INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

Determinação da Autocorrelação, HNR e NHR para Análise Acústica Vocal

Joana Filipa Teixeira Fernandes

Dissertação apresentada à

Escola Superior de Tecnologia e Gestão

Instituto Politécnico de Bragança

para obtenção do grau de Mestre em

Engenharia Industrial

Ramo de Engenharia Eletrotécnica

Este trabalho foi efetuado sob orientação de:

Professor Doutor João Paulo Teixeira

2018

AGRADECIMENTOS

Ao concluir esta tese, resta-me agradecer a todas as pessoas que contribuíram para que esta fosse possível.

Ao meu orientador, Professor Doutor João Paulo Teixeira, por todo o apoio prestado, pela disponibilidade que sempre demonstrou, pela paciência e conhecimentos transmitidos, fundamentais no desenvolver deste trabalho e enriquecimento pessoal.

A todos os meus amigos pelo incentivo para não desistir, por perceberem a minha falta de tempo/paciência e estarem sempre presentes.

Aos meus pais, irmã, tios e primas, por todo o apoio dado ao longo do meu percurso académico, nunca deixando de acreditar em mim, dando sempre motivação, força e mostrando que era possível chegar onde cheguei e ser quem sou hoje.

A todos, o meu Muito Obrigada.

RESUMO

Este trabalho teve como objetivo a determinação dos parâmetros: *Harmonic to Noise Ration* (HNR), *Noise to Harmonic Ratio* (NHR) e Autocorrelação. Estes parâmetros são usados como entradas de um sistema inteligente para diagnóstico de patologias da fala.

Foi realizada uma análise comparativa entre os valores do algoritmo e do *software* Praat, de modo a perceber qual a melhor janela e o seu comprimento, em número de períodos glotais. Desta análise resultou a decisão de se usar a janela de *hanning* com um comprimento correspondente a 6 períodos glotais. Através da comparação dos resultados chegou-se à conclusão que este algoritmo permite extrair os parâmetros HNR, NHR e Autocorrelação com valores suficientemente próximos dos valores de referência.

Foi ainda desenvolvido um algoritmo para selecionar apenas a parte do sinal onde ocorre fala, eliminando as zonas de silêncio iniciais e finais, para, posteriormente, se extrair os *Mel Frequency Cepstral Coefficientes* (MFCCs), os *Linear Prediction Coefficientes* (LPC) e os *Line Spectral Frequency* (LSF).

Ao longo do trabalho foi possível, embora não fosse o objetivo primordial, complementar uma base de dados curada, iniciada numa investigação anteriormente realizada, adicionando mais parâmetros e mais doenças. Esta base de dados ficou agora com os parâmetros MFCC com 13 coeficientes cepstrais, HNR, NHR, Autocorrelação, *jitter* absoluto, *jitter* relativo, *shimmer* absoluto, *shimmer* relativo, extraídos de 9 locuções correspondentes a 3 vogais em 3 tons e a uma frase, para sujeitos com 19 patologias, mais os sujeitos de controlo. Esta base de dados curada disponibiliza um conjunto de parâmetros sobre estes sinais de fala para a investigação sobre estas 19 patologias.

Palavras-Chave: HNR, *jitter, shimmer*, NHR, autocorrelação, MFCCs.

ABSTRACT

The objective of this work was to determine the parameters: *Harmonic to Noise Ration* (HNR), *Noise to Harmonic Ratio* (NHR) and autocorrelation. These parameters are used as inputs to an intelligent system for diagnosis of speech pathologies.

A comparative analysis was performed between the values of the algorithm and the Praat software, in order to understand the best window and its length, in number of glottal periods. This analysis resulted in the decision to use the hanning window with a length corresponding to 6 glottal periods. By comparing the results it was concluded that this algorithm allows to extract the parameters HNR, NHR and Autocorrelation with values close enough to the reference values.

An algorithm was developed to select only the part of the signal where speech occurs, eliminating the initial and final silence zones, to later extract the *Mel Frequency Cepstral Coefficientes* (MFCCs), *Linear Prediction Coefficientes* (LPC) and *Line Spectral Frequency* (LSF).

Throughout the work it was possible, although it was not the primary objective, to complement a cured database, initiated in a previous investigation, adding more parameters and more diseases. This database now has MFCC parameters with 13 ceptral coefficients, HNR, NHR, Autocorrelation, absolute jitter, relative jitter, absolute shimmer, relative shimmer, extracted from 9 phrases corresponding to 3 vowels in 3 tones and to one sentence, for subjects with 19 pathologies, plus control subjects. This cured database provides a set of parameters on these speech signals for research on these 19 pathologies.

Keywords: HNR, jitter, shimmer, NHR, autocorrelation, MFCCs.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO

Para que seja possível a produção da voz, é necessário a intervenção de vários sistemas. Entre estes sistemas está o respiratório, que é a fonte de energia, responsável pelo ar expelido pelos pulmões. O sistema fonador (ver Figura 1) é representado pelas pregas vocais, que são a fonte de vibrações; o sistema de ressonância, onde está incluída a cavidade oral e nasal; o sistema articulatório, que inclui a língua, os lábios, a mandíbula, o palato e os dentes, e o sistema nervoso central e periférico, como o córtex, permitindo a coordenação (Alves, 2016).

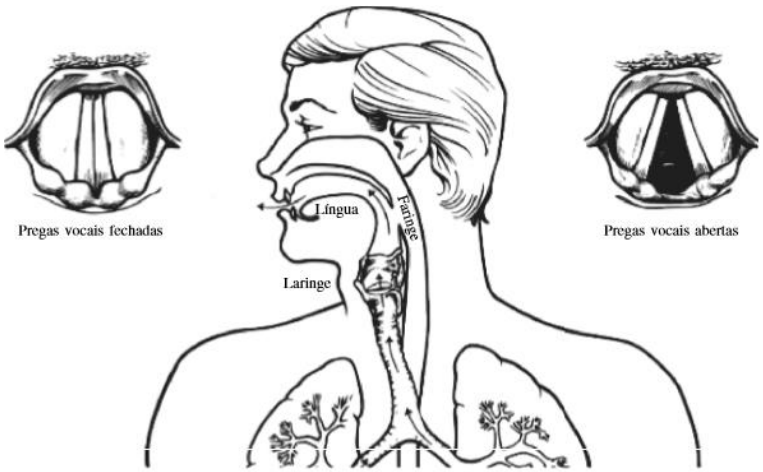

Figura 1 - Trato vocal e aparelho fonador (Guimarães, 2004)

O correto vozeamento dos fonemas é realizado através da vibração das cordas vocais. Estas encontram-se no interior da laringe e são constituídas por um tecido musculoso com duas pregas. Através da passagem do ar faz-se vibrar as duas pregas de fibras elásticas e pode-se modificar a sua forma e elasticidade através dos músculos da laringe, assim sendo, é possível produzir diferentes sons, consoante se quer, por exemplo, alterar o tom de voz ou cantar. O cérebro controla estes músculos enviando mensagens através dos nervos para controlar a aproximação e a tensão das pregas vocais, de modo a que estas

vibrem quando o diafragma e os músculos empurram o ar para fora dos pulmões (Cordeiro, 2016).

A voz humana tem, como qualquer outro som, qualidades próprias (Matuck, 2005):

- Tom – é a altura musical da voz. Segundo o tom, a voz humana classifica-se em aguda ou grave;
- Timbre – é a matriz pessoal da voz. É um fenómeno complexo e é determinado pela frequência fundamental (F0) e pelos seus harmónicos ou secundários. Reconhece-se a pessoa com quem se fala através do seu timbre característico. Existem vozes bem timbradas e agradáveis, assim como, roucas, agudas e chiadas;
- Quantidade – é a duração do som. Segundo a quantidade, os sons podem ser longos ou curtos, com toda a gama intermediária de semi-longos, semi-curtos, etc.. A quantidade depende, geralmente, das características de cada língua, dos costumes linguísticos das regiões ou países;
- Intensidade – é a maior força ou menor força com que se produz a voz. Há vozes fracas e vozes fortes.

Uma perturbação na voz traz implicações profundas na vida social e profissional de uma pessoa. Nos pacientes com patologias progressivas é importante ter acesso a um rápido diagnóstico para ser possível promover um melhor tratamento e prognóstico (Alves, 2016).

