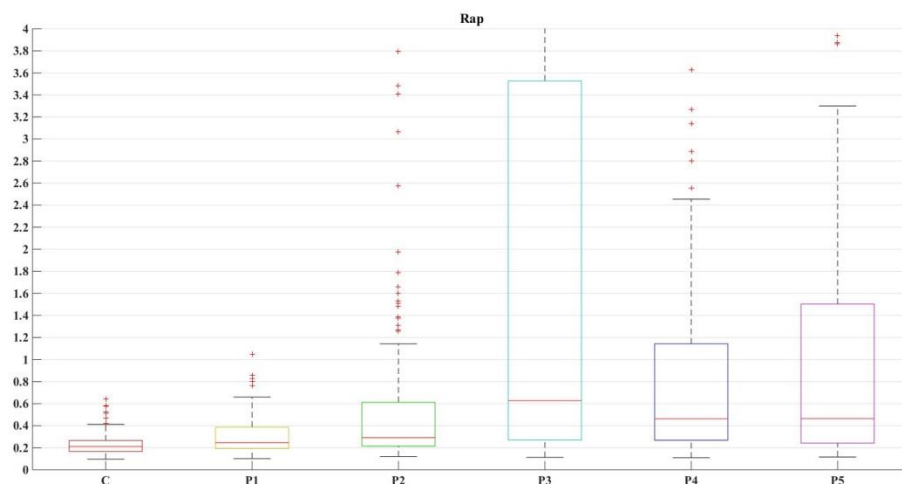


Figura 22 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Rap.

Na Figura 23 apresentam-se os dados relativos ao parâmetro ppq5, no qual existe distinção entre o grupo de controlo para as patologias Disfonia Espasmódica (P3), Pólipo das Cordas Vocais (P4) e Envelhecimento das Cordas Vocais (P5). As patologias P1 e P2 apresentam valores tendencialmente superiores aos do grupo de controlo.

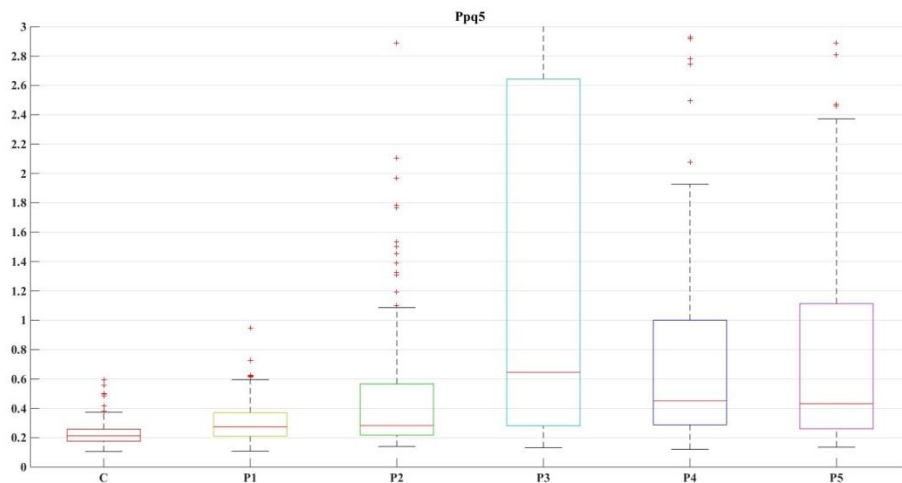


Figura 23 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Ppq5.

Para o parâmetro Shim (Figura 24), a patologia que apresenta valores mais distintos do grupo de controlo é a patologia Disfonia Espasmódica (P3), as restantes patologias têm valores tendencialmente superiores aos do grupo de controlo.