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

Existem vários exames que podem ser feitos na deteção de patologias associadas à voz, porém, estes são invasivos e tornam-se um pouco desconfortáveis para os pacientes, podendo até provocar o vómito, ou dependem da experiência do médico que faz a avaliação (exame auditivo) (Teixeira & Fernandes, 2015) (Teixeira, et al., 2011).

A análise acústica vocal pretende quantificar e caracterizar um sinal sonoro. Utilizando esta técnica, na análise da voz, é possível uma análise acústica que permite, de forma não invasiva, determinar a qualidade vocal do indivíduo, através dos diferentes parâmetros

acústicos que compõem o sinal – periodicidade, amplitude, duração e composição espectral (Teixeira, et al., 2011). Esta técnica é bastante utilizada na deteção de patologias da voz.

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

Neste trabalho constam os principais parâmetros acústicos usados na deteção de patologias, sendo estes, o *jitter* relativo (*jitt*), *jitter* absoluto (*jitta*), *shimmer* relativo (*shim)*, *shimmer* absoluto (ShdB), *Harmonic-to-Noise Ratio* (HNR), *Noise-to-Harmonic Ratio* (NHR), autocorrelação, *Mel Frequency Cepstral Coefficients* (MFCC), *Linear Prediction Coefficientes* (LPC) e *Line Spectral Frequency* (LSF).

Porém, apenas são realizadas análises para os parâmetros *Harmonic-to-Noise Ratio* (HNR), *Noise-to-Harmonic Ratio* (NHR) e autocorrelação, uma vez que, para os parâmetros *jitter* e *shimmer*, essa análise já foi realizada no trabalho de (Gonçalves, 2015).

A determinação do HNR teve por base um algoritmo anteriormente desenvolvido, porém, os resultados não eram os melhores.

Por este motivo, foi necessário reformular a determinação do HNR. A este algoritmo adicionou-se a determinação da autocorrelação e NHR, por forma a obter-se uma análise correta dos sinais de voz provenientes da base de dados *Saarbrucken Voice Database* (SVD). O objetivo de comprovar a exatidão do algoritmo fez-se através de uma análise comparativa entre os valores obtidos pelo algoritmo e utilizaram-se como referência os valores obtidos pelo *software* Praat para sinais de controlo e patológicos.

O *software* Praat foi o escolhido como termo de comparação, uma vez que existem imensos trabalhos realizados no âmbito da análise de sinais de fala em que é utilizado, é

gratuito e é comumente aceite pela comunidade científica como uma medida precisa. Este *software* foi desenvolvido por (Boersma & Weenink, s.d.) e é usado para análise de fala.

Para a identificação dos MFCCs, LPC e LSF utilizou-se as funções disponíveis na tollbox de processamento de sinal do MatLab (MathWorks, 2011) (MathWorks, 2006) (MathWorks, 2006).

1.2. OBJETIVOS

A investigação no reconhecimento de patologias da voz, onde é identificada qual a patologia das pregas vocais, ainda é escassa, porém, trabalhos para a deteção da existência de uma doença já existem desde há alguns anos. Assim sendo, o reconhecimento das patologias da voz, com recurso a métodos de reconhecimento de fala, é viável, tendo em conta diversos trabalhos referidos no estado da arte desta tese.

Para o reconhecimento de vozes patológicas pode utilizar-se como recurso as medidas que caracterizam o movimento das pregas vocais, tais como o *jitter* e o *shimmer*. Contudo, apesar de as patologias afetarem tipicamente as cordas vocais, só estas medidas não produzem resultados conclusivos no reconhecimento de patologias da voz, daí a necessidade de utilizar mais medidas, tais como parâmetros harmónicos.

O objetivo deste trabalho é desenvolver um algoritmo para determinar o HNR, a autocorrelação e NHR. De forma a validar os resultados obtidos através do algoritmo desenvolvido neste trabalho, estes foram comparados com os resultados de um *software* comummente aceite pela comunidade científica. Pretende-se, ainda, completar a base de dados curada criada por (Fernandes, et al., 2018), porém, a esta base de dados, para além dos parâmetros referidos anteriormente, ainda vão ser adicionados os *Mel Frequency Cepstral Coefficients* (MFCC).

Nesta base de dado curada pensou-se adicionar os parâmetros LPC e LSF, contudo, o número de coeficientes pode variar de investigador para investigador, daí não se terem colocado.

1.3. ESTADO DA ARTE

O trabalho de (Yumoto & Gould, 1982) foi realizado para desenvolver a relação harmónica-ruído, HNR (*harmonics-to-noise ratio*) como uma avaliação objetiva e quantitativa do grau de disfonia. Neste, foram analisadas gravações de 42 vozes normais

e 41 com diferentes graus de disfonia. Sendo que dois especialistas também classificaram o espectrograma de cada amostra de voz, a vogal sustentada /a/, com base na quantificação de ruído em relação à componente harmónica. A relação harmónica-ruído mostrou-se útil na avaliação quantitativa da disfonia, uma vez que, quanto maior for o ruído no espectro do sinal de fala, maior será a ausência de vozeamento.

Em (Murphy, 1999) foram comparados parâmetros glotais como o *jitter* do período fundamental, *shimmer* e HNR, como características do trato vocal, valores e desvios das frequências dos primeiros e segundos formantes. Para esta análise utilizou-se a vogal /a/ produzida por 223 sujeitos de controlo, 472 pacientes diagnosticados com nódulos e 195 com paralisia das cordas vocais. Os resultados obtidos mostram que as características glotais obtêm melhores resultados que as do trato vocal, porém, se estas forem combinadas, existe uma melhoria dos resultados. Assim sendo, foi possível concluir que os formantes contêm informação que permite classificar as vozes patológicas.

Em (Henríquez, et al., 2009) foi proposto quantificar a qualidade de vozes gravadas utilizando medidas não-lineares objetivas. Foram utilizadas seis medidas caóticas não-lineares baseadas em teoria dinâmica não-linear onde é feita a discriminação entre vozes saudáveis e patológicas. Foram analisadas as entropias de *Rényi* de primeira e segunda ordem, a entropia da correlação e a dimensão da correlação. Estas medidas foram obtidas a partir do sinal de fala no domínio espaço-frase, que é criado através da estabilização de vetores $\mathbb{R}^m$, onde os elementos são versões com atraso de tempo da série temporal original. Foram, também, analisados os valores do primeiro mínimo da função de informação mútua e entropia de *Shannon*. Utilizaram-se duas bases de dados para fazer a avaliação: uma com quatro níveis de qualidade de voz (1 nível de voz saudável e 3 de vozes patológicas); e a MEEI *Voice Disorders* (*Massachusetts Eye and Ear Infirmary*) composta por dois níveis (saudáveis e patológicos). Os testes foram feitos através de redes neuronais artificiais (RNA) onde se obtiveram taxas de acerto de 82,47% para a base de dados com quatro níveis de qualidade de voz e 99,69% para a base de dados MEEI.

(Silva, et al., 2009) focou-se na avaliação de diferentes métodos para estimar a quantidade de *jitter* presente nos sinais de fala. O valor de *jitter* é um bom indicador para a deteção de patologias na laringe, como nódulos das cordas vocais ou pólipos. Utilizaram duas bases de dados: a base de dados da MEEI e uma adquirida para o trabalho. Para a base de dados do MEEI os melhores resultados foram obtidos com o *jitter* absoluto.

(Fonseca & Pereira, 2009) desenvolveu um método para analisar as características tempo-frequência (DWT – *Discrete Wavelet Transform*) para distinguir vozes patológicas de pacientes com edema de *Reinke* e nódulos nas cordas vocais, de pacientes com voz normal. Neste trabalho utilizou *wavelets*. Foram utilizados 71 sinais de voz, tanto do sexo masculino como do feminino. Destes, 30 não tinham patologia, 25 pacientes com nódulos nas cordas vocais e 16 pacientes com edema de *Reinke*. Utilizaram um classificador de máquinas de vetor de suporte (LS-SVM), onde obtiveram uma taxa de acerto de 90% na classificação entre vozes normais e vozes patológicas com nódulos nas cordas vocais, mais de 85% entre vozes normais e vozes patológicas com edema de *Reinke* e mais de 80% na distinção entre vozes patológicas com nódulos nas cordas vocais e vozes patológicas com edema de *Reinke.*

(Almeida, 2010) desenvolveu um sistema de classificação de voz para auxiliar no pré-diagnóstico de patologias na laringe, bem como no acompanhamento de tratamentos farmacológicos e pós-cirúrgicos. Os coeficientes utilizados para a extração de características relevantes do sinal de voz são os coeficientes de predição linear (LPC), coeficientes cepstrais de frequência mel (MFCC) e os coeficientes obtidos através da transformada de *wavelet packet* (WPT). Como classificador utilizou máquinas de vetor de suporte (SVM). Foram utilizadas duas bases de dados, porém, não é referido o número de pessoas utilizadas nos testes. Foram obtidas taxas de acerto de 98,46% na classificação entre voz normal e voz patológica e 98,75% na classificação de patologias entre edemas e nódulos.

(Muhammad, et al., 2011) propõe um sistema automático de classificação de voz usando os dois primeiros formantes da vogal /a/ e /i/. Era proposto reconhecer cinco patologias, não existindo uma classe para vozes normais, e, para tal, foram utilizados dois métodos de classificação: quantificação vetorial (VQ) e redes neuronais artificiais (RNA). São utilizados 71 pacientes, dos quais 50 são do sexo feminino e os restantes são do sexo masculino, num total de 710 sinais, onde 80% são usados para treino e 20% para testes. Com as quatro características retiradas do sinal de fala e usando uma rede neuronal artificial são conseguidas taxas de acerto de 67,8% para pacientes masculinos e 52,5% para pacientes femininos. Relativamente ao classificador vectorial, obtiveram uma taxa de acerto de 35% para os pacientes do sexo masculino e 28% para os do sexo feminino.

(Teixeira, et al., 2011) implementou um algoritmo onde pretendia determinar a F_0 de um sinal de voz utilizando o método do Cesptro e através do método da autocorrelação.

Posteriormente foram avaliados os valores obtidos comparando-os com os resultados obtidos pelo *software* Praat. Produziu também um algoritmo onde são identificados os parâmetros *jitter* e *shimmer* utilizando o *software* MatLab.

(Teixeira, et al., 2013) este trabalho teve como base a análise do sinal de fala no processamento de sinal e, como principal objetivo, tornar automático o diagnóstico das patologias da laringe. O autor pretendia analisar o sinal/som correspondente à pronúncia continuada de uma vogal (/a/) e estiveram envolvidas pessoas, estudantes, entre os 20 e 23 anos. Foi selecionado para análise o som de um elemento masculino e outro feminino e não apresentavam qualquer sinal ou sintoma de distúrbios na voz. Para esta gravação teve-se em atenção algumas condições como: tempo e espaço onde foi gravado com determinadas condições acústicas procurando criar um ambiente o melhor possível. Foram analisados os algoritmos e a sua implementação para determinar os parâmetros, associados com o *jitter*, *shimmer* e HNR e as várias medidas.

No trabalho de (Teixeira & Fernandes, 2014) foi utilizada a base de dados *Saarbrucken Voice Database* (SVD) e são utilizados para vozes saudáveis 34 elementos do sexo feminino e 7 do sexo masculino. Como parâmetros de avaliação foi utilizado o *Jitta*, *Jitt*, rap, ppq5, S*him*, ShdB, apq3, apq5 e HNR e são analisadas três vogais, /a/, /i/ e /u/ nos tons alto, baixo e normal. É apresentada uma análise estatística de vozes saudáveis, onde apresenta a média e desvio padrão para cada parâmetro por gênero. Nesta análise apenas o *Jitta* apresenta diferenças estatisticamente evidentes entre o sexo masculino e o feminino. Foi também feita outra análise. Nesta, inicialmente, agruparam os parâmetros correspondentes a cada um dos três tons e depois os parâmetros por vogais e verificaram que cada tom possui diferenças estatisticamente significativas em relação ao outro.

O trabalho de (Teixeira & Gonçalves, 2014) teve como objetivo analisar as medidas de *jitter* e *shimmer* produzidas por um algoritmo desenvolvido. O algoritmo baseia-se no uso da média móvel do sinal de fala e encontrar os picos que serão o centro para encontrar a amplitude máxima das ondas de fala. Procuraram verificar a fiabilidade que tem o algoritmo desenvolvido e que já foi melhorado. Para tal, produziram um sinal sintetizado com os valores controlados, de seguida, determinaram os parâmetros de *jitter* e *shimmer* usando o sistema desenvolvido e o *software* Praat e, posteriormente, comparados com os valores determinados analiticamente. Foram realizadas várias experiências com diferentes tipos de perturbações de *jitter* e *shimmer* e com diferentes valores de F_0, assim

como a influência que as variações de F0 tem nas medidas restantes do *jitter* e do *shimmer*.

(Gonçalves, 2015) produziu um algoritmo capaz de medir os parâmetros da voz, *jitter*, *shimmer* e HNR, em vozes patológicas para posterior diagnóstico de patologias da fala. O algoritmo desenvolvido teve por base o trabalho apresentado em (Teixeira & Gonçalves, 2014) tornando-o mais robusto aplicando também em vozes patológicas. A medição do *jitter* foi medida em quatro parâmetros: *Jitt*, *Jitta*, Rap e Ppq5; e a do *Shimmer* em: *Shim*, ShdB, Apq3 e Apq5. Para comparar os valores obtidos pelo algoritmo utilizaram-se os obtidos pelo *software* Praat. Para testar o algoritmo utilizou sinais sintetizados, com valores bem conhecidos para os parâmetros, sinais de voz normal (controlo) e sinais de voz patológicos retirados da base de dados *Saarbrucken Voice Database* (SVD). Na comparação realizada, utilizando o sinal sintetizado, o algoritmo produziu um erro inferior a 5 µs 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 insignificantes. Realizou, também, uma comparação estatística do comportamento dos parâmetros *jitter* e *shimmer* em sinais de voz de controlo e sinais de pacientes com laringite, disfonia hiperfuncional, disfonia espasmódica, pólipos das cordas vocais e envelhecimento das cordas vocais, apesar de só as três últimas apresentarem distinção estatisticamente significativa dos parâmetros em relação ao grupo de sinais de voz de controlo.

(Panek, et al., 2015) utilizou um vetor de 28 parâmetros acústicos que avaliou utilizando a análise de componentes principais (PCA), a análise da componente principal de *Kernel* (kPCA) e uma rede neuronal auto associativa (NLPCA). Estas avaliações foram utilizadas na deteção de quatro tipos de patologias (disfonia hipertónica, disfonia funcional, laringite e paralisia das cordas vocais), utilizando as vogais /a/,/i/ e /u/ em três tons, alto, baixo e normal. De acordo com os resultados os métodos kPCA e NLPCA podem ser considerados para a deteção da patologia das cordas vocais. Através dos resultados observou-se que com esta abordagem conseguiram-se níveis de eficiência de cerca de 100%. Sendo que, a classificação entre saudável e patológico foi feita separadamente para cada doença e gênero e teve-se em conta o número de pessoas de controlo ser igual ao número de pacientes para cada patologia, porém, não é referido o número de pessoas que foram utilizadas.

(Teixeira & Fernandes, 2015) fizeram uma análise acústica da disfonia vocal e utilizaram como parâmetros acústicos o *jitter*, o *shimmer* e o HNR. Nesta análise utilizaram a base de dados *Saarbrucken Voice Database* (SVD) e observaram um grupo de controlo e 4 grupos de pessoas com patologia: disfonia, disfonia funcional, disfonia hiperfuncional e disfonia psicogénica. Fizeram uma análise estatística dos três parâmetros utilizando 3 vogais /a/, /i/ e /u/ e cada vogal em 3 tons, alto, baixo e normal, para os 4 grupos patológicos e grupo de controlo. O grupo de controlo é composto por 34 pacientes do sexo feminino e 7 do sexo masculino. Para a disfonia funcional foram utilizados 22 sujeitos do sexo feminino e 7 do masculino, na disfonia hiperfuncional foram utilizados 23 pacientes do sexo feminino e 6 do masculino, para a disfonia psicogénica utilizou 21 mulheres e 8 indivíduos do sexo masculino e para a disfonia utilizou 15 mulheres e 11 homens. Apenas foram apresentados os resultados *jitter* relativo e parâmetros *shimmer*, apesar de as conclusões serem as mesmas para o *jitter* remanescente e parâmetros *shimmer*. Verifica-se que o parâmetro HNR não mostra uma forte capacidade discriminante, enquanto que, *jitter* e *shimmer* são relevantes para serem utilizados num sistema de diagnóstico de patologias da disfonia.