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

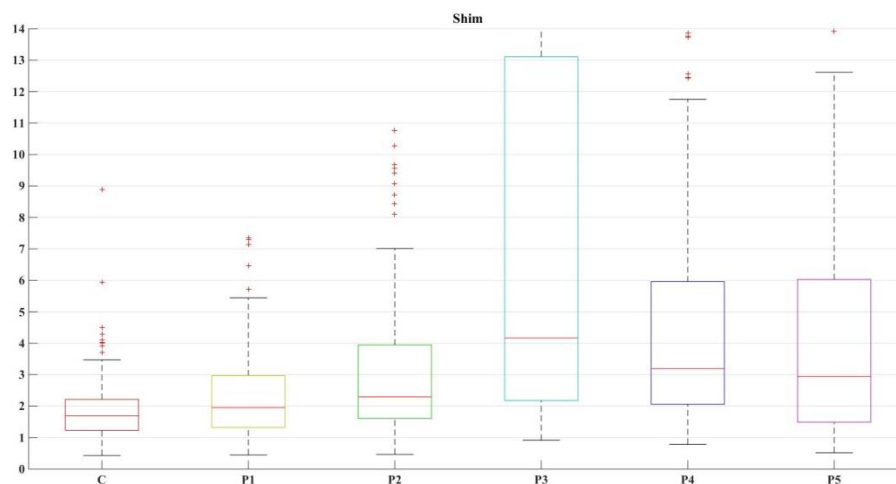


Figura 24 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Shim.

Para o parâmetro Shdb ilustrado abaixo, só existe distinção entre o grupo de valores da patologia Disfonia Espasmódica (P3) e o grupo controlo. As restantes patologias também apresentam valores tendencialmente superiores aos do grupo de controlo.

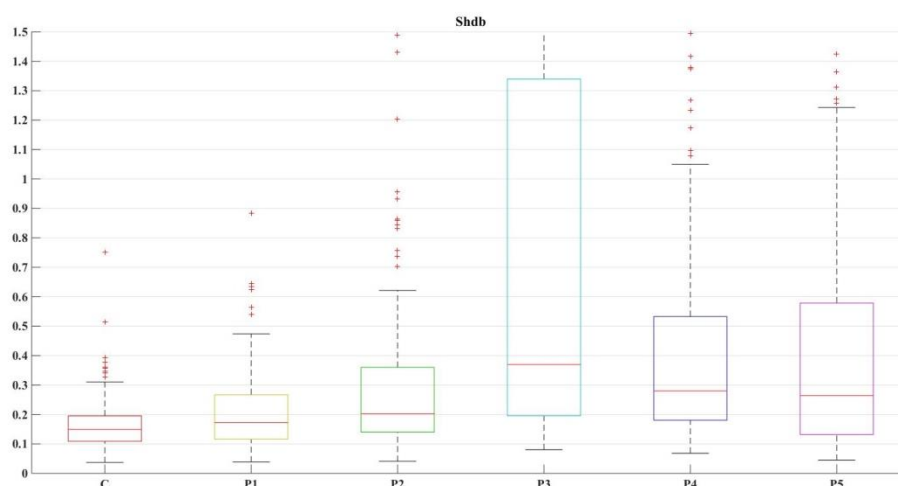


Figura 25 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Shdb.

Os parâmetros Apq3 (Figura 26) e Apq5 (Figura 27), não apresentam nenhum grupo patológico claramente distinto do grupo de controlo. Contudo, nos dois parâmetros, todos os grupos patológicos apresentam valores tendencialmente superiores aos do grupo de controlo.

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

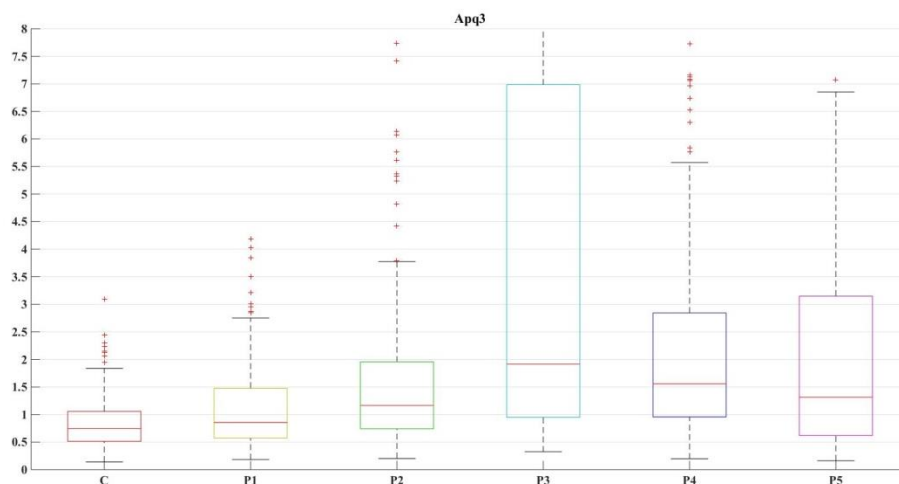


Figura 26 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Apq3.