(Baravieira, 2016) pretendia validar um sistema automático onde utiliza redes neuronais artificiais para avaliar vozes rugosas e soprosas. A base de dados utilizada é da clínica de Fanoaudiologia da Faculdade de Odontologia de Bauru (FOB/USP). Desta base de dados foram utilizadas um total de 123 vozes com e sem patologia e a vogal sustentada /a/. Neste trabalho foi feita a extração das características do sinal de voz através da transformada *wavelet packet* e dos parâmetros acústicos: *jitter*, *shimmer*, frequência fundamental, amplitude da derivada e amplitude do *pitch*. Para a rede neuronal artificial, na identificação da soprosidade e da rugosidade e dos seus graus, obteve-se o melhor resultado para a soprosidade no subconjunto composto pelo *jitter*, amplitude do *pitch* e frequência fundamental, conseguindo obter uma taxa de acerto de 74%, relativamente à rugosidade, o melhor subconjunto foi o composto pela transformada *wavelet packet* com 1 nível de decomposição, *jitter*, *shimmer*, amplitude do *pitch* e frequência fundamental, obtendo uma taxa de acerto de 73%.

(Forero, et al., 2016) utilizou parâmetros do sinal glotal, uma vez que são mais propensos na identificação da voz saudável ou de distúrbios da voz: nódulos nas cordas vocais e paralisia unilateral das cordas vocais. Através do algoritmo de filtragem inversa adaptativa iterativa sincronizada de afinação (PSIAIF) conseguimos eliminar a influência

do trato vocal e a radiação da voz causada pela boca, pré-servindo as características do sinal glotal. Este método foi escolhido devido ao seu alto desempenho. Foi utilizado o MatLab onde de implementou uma *toolbox* – Aparat, que foi desenvolvida com base neste método permitindo obter o sinal glotal e extrair as principais características. Foram utilizados parâmetros que podem ser divididos em três grupos: domínio do tempo, domínio das frequências e os que representam as vibrações da frequência fundamental. A base de dados utilizada foi cedida por um terapeuta e contém 248 gravações de voz, em que cada paciente tem 8 gravações, e existem 12 pessoas com nódulos, 8 com paralisia das cordas vocais e 11 saudáveis. Os métodos de classificação utilizados foram: redes neuronais artificiais, máquinas de vetor de suporte e cadeias de *Markov* escondidas, onde se obtiveram taxas de acerto de 95,8%, 82% e 96,2% respetivamente.

(Alves, 2016) começou por utilizar um primeiro conjunto de parâmetros constituídos por HNR e quatro medidas de *jitter* e *shimmer*. Foi avaliada a capacidade de predição deste conjunto de parâmetros quando usados com apenas uma vogal e um tom e quando usados com várias vogais e vários tons. Foram também analisados um segundo conjunto de parâmetros: 12 coeficientes cepstrais, frequência e largura de banda dos três primeiros formantes, frequência fundamental, energia, potência, momentos espectrais de ordem zero, um, dois, três e curtose. Sendo que, estes parâmetros serviram para aferir utilidade de outro tipo de parâmetros na deteção de patologias da laringe. Foram aplicadas técnicas de seleção de variáveis e redução da dimensão como a regressão linear passo a passo e análise dos componentes principais (PCA). Para a classificação entre saudável e patológico utilizou dois tipos de sistemas inteligentes: redes neuronais artificiais (RNA) e máquinas de vetor de suporte (SVM). Neste trabalho foi utilizada a base de dados *Saarbrucken Voice Database* (SVD) e separou o género feminino do masculino. O número de pacientes saudáveis selecionados foi o mesmo que o grupo patológico. Esta análise teve por base, como patologia, a paralisia das cordas vocais e a disfonia. Utilizou um total de 334 pessoas do sexo feminino e 196 do sexo masculino. Dos 334 pacientes do sexo feminino, 126 tinham paralisia das cordas vocais e 41 disfonia e no sexo masculino 69 tinham paralisia das cordas vocais e 29 disfonia. Foram obtidas precisões de 100% para o primeiro conjunto de parâmetros, usando a disfonia feminina e a masculina como grupo patológico; 78,9% usando a paralisia das cordas vocais feminina como grupo patológico; 81,8% usando a paralisia masculina como grupo patológico.

(Cordeiro, 2016) propõe soluções que permitam identificar patologias da voz através do processamento de fala. Para os métodos utilizados envolveu classificadores inteligentes geralmente usados em reconhecimento de fala, tais como, máquinas de vetor de suporte (SVM) e modelo de misturas Gaussianas. Com os parâmetros utilizados nos classificadores pretendeu modelar o trato vocal, como por exemplo os *mel-frequency cepstral coeficientes*, os *line spectral frequencies* e *mel-line spectral frequencies*. Foram usadas duas bases de dados: a base de dados da Universidade de São Paulo e a MEII (*Massachusetts Eye and Ear Infirmary*). A base de dados da Universidade de São Paulo é constituída por 47 pessoas, divididas em 3 grupos: 16 pessoas saudáveis, 16 pessoas diagnosticadas com edema de *Reinke* e 15 diagnosticadas com nódulos. Da base de dados MEEI fazem parte 53 sujeitos saudáveis e 724 com patologia da voz. Propôs, também, o uso de fala contínua como sinal para a identificação de patologias. Nesta abordagem realizou testes onde usou três grupos: sujeitos saudáveis, sujeitos com patologia laríngeas fisiológicas (edemas e nódulos); e sujeitos com patologias laríngeas neuromusculares (paralisia unilateral das pregas vocais). Obteve uma taxa de acerto de 84% para os três grupos e, ainda, foi tido em conta outra abordagem, tendo por base a análise dos formantes e a relação harmónica-ruído. Deste modo, efetuou a implementação de um algoritmo simples baseado em árvores de decisão que permitiu uma taxa de reconhecimento de 95%.

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

para usar como ferramenta de decisão inteligente para classificar entre laringite crónica e saudável.

1.4. ORGANIZAÇÃO DOS CAPÍTULOS

Este trabalho está dividido em 7 capítulos.

No **Capítulo 1** é feita uma introdução ao tema do trabalho onde se refere a vantagem da utilização da análise acústica vocal na deteção de patologias da fala. São igualmente enunciados os objetivos, metodologias utilizadas no desenvolvimento deste trabalho e ainda uma revisão da literatura.

O **Capítulo 2** diz respeito a uma breve descrição das patologias da fala utilizadas para completar a base de dados curada com parâmetros acústicos extraídos dos sinais de fala patológica e de controlo, iniciada por (Fernandes, et al., 2018).

No **Capítulo 3** consta a base de dados utilizada e o número de pacientes de cada doença e de controlo e ainda a descrição teórica de todos os parâmetros usados neste trabalho.

No **Capítulo 4** são referidos os algoritmos desenvolvidos para determinação dos parâmetros.

O **Capítulo 5** apresenta os resultados e discussão sobre os testes feitos aos algoritmos desenvolvidos neste trabalho.

No **Capítulo 6** consta a descrição da base de dados curada.

No **Capítulo 7** estão presentes as conclusões retiradas sobre todas as análises feitas ao longo deste trabalho e algumas sugestões de trabalhos futuros.

2. PATOLOGIAS DA FALA

Hoje em dia, cada vez mais, as patologias da fala perturbam a vida das pessoas, daí a necessidade de se ter acesso a um rápido diagnóstico, tendo em vista um tratamento mais eficaz e eficiente.

Existem vários tipos de exames utilizados como diagnósticos de distúrbios fonológicos. Inicialmente a avaliação vocal realizava-se de forma subjetiva através da análise preceptiva auditiva, porém, a falta de consenso entre os examinadores experientes, mesmo com o uso de diferentes escalas de alterações vocais, tornou necessária a pesquisa de uma avaliação objetiva, onde a voz fosse analisada através de aparelhos capazes de medir parâmetros acústicos.