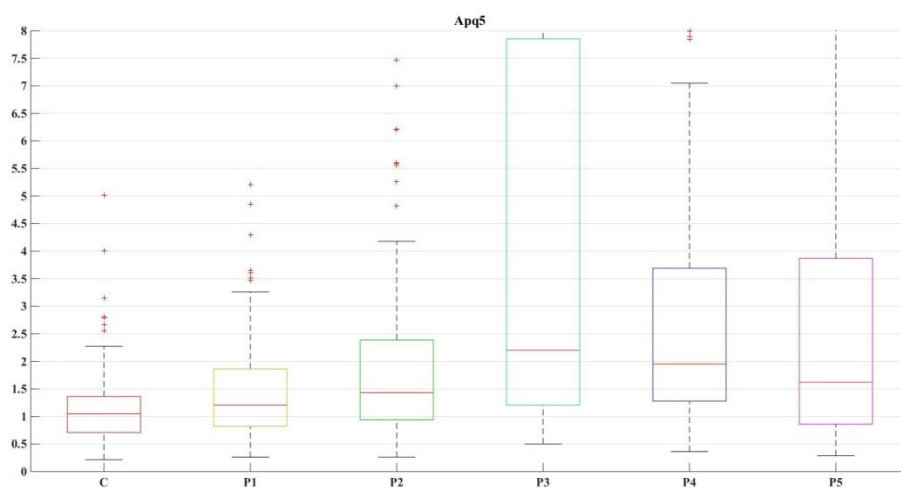
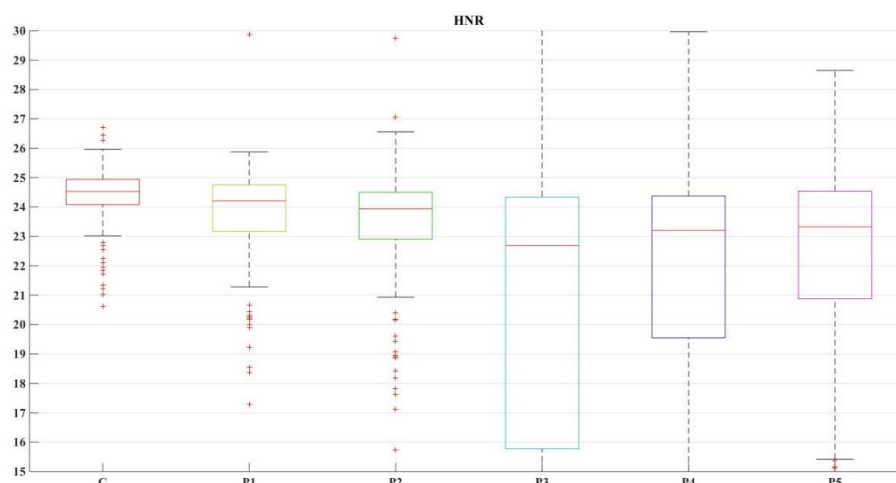


Figura 27 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro Apq5.

Para o parâmetro HNR (Figura 28) nenhum dos grupos patológicos apresenta uma diferença estatisticamente significativa do grupo controlo. Contudo, todos os grupos de patologias têm uma tendência para terem menor valor de HNR que o grupo de controlo.

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos



**Figura 28 - Gráfico ilustrativo dos vários grupos de pacientes para o parâmetro HNR.**

Na tabela 36 apresenta-se um resumo dos parâmetros que apresentaram relevância estatística de evidentes diferenças entre cada patologia e o grupo de controlo. Pode verificar-se que as medidas de Jitter são mais reveladoras dos grupos patológicos. Pode também verificar-se que para as patologias P1 e P2, nenhum dos parâmetros evidenciou diferenças significativas, apesar de haver tendências para valores mais elevados nos grupos com patologias.

**Tabela 36 – Resumo da variação dos parâmetros dos grupos patológicos relativamente ao grupo de controlo.**

	<b>P1</b>	<b>P2</b>	<b>P3</b>	<b>P4</b>	<b>P5</b>
<b>Jitta (us)</b>				↑	
<b>Jitt (%)</b>			↑	↑	↑
<b>rap (%)</b>			↑	↑	
<b>ppq5 (%)</b>			↑	↑	↑
<b>Shim (%)</b>			↑		
<b>Shdb (dB)</b>			↑		
<b>apq3 (%)</b>					
<b>apq5 (%)</b>					
<b>HNR (dB)</b>					

Foi também feita uma comparação entre sinais do grupo de controlo e sinais com os grupos das patologias para cada um dos parâmetros agrupados por vogal e tom como ilustrado na Figura 29 para o exemplo do tom baixo para o parâmetro Jitta. Nesta análise não se repetem todos os diagramas de caixas devido ao seu elevado número. Apresenta-se apenas o exemplo da Fig. 29, resumindo-se os resultados das tabelas a baixo.

Nestes casos, o número de medidas usado em cada caixa corresponde aos 22 elementos do grupo num total de 22 medidas.

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

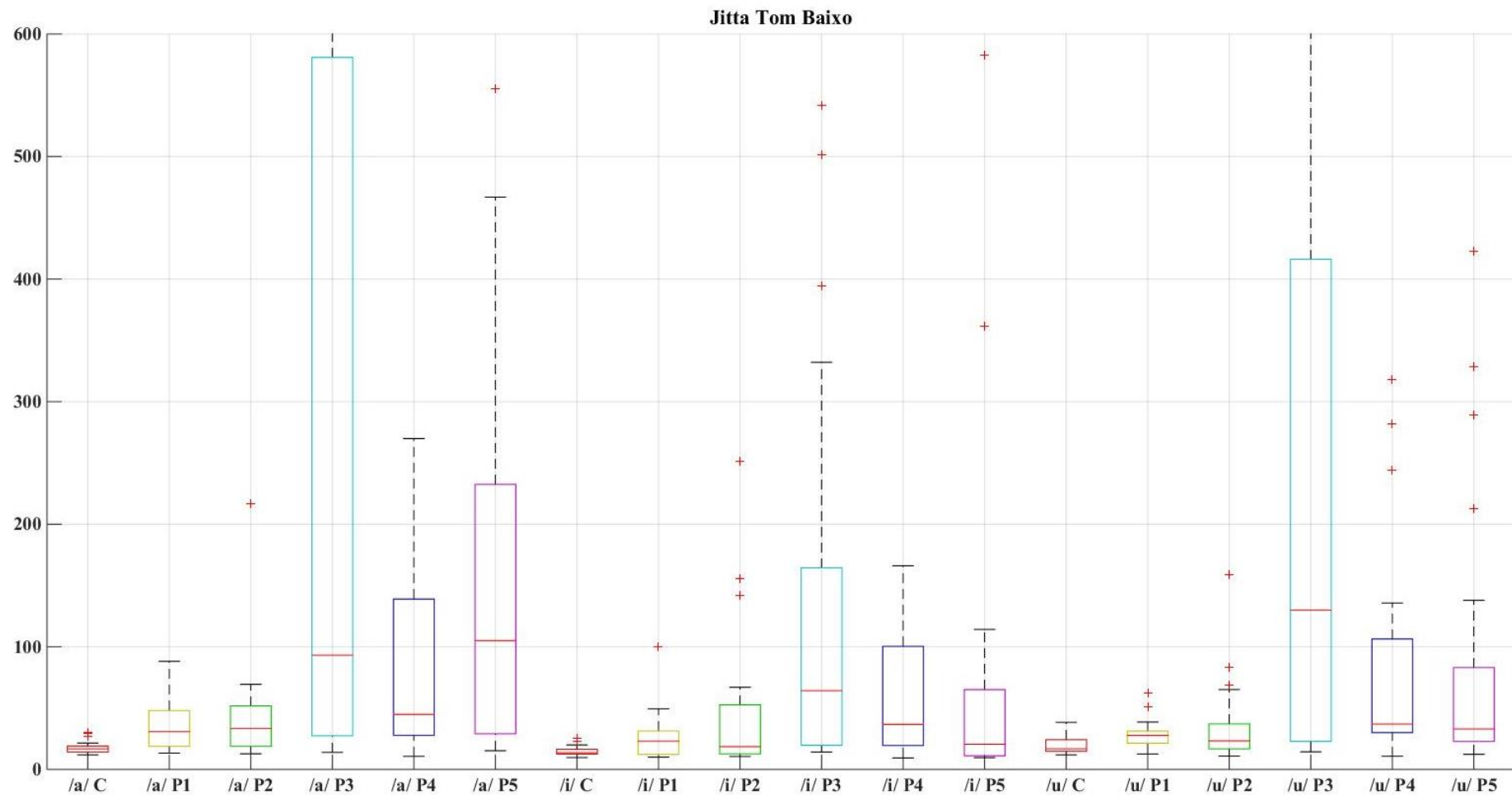


Figura 29 – Comparação entre sinais do grupo controlo com sinais de grupos patológicos de Jitta para o tom baixo.