A existência de uma alteração na voz devido a lesões nas cordas vocais altera o processo de fonação, uma vez que, os padrões de vibração durante a fase de abertura e fecho das cordas vocais são irregulares (Teixeira, et al., 2011). Considerando este princípio, neste trabalho são utilizadas técnicas que permitem analisar vozes retiradas de uma base de dados, que foram analisadas, posteriormente, pelo sistema, de forma a obter medidas quantitativas das alterações da voz.

Neste capítulo serão elencadas e analisadas várias doenças, contudo, neste trabalho só três servirão de base para as análises realizadas: laringite crónica, disfonia e paralisia das cordas vocais. Todavia, todas as aqui descritas servirão para a criação da base de dados curada.

2.1. LARINGITE CRÓNICA

A laringite crónica corresponde a uma inflamação persistente da mucosa laríngea, por vezes, com muitos anos de evolução. Esta doença é, geralmente, provocada por infeções agudas repetidas. As pessoas que estão em constante exposição a fatores irritantes, nomeadamente tabaco, álcool, ambientes repletos de fumo, pó e vapores irritantes, abuso ou má utilização da voz entre outros, são geralmente mais afetadas. A inflamação provoca, além de uma tumefação da mucosa laríngea, uma abundante produção de secreções, fatores responsáveis pelas manifestações que dão sob a forma de rouquidão tenaz, permanente, embora variável (Medipédia, 2012).

2.2. DISFONIA

Disfonia é um termo médico que significa desordem (dis-) da voz (-fonia) e é um distúrbio na comunicação, que dificulta a produção vocal, onde ocorre um impedimento da produção da voz. Embora existam muitas causas de disfonia, esta pode ser causada por uma disfunção, uso intensivo ou mau uso da voz, é mais frequente em indivíduos que usam a voz diariamente de forma abundante e incorreta. Esta patologia pode ter como sintomas rouquidão, dor de garganta ou garganta seca. Um cantor ou cantora pode notar que já não é capaz de cantar em tons mais altos. Pode, ainda, ocorrer outros sintomas associados como um gotejamento contínuo na parte de trás da garganta (catarro nasal) e azia (Teixeira & Fernandes, 2015).

Entre saúde vocal, distúrbios da voz (disfonia) e condições de trabalho existe uma relação, pois a disfonia pode manifestar-se através de uma série de mudanças: dificuldade em manter a voz; fadiga vocal; variações na frequência usual; falta de volume e projeção; perda de eficiência vocal e pouca resistência ao falar (Teixeira & Fernandes, 2015).

A disfonia é uma patologia que está relacionada a vários distúrbios e sintomas, manifestando-se como sintoma secundário ou como principal. Pode, também, ser orgânica ou funcional, sendo que, a orgânica deve-se a uma alteração anatómica nas cordas vocais, como nódulos ou tumores benignos; e a funcional ocorre quando não existem alterações anatómicas (Teixeira & Fernandes, 2015).

2.3. PARALISIA DAS CORDAS VOCAIS

A paralisia das cordas vocais é um distúrbio da voz que ocorre quando os músculos laríngeos não conseguem executar a sua função. A paralisia pode ser unilateral – ocorre devido à imobilidade de uma das cordas vocais; ou bilateral – quando as duas cordas vocais paralisam. A paralisia unilateral é a mais comum, enquanto a bilateral é mais rara e pode implicar risco de vida (Teixeira, et al., 2011).

Aquando de uma paralisia, as cordas vocais podem permanecer abertas ficando as vias respiratórias e os pulmões desprotegidos. Esta patologia pode ocorrer após um trauma na cabeça, pescoço ou peito como em pessoas com problemas neurológicos como esclerose múltipla, doença de *Parkinson* ou quem tenha sofrido um acidente vascular cerebral (AVC) (ServiceS, 2011).

Quando a paralisia afeta apenas uma prega vocal, as pregas vocais vibram com frequências diferentes. Nestes casos, a voz apresenta um som bitonal e o paciente não consegue falar alto, perdendo o poder de amplificação vocal. Quando ocorre a paralisia nas duas pregas vocais pode existir o risco da abertura da glote não ser total, provocando com isso dificuldades respiratórias assim como ruído à passagem do ar respirado.

Pode manifestar-se sobre a forma de rouquidão, soprosidade, dificuldades em respirar, respiração ruidosa e problemas de deglutição, podendo, também, ainda ocorrer alterações na qualidade de voz como perda de volume ou frequência fundamental (ServiceS, 2011).

2.4. TUMOR DA LARINGE

O tumor da laringe refere-se a um tumor que tem origem no revestimento desta estrutura (mucosas), sendo as principais causas deste, o fumo do tabaco e o álcool. (Estibeiro & Trindade, s.d.)

Os sintomas para o tumor da laringe são a rouquidão, dificuldades em engolir, dificuldade em respirar, dor (às vezes correndo para o ouvido) e nódulos no pescoço (Estibeiro & Trindade, s.d.).

2.5. CARCINOMA DAS CORDAS VOCAIS

Quando falamos em carcinoma das cordas vocais, estamos a falar de cancro da laringe, maligno, situado nas cordas vocais. Por norma este tumor está associado ao hábito de fumar, ao consumo excessivo de bebidas alcoólicas e também à laringite crónica, provocada por infeções agudas repetidas, má utilização da voz ou inalação persistente de ar contaminado por pó ou vapores irritantes (Silva, 2014).

Os sintomas associados às cordas vocais é a rouquidão.

2.6. QUISTO

Os quistos ou cistos têm uma forma arredondada, com paredes finas cheias de líquido, que pode ser congênito ou adquirido e pode localizar-se na laringe ou na faringe (Valiullina, s.d.).

Existem fatores de risco que podem levar à formação de quistos tais como fumar, bebidas alcoólicas, trabalhar em ambientes perigosos onde se inalam pequenas partículas de substâncias nocivas, má higiene oral e, predisposição hereditária (Valiullina, s.d.).

Como sintomas podemos ter uma forte sensação de corpo estranho na garganta, causando desconforto ao engolir, tosse irritativa, rouquidão, uma sensação de que rasga o pescoço (Valiullina, s.d.).

2.7. DISFONIA ESPASMÓDICA

A disfonia espasmódica é uma doença neurológica que envolve as cordas vocais, onde os músculos se contraem de forma intensa e irregular (espasmos), tornando a voz "tensa" e "fragmentada", condicionando a qualidade e fluência da voz (Cuf, s.d.).

Esta doença pode caracterizar-se pela fonação tensa-estrangulada, com quebras na produção de palavras ou a dificuldade de iniciar a comunicação, e também como a manutenção de qualidade vocal normal seguida por momentos de voz soprosa ou sussurrada (Fabron, et al., 2012).

2.8. DISFONIA FUNCIONAL

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

2.9. DISFONIA HIPERFUNCIONAL

A disfonia hiperfuncional ocorre quando há uma contração involuntária excessiva da musculatura fonatória, como consequência do uso inapropriado da voz. Esta doença tem como sintomas uma voz rouca e forçada, com uma sensação de limpar a garganta e sensação de corpo estranho na faringe (Navarra, 2015).

2.10. DISFONIA HIPOFUNCIONAL

Na disfonia hipofuncional ocorre uma fraqueza na musculatura laríngea, com o fecho incompleto da glote, devido à fraca musculatura generalizada ou laríngea. A voz torna-se

sussurrada e sem timbre. Existe necessidade de limpeza da garganta, uma sensação de um corpo estranho e uma dor cervical (Navarra, 2015).

2.11. DISFONIA HIPOTÓNICA

A disfonia hipotónica deve-se a um defeito no tônus muscular ao nível da musculatura intrínseca da laringe (falta de força das cordas vocais). Por norma ocorre após longos períodos de silêncio, provocando um início de fonação difícil, melhorando à medida que se vai falando. Normalmente ocorrem alterações na postura acompanhadas de esforço localizado na nuca, maxila e ressonantes. Geralmente ocorrem alterações na respiração e dos parâmetros aerodinâmicos (Castellon, s.d.).

2.12. DISFONIA PSICOGÉNICA

A disfonia psicogénica é um distúrbio de natureza psicológica, caracteriza-se essencialmente por alterações da voz sem uma lesão estrutural laríngea ou doença neurológica, sendo uma doença que surge com maior frequência em mulheres (Santiago, s.d.). Na disfonia psicogénica a voz, a articulação e a fluência são sensíveis às oscilações psicológicas, isto leva a que fatores stressantes possam estar relacionados às alterações vocais (Bergamini, et al., 2015).