## V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

As diferenças significativas registadas nos vários parâmetros relativamente ao grupo de controlo foram resumidas nas Tabelas 37, 38 e 39. Analisando as tabelas, para a patologia P1 não existe nenhum parâmetro que permita distinguir esta patologia do grupo de controlo. A patologia P2 só apresenta dados distintos para o parâmetro Jitt na vogal /i/ com tom baixo, dando a percepção que estes parâmetros não são os mais indicados para fazer a identificação destas patologias, ou devem ser combinados de forma inteligente uma vez que apesar de não apresentarem valores estatisticamente diferentes apresentam valores tendencialmente diferentes, como já acontecia na análise anterior.

Quanto as patologias P3, P4 e P5 confirma-se uma distinção estatisticamente significativa relativamente ao grupo de controlo para vários parâmetros, sendo essa diferença mais evidente para a vogal /a/, seguida da vogal /i/. Quanto ao tom, o tom alto é o que apresenta maior distinção seguido do tom baixo.

V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

Tabela 37 – Variação dos parâmetros dos sinais patológicos comparativamente ao grupo de controlo para a vogal /a/.

	P1	P2	P3	P4	P5	
<b>/a/ Baixo</b>	Jitta (us)		↑	↑	↑	
	Jitt (%)		↑	↑	↑	
	rap (%)		↑	↑	↑	
	ppq5 (%)		↑	↑	↑	
	Shim (%)		↑	↑	↑	
	Shdb (dB)		↑	↑	↑	
	apq3 (%)		↑	↑		
	apq5 (%)		↑	↑		
	HNR (dB)				↓	↓
	<b>/a/ Normal</b>	Jitta (us)			↑	↑
Jitt (%)				↑	↑	
rap (%)					↑	
ppq5 (%)				↑	↑	
Shim (%)				↑	↑	
Shdb (dB)				↑	↑	
apq3 (%)					↑	
apq5 (%)					↑	↑
HNR (dB)				↓	↓	↓
<b>/a/ Alto</b>		Jitta (us)		↑	↑	↑
	Jitt (%)		↑		↑	
	rap (%)				↑	
	ppq5 (%)				↑	
	Shim (%)		↑	↑	↑	
	Shdb (dB)		↑	↑	↑	
	apq3 (%)		↑		↑	
	apq5 (%)		↑	↑	↑	
	HNR (dB)			↓	↓	↓

V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

Tabela 38 - Variação dos parâmetros dos sinais patológicos comparativamente ao controlo para a vogal /i/.

	P1	P2	P3	P4	P5
<b>/i/ Baixo</b>	Jitta (us)			↑	
	Jitt (%)		↑	↑	
	rap (%)				
	ppq5 (%)				↑
	Shim (%)				↑
	Shdb (dB)				↑
	apq3 (%)				
	apq5 (%)				
	HNR (dB)			↓	
	<b>/i/ Normal</b>	Jitta (us)		↑	↑
Jitt (%)			↑	↑	
rap (%)					↑
ppq5 (%)			↑	↑	
Shim (%)					
Shdb (dB)					
apq3 (%)					
apq5 (%)					
HNR (dB)					
<b>/i/ Alto</b>		Jitta (us)		↑	↑
	Jitt (%)		↑	↑	
	rap (%)		↑		
	ppq5 (%)		↑	↑	
	Shim (%)		↑	↑	
	Shdb (dB)		↑	↑	
	apq3 (%)		↑		
	apq5 (%)		↑	↑	
	HNR (dB)		↓		

V – Comparação dos Sinais do Grupo de Controlo com Sinais dos Grupos Patológicos

Tabela 39 - Variação dos parâmetros dos sinais patológicos comparativamente ao controlo para a vogal /u/.

	P1	P2	P3	P4	P5	
<b>/u/ Baixo</b>	Jitta (us)				↑	
	Jitt (%)				↑	
	rap (%)					
	ppq5 (%)			↑	↑	↑
	Shim (%)					
	Shdb (dB)					
	apq3 (%)					
	apq5 (%)					
	HNR (dB)					↓
	<b>/u/ Normal</b>	Jitta (us)			↑	
Jitt (%)						
rap (%)						
ppq5 (%)						
Shim (%)						
Shdb (dB)						
apq3 (%)						
apq5 (%)						
HNR (dB)						
<b>/u/ Alto</b>		Jitta (us)			↑	
	Jitt (%)			↑		
	rap (%)			↑		
	ppq5 (%)			↑	↑	
	Shim (%)			↑		
	Shdb (dB)			↑	↑	
	apq3 (%)			↑	↑	
	apq5 (%)			↑	↑	
	HNR (dB)					

# CAPÍTULO VI

---

## 6. CONCLUSÃO E TRABALHOS FUTUROS

### 6.1. CONCLUSÃO

Na realização desta tese foi imprescindível o estudo e apreensão de um conjunto de conceitos novos relacionados com o sinal de fala. A parte inicial teve como objetivo a pesquisa de parâmetros que estejam relacionados com o processo de produção de fala. Foi necessário fazer uma pesquisa a nível anatómico da laringe, assim como a procura dos critérios envolvidos diretamente no processo de produção de fala, sendo estes o Jitter, Shimmer e HNR.

Foi utilizada a base de dados SVD, pois esta apresenta sinais de voz de vogais sustentadas que é condição essencial para o tipo de análise pretendida neste trabalho e por ser de acesso gratuito.

Foi desenvolvido um algoritmo para determinação dos diversos parâmetros do Jitter e do Shimmer que utiliza o método da autocorrelação para determinar a  $F_0$  com o intuito de definir o período glotal associado a cada sinal de voz. Com base no período glotal determinam-se os impulsos glotais positivos dos sinais de voz. Foi necessário proceder a uma avaliação dos picos negativos, quando o sinal apresentava mais do que um pico de amplitude semelhante por ciclo na parte positiva.

Tornou-se o algoritmo mais robusto usando a média deslizante para os sinais com dois ou mais picos positivos e negativos em simultâneo. Para estes casos os picos foram identificados no sinal resultante da média deslizante.

Implementou-se um método de decisão de quais os sinais para que se usaria a média deslizante, com base no número de picos num período glotal.

O sistema de determinação do valor e posição dos impulsos glotais é feito através de uma janela de procura definida entre 1/3 do PG antes e depois de um ponto de referência. Sendo esse ponto de referência localizado 1 PG à frente do último impulso glotal identificado. O procedimento é repetido para os sucessivos impulsos glotais. O primeiro impulso é identificado como sendo o pico de maior amplitude (positiva ou negativa) num intervalo de um período glotal.

## VI – Conclusão e Trabalhos Futuros

Os diversos parâmetros do Jitter e do Shimmer são depois medidos com base nos instantes e amplitudes dos períodos glotais identificados.

Após sucessivas melhorias para tornar o algoritmo mais robusto passou a ser possível avaliar corretamente qualquer tipo de voz.

Posteriormente foram implementados alguns testes ao algoritmo usando sinais de voz sintetizados. Nestes foram utilizados dois sinais simulando dois tipos de perturbação para o parâmetro Jitter e Shimmer sendo os resultados do algoritmo posteriormente comparados com os resultados obtidos pelo programa Praat e com os valores obtidos analiticamente. Neste teste o algoritmo revelou-se bastante preciso nas medidas do Jitter e do Shimmer. Concretamente o erro máximo do Jitter foi de 5  $\mu$ s, ou 0,05%, e no Shimmer de 0,04 dB, ou 0.46%. Assim, o algoritmo revelou ser mais preciso que o Praat na determinação do Jitter e igual na determinação do Shimmer. Em conclusão pode-se considerar que as medidas de Jitter e de Shimmer efetuadas pelo algoritmo são de elevada precisão.

Posteriormente foi feita uma análise às medidas do algoritmo para os parâmetros Jitter, Shimmer e HNR utilizando sinais de voz reais. Foram utilizados sinais de voz de controlo e sinais de voz com a patologia Laringite nesta avaliação. Nesta análise não são conhecidos os valores reais de Jitter, Shimmer e HNR pelo que apenas se pode fazer uma avaliação comparativa com as medidas efetuados por outro software. As diferenças de valores apresentadas entre o algoritmo e o Praat para ambas as situações foram pouco significativas para os parâmetros Jitter e Shimmer, sendo ligeiramente significativas para o parâmetro HNR. Com a obtenção destes resultados foi confirmada a precisão do algoritmo desenvolvido para sinais de voz reais.