Esta doença tem como sintomas uma fonação sussurrada, comportamentos bizarros ou atípicos durante a fonação e uma movimentação específica das pregas vocais no momento em que se tenta falar (Santiago, s.d.).

2.13. EDEMA DE REINKE

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

Por norma, esta doença desenvolve-se principalmente em fumadores crónicos, provocando uma voz mais grave e há um maior número de casos detetados no sexo feminino (Cordeiro, 2016) (Martins, et al., 2009).

2.14. FIBROMA

O fibroma localizado na zona da garganta é dos tumores benignos, o mais comum. Pode localizar-se em várias partes da garganta, porém, o local onde mais ocorre é nas cordas vocais. Esta doença surge essencialmente em pessoas que utilizam muito a voz para trabalhar, como por exemplo professores e cantores, e surge como consequência de uma sobrecarga vocal, ou seja, o alongamento excessivo dos ligamentos. Surge também em fumadores ou pessoas que permaneçam durante muito tempo em lugares com poeiras, ar seco ou vapores perigosos (Anon., 2015).

2.15. GRANULOMA

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

A origem dos granulomas tem vários fatores e pode ser funcional, orgânica ou mista. Quando a origem é funcional deve-se a desvios específicos no comportamento vocal, que envolvem pressão e atrito na região posterior da laringe durante a fonação. O início do granuloma ocorre através de uma lesão na mucosa e desenvolve-se através de sucessivos traumas da captação glótica na região posterior da laringe (Virtual, 2010).

Dos fatores de natureza orgânica pode-se distinguir as seguintes situações: granuloma por trauma químico quer seja por refluxo gastresofágico ou inalação de substâncias irritantes, por trauma direto pós-intubação, granuloma por cicatrização de área cirúrgica da laringe. Às vezes é observado no pós-operatório de laringectomias parciais, ou ainda, por traumas

fechados, como os provocados por choques na parte anterior do pescoço em acidentes automobilísticos ou em brigas (Virtual, 2010).

2.16. GRANULOMA DE INTUBAÇÃO

O granuloma de intubação é igual ao descrito anteriormente, porém, este ocorre devido a longos períodos de intubação, uma vez que, a configuração da glote em "V", a cânula orotraqueal fica apoiada nos processos vocais das cartilagens aritenóideas, exercendo determinada pressão sobre a mucosa da região. Os fatores envolvidos no desenvolvimento do granuloma de intubação são: intubações prolongadas e traumáticas, utilização de tubos traqueais de tamanho inadequado para o diâmetro da via aérea, pressões elevadas nos balonetes dos tubos traqueias, e plano de sedação inadequado, o que proporciona atrito entre o tubo e a mucosa da laringe e da traqueia (Martins & Dias, s.d.).

2.17. LARINGE DISPLÁSICA

Grande parte dos cancros de células escamosas da laringe e hipofaringe iniciam-se como uma condição pré-cancerígena. Apesar de estas células parecerem anormais, vistas a microscópio, elas não se assemelham às células cancerígenas. Na maioria dos casos, a displasia não evolui para cancro e costuma desaparecer sem qualquer tratamento, especialmente se a causa for interrompida, como por exemplo, fumar (Oncoguia, 2018).

Por vezes a displasia evolui para carcinoma *in situ*. Neste caso, as células cancerígenas são vistas apenas no epitélio (Oncoguia, 2018).

2.18. TUMOR DA HIPOFARINGE

O tumor da hipofaringe por norma é uma neoplasia altamente diferenciada (carcinoma espinocelular) localizada na parte de trás da faringe e surge essencialmente em pessoas que consomem bebidas alcoólicas, sistematicamente, ou em fumadores (Rapoport, et al., 2011) (Oncoguia, 2018).

2.19. PÓLIPO DAS CORDAS VOCAIS

Pólipos das cordas vocais é uma massa não cancerígena que surge como consequência do uso excessivo da voz, de reações alérgicas crónicas na laringe ou da inalação crónica de substâncias irritantes, tais como poluentes industriais ou o tabaco. Esta massa, por norma,

dilata durante o crescimento, até se encontrar unida à superfície através de um pendulo. Por norma, ocorre apenas a formação de um pólipo isolado, contudo, há casos, em que pode ocorrer mais que uma formação (Merdau, s.d.).

Normalmente, os pólipos surgem devido ao mau uso da voz, quer isto dizer, quando se utiliza um timbre ou um tom de voz forçado, especialmente quando se utiliza, forçadamente, uma frequência baixa. Assim sendo, é comum o aparecimento dos pólipos em crianças que falam aos gritos, em vendedores ambulantes, telefonistas, cantores, políticos e professores (Merdau, s.d.).

3. SINAIS ACÚSTICOS

3.1. BASE DE DADOS *SAARBRUCKEN VOICE DATABASE*

Neste trabalho foi utilizada a base de dados de ficheiros de fala alemã *Saarbrucken Voice Database* (SVD). Esta base de dados está disponível *online* de forma gratuita pelo Instituto de Fonética da Universidade de *Saarland*.

A base de dados é composta por sinais de vozes de mais de 2000 pessoas com alguma patologia e sem patologias. Cada pessoa tem a gravação dos fonemas /a/, /i/ e /u/ nos tons baixo, normal e alto, variando entre tons, e a frase em alemão "*Guten Morgen, wie geht es Ihnen*?" ("Bom dia, como estás?"). O tamanho dos ficheiros de som situa-se entre 1 e 3 segundos e têm uma resolução de 16 bits e uma frequência de amostragem de 50 kHz.

3.1.1. Sinais Utilizados

Como no trabalho de (Teixeira, et al., 2018) ficou provado que não há diferença entre o género masculino e feminino, neste não se fez a separação por género. Foram utilizadas várias patologias da laringe que podem ser observadas na Tabela 1, juntamente, com o número de pacientes para cada uma das patologias, a média de idades e o desvio padrão.

Tabela 1 - Grupos utilizados, tamanho da amostra, média e desvio padrão das idades

Grupos de Teste	Tamanho da Amostra	Média de Idades	Desvio Padrão Idade
Controlo	194	38,06	14,36
Disfonia	69	47,38	16,27
Laringite Cronica	41	49,69	13,47
Paralisia das Cordas Vocais	169	57,75	13,77
Quisto	3	47,5	15,56
Pólipo das Cordas vocais	27	52,28	13,41
Carcinoma das Cordas Vocais	19	57	6,60
Tumor Laríngeo	4	53,5	8,17
Granuloma	2	44,5	4,5
Granuloma de Intubação	3	53	11,22
Tumor da Hipofaringe	5	59,5	9,29
Fibroma	1	46	0
Laringe Displásica	1	69	0
Edema de Reinke	34	56,10	11,37
Disfonia Funcional	75	47,12	14,54
Disfonia Hipofuncional	12	41,63	15,07
Disfonia Hiperfuncional	127	42,32	13,62
Disfonia Hipotónica	2	49,5	12,5
Disfonia Psicogénica	51	51,40	9,40
Disfonia Espasmódica	62	57,15	15,75

3.2. PARÂMETROS EXTRAÍDOS DO SINAL ACÚSTICO

Nesta secção é feita a descrição de todos os parâmetros utilizados.

Foi utilizado o algoritmo desenvolvido por (Gonçalves, 2015) para extrair o *jitter* e o *shimmer*. Para o HNR modificou-se este algoritmo e ainda se acrescentou o NHR e a autocorrelação

Na Figura 2 pode-se ver uma ilustração do conceito de *jitter* e *shimmer*, onde *jitter* corresponde à medida da variação da duração dos períodos glotais e o *shimmer* corresponde à variação da amplitude dos sucessivos períodos glotais.

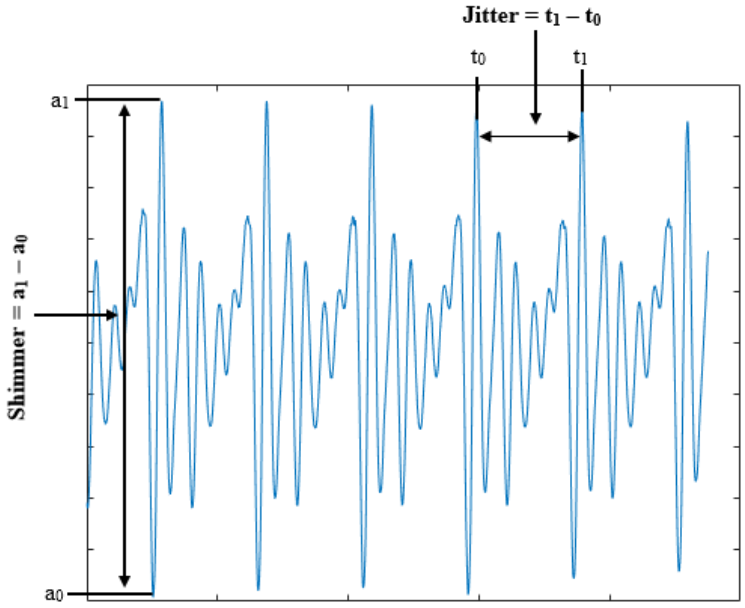

Figura 2 - Representação dos parâmetros *jitter* e *shimmer* para um sinal de fala

Foram extraídos um segundo conjunto de parâmetros, do qual consta os coeficientes mel cepstrais (MFCC), *Linear Prediction Coding* (LPC) e *Line Spectral Frequency* (LSF).

3.2.1. Jitter

O *jitter* é definido como uma medida de variação glotal entre ciclos de vibração das cordas vocais. Sujeitos que não consigam controlar a vibração das cordas vocais têm tendência a ter valores maiores de *jitter*. O *jitter* pode ser medido de quatro maneiras diferentes (Teixeira & Gonçalves, 2014). Porém, neste trabalho, apenas são utilizadas duas dessas formas, *jitter* relativo (*jitter*) e *jitter* Absoluto (*jitta*). Sendo que as outras medidas são a Perturbação Média Relativa (rap) e o Quociente de Perturbação do Período (ppq5) que mede a mesma variabilidade dentro de uma janela de 3 e 5 períodos glotais. Estas duas medidas não são utilizadas, uma vez que, a análise estatística anterior (Teixeira & Fernandes, 2015) mostrou que o *jitter* relativo tem resultados semelhantes com o rap e ppq5.

***Jitter* relativo (*jitter*)** é a diferença absoluta média entre os períodos glotais consecutivos divididos pelo período médio e expresso em percentagem (Eq.1).

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

***Jitter* absoluto (*jitta*)** é a variação do período glotal entre ciclos, ou seja, a diferença absoluta média entre períodos consecutivos, expressa pela Eq. 2.

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

Na equação 1 e 2 T_i é o tamanho do período glotal i e N é o número total de períodos glotais.

3.2.2. Shimmer

O *shimmer* está relacionado com a variação de magnitude ao longo dos períodos glotais. Uma redução na resistência glotal e lesões podem causar variações na magnitude glotal correlacionada com a respiração e a emissão de ruído, dando origem a valores maiores de *shimmer*. Este pode ser medido de quatro maneiras diferentes (Teixeira & Gonçalves, 2014), no entanto, neste trabalho, apenas duas vão ser utilizadas, *Shimmer* Relativo (*Shim*) *e Shimmer* Absoluto (*ShsB*). As outras duas medidas são Quociente de Perturbação de Amplitude em 3 ciclos (APQ3) e Quociente de Perturbação da Amplitude em 5 ciclos (APQ5) que mede a mesma variabilidade dentro de uma janela de 3 ou 5 períodos glotais, respetivamente. A análise estatística anterior (Teixeira & Fernandes, 2015) mostrou que o *shimmer* relativo tem resultados semelhantes aos APQ3 e APQ5.

***Shimmer* relativo (Shim)** é definido como a diferença absoluta média entre as magnitudes de períodos consecutivos, dividida pela magnitude média, expressa em percentagem (Eq. 3).

$$Shim = \frac{\frac{1}{N-1}\sum_{i=2}^{N}|A_i - A_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \quad (3)$$

***Shimmer* absoluto (ShdB)** é expresso como a variação da magnitude pico-a-pico em decibel, ou seja, o algoritmo de base 10 da média absoluta da razão de magnitude entre períodos consecutivos, multiplicada por 20. É expressa em decibel (Eq. 4).

$$ShdB = \frac{1}{N-1}\sum_{i=2}^{N}\left|20 \times \log\left(\frac{A_i}{A_{i-1}}\right)\right| \quad (4)$$

Na equação 3 e 4 A_i é a magnitude do período glotal *i* e N é o número total de períodos glotais.

3.2.3. Parâmetros Harmónicos

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

3.2.3.1. HNR

HNR é um parâmetro em que a relação entre componentes harmónicos e de ruído fornece uma indicação da periodicidade geral do sinal de fala, quantificando a relação entre a componente periódica (parte harmónica) e a componente aperiódica (ruído), expresso em dB (Alves, 2016) (Cordeiro, 2016). A primeira componente decorre da vibração das cordas vocais e a segunda decorre de ruído glotal. Esta avaliação traduz a eficiência do processo de fonação, ou seja, quanto maior for a eficiência na utilização do fluxo de ar expelido pelos pulmões em energia de vibração das cordas vocais, e quanto mais íntegro (saudável) for o ciclo vibratório destas pregas, maior será a relação HNR. Contudo, quanto menor for o ciclo vibratório, menor será o ruído glótico e mais baixa será o valor de HNR (Lopes, et al., 2008). Diferentes autores propõem sua própria maneira de medir o HNR (Boersma, 1993) (Shama & Cholayya, 2007).

Em termos matemáticos um sinal vozeado com estrutura harmónica no domínio das frequências pode ser representado pela Eq.5.

$$X(w) = H(w) + N(w) \tag{5}$$

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

Por definição o HNR é uma medida logarítmica da relação das energias que estão associadas à componente harmónica e de ruído. Através da Eq. 6 é possível a integração espectral ao longo da gama audível de frequências (Gonçalves, 2015).

$$HNR = 10 \times \log_{10} \frac{\int_w |H(w)|^2}{\int_w |N(w)|^2} \tag{6}$$

Tendo em conta os trabalhos publicados por Boersma,1993 (Boersma, 1993), desenvolveu-se este algoritmo. Boersma usa um procedimento baseado nas propriedades da função autocorrelação, para obter a separação de componentes descritas anteriormente. A autocorrelação consiste na correlação de um sinal consigo mesmo. Se considerarmos um sinal de voz x(t), a função autocorrelação $r_x(\tau)$ está representada na Eq. 7.

$$r_x(\tau) \equiv \int x(t)x(t+\tau)dt\,. \tag{7}$$

Nesta função existe um máximo global para τ=0. Caso existam 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 F0 deste sinal periódico é definida pela Eq. 8.

$$F0 = \frac{1}{T_0} \tag{8}$$

Caso não haja máximos globais fora de 0, pode ocorrer haver máximos locais. Caso o maior deles seja no desfasamento τ_{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 harmonica R0 é um número entre 0 e 1, igual ao máximo local $r'_x(\tau_{max})$ da autocorrelação normalizada (Eq. 9).

$$r'_x(\tau) \equiv \frac{r_x(\tau)}{r_x(0)} \tag{9}$$

A autocorrelação total do sinal é a soma das autocorrelações das componentes harmónicas e de ruído como se pode ver na Eq. 10.

$$r_x(0) = r_H(0) + r_N(0) \tag{10}$$

Caso o ruído seja branco (não é possível a correlação com o mesmo), obtém-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).

Com isto, a função autocorrelação de um sinal de voz sustentada exibe máximos locais para valores múltiplos de τ, múltiplos inteiros do período fundamental. Deste modo, para determinar o HNR apenas é necessário calcular a função autocorrelação do sinal de voz e identificar o primeiro máximo local que será correspondente à componente harmónica do sinal, e considerar a restante energia como de ruído, dada pela diferença entre 1 e a energia harmónica. O valor de HNR determina-se através da Eq.11 (Gonçalves, 2015) .

$$HNR(dB) = 10 \times \log_{10} \frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})} \tag{11}$$

3.2.3.2. Autocorrelação

Autocorrelação fornece uma medida das partes de fala semelhantes repetidas ao longo do sinal. Quanto maior o valor de autocorrelação, maior é a repetição de eventos semelhantes ao longo do sinal.