Após esta confirmação procedeu-se à comparação dos parâmetros dos sinais de controlo com os sinais provenientes de cinco patologias. Primeiro foi feita uma avaliação sem distinção de tom ou vogal, na qual se concluiu que as medidas de Jitter são as mais reveladoras dos grupos patológicos. Das cinco patologias avaliadas só três (Disfonia Espasmódica, Pólipos das Cordas Vocais e Envelhecimento das Cordas Vocais) apresentaram diferenças significativas nos valores dos parâmetros apesar de haver sempre tendência para valores mais elevados nos grupos com patologia.

Na avaliação com distinção de vogal e tom, só houve distinção estatisticamente significativa para as mesmas três patologias que na situação anterior, sendo essa evidencia mais evidente para a vogal /a/, seguida da vogal /i/ e sendo o tom alto o que apresenta maior distinção seguido do tom baixo. As restantes patologias (Laringite e

Disfonia Hiperfuncional) evidenciam que estes parâmetros não são os mais indicados para fazer a identificação destas patologias.

O algoritmo aqui desenvolvido e aferido faz a determinação com elevada precisão dos parâmetros de Jitter, Shimmer e HNR que podem ser usados num sistema inteligente que combinando os parâmetros mais relevantes numa ferramenta de inteligência artificial pode ser uma ajuda ao diagnóstica de patologias da laringe. Este sistema terá sempre a vantagem de baixo custo e de ser não invasivo em oposição aos métodos tradicionais.

### **6.2. TRABALHOS FUTUROS**

Durante a elaboração desta tese tentou-se criar um programa, através da implementação de uma rede neuronal, capaz de identificar os sinais de voz saudáveis dos sinais de voz patológicos. Esta experiência não foi ainda bem sucedida devido ao elevado número de parâmetros de entrada da rede neuronal, sem estar ainda corretamente estudada a sua importância estatística para esta identificação. Refira-se que na rede neuronal experimentada a entrada consistiu nas 4 medidas de Jitter, mais as 4 medidas de Shimmer, o HNR, multiplicado pelas 3 vogais e pelos 3 tons, mais a identificação do género (masculino ou feminino). Fica assim como trabalho futuro esta análise estatística para identificar quais os parâmetros mais significativos para o diagnóstico de uma determinada patologia.

Como trabalhos futuros, e tendo em conta esta experiência, seria interessante aumentar o número de parâmetros a avaliar. Com um número alargado de parâmetros, e após avaliação destes no que diz respeito à comparação de sinais de voz normais e patológicos seria interessante proceder a uma nova implementação de uma rede neuronal com o objetivo de se criar uma ferramenta de ajuda ao diagnóstico de vozes patológicas ou se possível de diagnóstico mesmo das patologias.



# BIBLIOGRAFIA

---

**Almeida, N. C. (2010).** Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor de Suporte. Dissertação de Mestrado. Universidade Federal do Rio Grande do Norte, Natal, Brasil.

**Arias-Londoño, J. D., & Juan I. Godino Llorente. (2011).** Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients. *IEEE Transactions on Biomedical Engineering*, V. 58(2), p. 370.

**Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., & Berke, G. S. (1996).** Comparison of Voice Analysis Systems for Perturbation Measurement. *Jornal of Speech and Hearing Research*, p. 126–134.

**Boersma, P. (1993).** Accurate Short-term Analysis of the Fundamental Frequency and the Harmonics-to-noise Ratio of a Sampled Sound. *Institute of Phonetic Sciences, University of Amsterdam, Proceedings 17*, p. 97–110.

**Boersma, P., & Heuven, V. van. (2001).** Speak and unSpeak with Praat. *Glott International*, V. 5(No. 9/10), p. 341–347.

**Boersma P, Weenink D.** Praat: doing phonetics by computer. *Phonetic Sciences, University of Amsterdam*. Obtido a 20 de Junho de 2015 <http://www.fon.hum.uva.nl/praat/>

**Coelho, M. S., Macedo, E., Oliveira, M. S. B. de, Lobo, P., Soccol, A. T., & Koerner, H. N. (2010).** Disfonia Espasmódica: Descrição da Doença e dos Distúrbios Neurológicos Associados. *Arq. Int. Otorrinolaringol. / Intl. Aech. Otorhinolaryngol*, p. 163–166. São Paulo Brasil.

**Cunha, C., & Cintra, L. (2010).** Nova gramática do português contemporâneo. Lisboa: Edições João Sá da Costa.

**Dajer, M. E. (2010).** Análise de sinais de voz por padrões visuais de dinâmica vocal. Tese de Doutorado. Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, Brasil.

**Drake, R. L., & Vogl, W., & Mitchell, A. W. (2005).** Gray's Anatomia para Estudantes, Rio de Janeiro: Elsevier Editora Ltda.

**Fernandes, A. P. (2012).** Sintetizador de Fala Didático – Módulo Acústico Modelo de Formantes. Dissertação de Mestrado. Instituto Politécnico de Bragança, Bragança, Portugal.

**Freitas, S. A. (2010).** Correlação entre a avaliação acústica e perceptual na caracterização de vozes patológicas. Relatório do Estado da Arte. Faculdade de Engenharia da Universidade do Porto.

**Freitas, S. A. (2012).** Avaliação Acústica e Áudio Perceptiva na Caracterização da Voz Humana. Tese de Doutorado Faculdade de Engenharia da Universidade do Porto.

**Godino-Llorente, J. I., & Gómez-Vilda, P. (2006).** 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), p. 1943-1953.

**Guimarães, I. (2007).** A ciência e a arte da voz humana. Alcoitão: Edição da ESSA - Escola Superior de Saúde do Alcoitão.

**Lopes, J. M. (2008).** Ambiente de análise robusta dos principais parâmetros qualitativos da voz. Dissertação de Mestrado. Faculdade de Engenharia da Universidade do Porto.

**Miranda, S. V., Mello, R. J., & Silva, H. J. (2011).** Correlação entre o Envelhecimento e as Dimensões das Pregas Vocais. CEFAC. p. 444-451.

**Oliveira, C. M., & Lopes, C. N. (2012).** Determinação de Parâmetros para Diagnóstico Automático de Patologias da Laringe. Projeto de fim de curso. Instituto Politécnico de Bragança.

**Parraga, A. (2002).** Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.

**Pützer, M., & Barry, W. J.** Saarbruecken Voice Database - Handbook. Obtido a 20 de Junho de 2015, de [http://www.stimmdatenbank.coli.uni-saarland.de/help\\_en.php4](http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4)

**Robbins, S. L., & Cotran, R. S. (2005).** Robbins e Cotran/Patologia – Bases Patológicas das Doenças. Rio de Janeiro, Brazil: Elsevier Editora Ltda.

**Silva, M. A. (2010).** Uma Contribuição para a Caracterização do Sinal de Voz Envelhecida. Dissertação de Mestrado. Universidade Federal Fluminense, Brasil.

**Teixeira, J. P. (1995).** Modelação Paramétrica de sinais para aplicação em sistemas de conversão Texto-Fala. Dissertação de Mestrado. Faculdade de Engenharia da Universidade do Porto.

**Teixeira, J. P. (2013).** Slides Teóricos de Tecnologias de Reabilitação – Aplicações de Processamento de Sinal. Escola Superior de Tecnologia e Gestão de Bragança.

**Teixeira, J. P., & Gonçalves, A. (2014).** Accuracy of Jitter and Shimmer Measurements. Procedia Technology, V. 16, 1190–1199.

**Teixeira, J. P., Ferreira, D. B., & Carneiro, S. M. (2011).** Análise Acústica Vocal - Determinação do Jitter e Shimmer para Diagnóstico de Patologias da Fala. Atas do VI Congresso Luso-Moçambicano de Engenharia. Maputo, Moçambique.

**Teixeira, J. P. Fernandes, A. (2013).** Didactic Speech Synthesizer – Acoustic Module – Formants Model. Proceedings of 6<sup>th</sup> International Joint Conference on Biomedical Engineering Systems and Technologies. p. 356-359.

**Klatt, D. H. (1987).** Review of text-to-speech conversion for English . Journal of Acoustical Society of America, 82(3). p. 737-783.

**Wertzner, H. F., Screiber, S., & Amaro, L. (2005).** Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders. Brazilian Journal of Otorhinolaryngology, p. 582–588. Fortaleza, Brasil.

**Zwetsch, I. C., Fagundes, R. D. R., & Scolari, D. (2006).** Processamento digital de sinais no diagnóstico diferencial de doenças laríngeas benignas. Scientia Medica, V. 16(n.3), p. 109–114.