Tendo em conta os trabalhos feitos por (Boersma, 1993) acrescentou-se este parâmetro ao algoritmo desenvolvido por (Gonçalves, 2015).

Inicialmente temos o sinal *x(t)* do qual vamos utilizar uma parte com uma duração T, centrado em t_{mid}. Dessa parte selecionada subtraímos a média μ_x e multiplicamos o resultado por uma função de janela *w(t)* de modo a obtermos uma janela do sinal:

$$a(t) = \left(x\left(t_{mid} - \frac{1}{2}T + t\right) - \mu_x\right) w(t) \tag{12}$$

A função de janela *w(t)* é simétrica em torno de $t = \frac{1}{2}T$ e 0 em todos os lugares fora do intervalo de tempo [0,T]. (Boersma, 1993) diz que a janela a ser utilizada deve ser uma janela sinusoidal ou de *Hanning*, dada pela equação 13.

$$w(t) = \frac{1}{2} - \frac{1}{2}cos\frac{2\pi t}{T} \tag{13}$$

De seguida é necessário calcular a autocorrelação normalizada $r_a(\tau)$ da parte do sinal selecionada. Esta é uma função simétrica ao atraso τ:

$$r_a(\tau) = r_a(-\tau)\frac{\int_0^{T-\tau} a(t)a(t+\tau)\,dt}{\int_0^T a^2(t)\,dt} \tag{14}$$

Na figura 3 é possível ver os passos dados para obter a autocorrelação normalizada da parte selecionada do sinal.

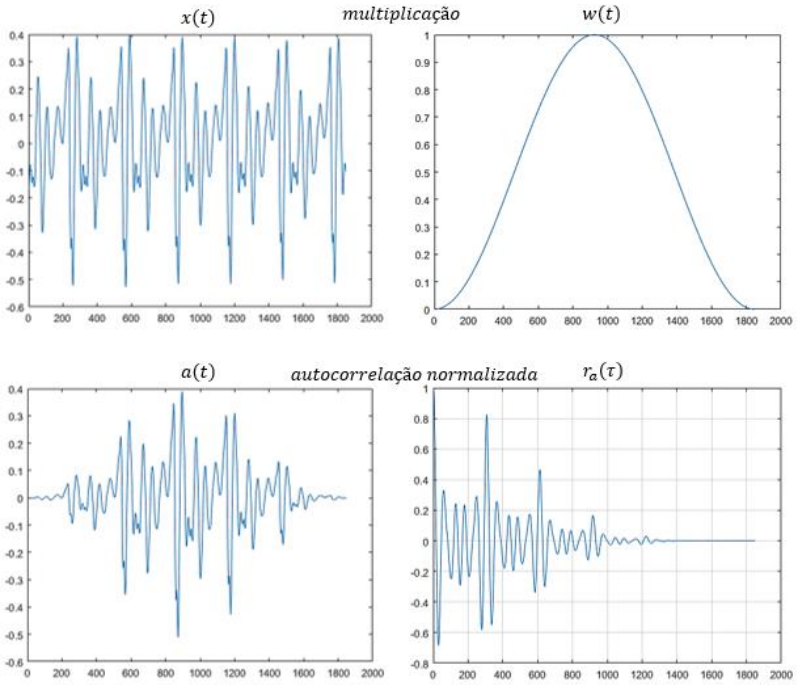

Figura 3 – *x(t)* Segmento do sinal de som, *w(t)* janela de *hanning*, *a(t)* multiplicação do segmento de sinal de som com a janela de *hanning*, $r_a(\tau)$ resultado da autocorrelação normalizada do segmento do sinal de som multiplicado pela janela de *hanning*

Por último é necessário calcular a autocorrelação normalizada $r_w(\tau)$ da função de janela utilizada. Utilizando a janela de *hanning* a autocorrelação é obtida através da equação 15.

$$r_w(\tau) = \left(1 - \frac{|\tau|}{T}\right)\left(\frac{2}{3} + \frac{1}{3}cos\frac{2\pi\tau}{T}\right) + \frac{1}{2\pi}sin\frac{2\pi|\tau|}{T} \qquad (15)$$

Na figura 4 é possível observar a determinação da autocorrelação normalizada da janela de *hanning* $r_w(\tau)$. À esquerda está representada a autocorrelação normalizada da janela de *hanning* e à direita, apenas, a parte positiva da autocorrelação normalizada.

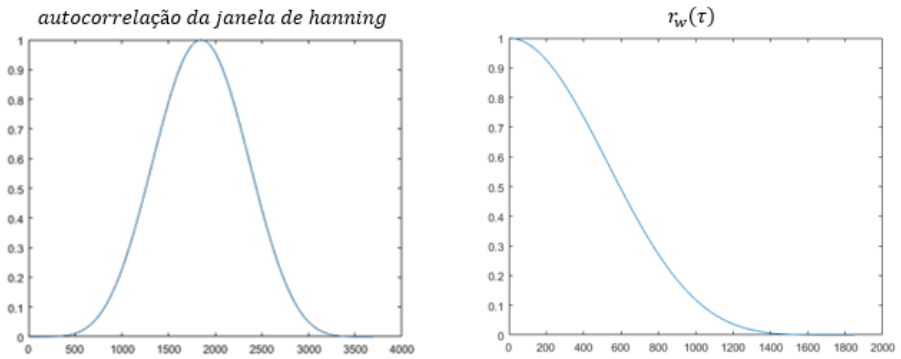

Figura 4 – Representação de $r_w(\tau)$

Para estimar a autocorrelação $r_x(\tau)$ do segmento de sinal original, dividimos a autocorrelação $r_a(\tau)$ da janela do sinal pela autocorrelação $r_w(\tau)$ da janela utilizada (Eq.16)

$$r_x(\tau) = \frac{r_a(t)}{r_w(t)} \tag{16}$$

Na figura 5 estão descritos os passos para se obter a autocorrelação do sinal original $r_x(\tau)$. Na parte superior, do lado esquerdo, está representada a Autocorrelação normalizada do sinal. Do lado direita, da parte superior, temos a parte positiva da Autocorrelação normalizada da janela de *hanning* e por fim, na parte de baixo, temos o resultado da divisão da autocorrelação normalizada do sinal pela autocorrelação da janela de *hanning*.

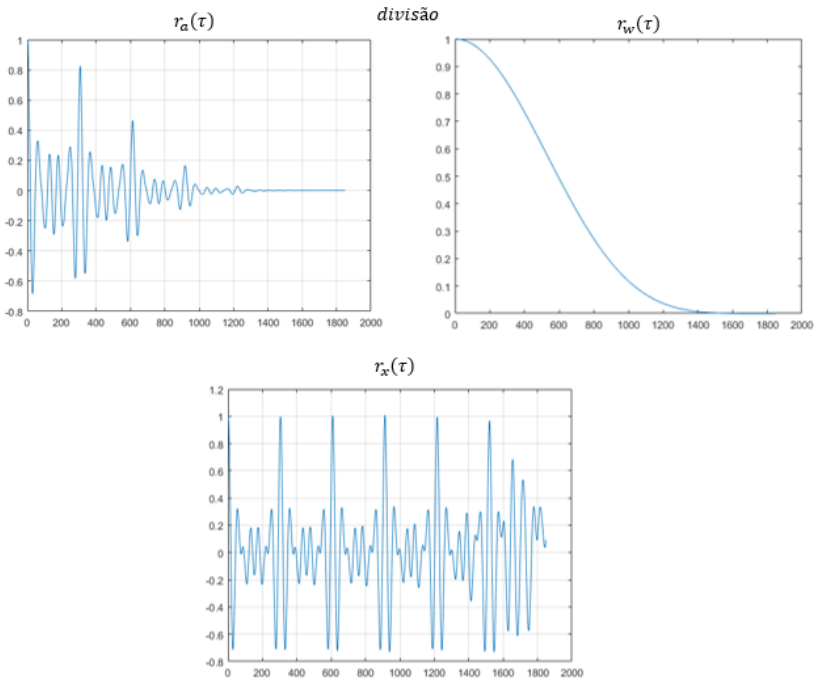

Figura 5 – $r_x(\tau)$ obtêm através da divisão de $r_a(\tau)$ por $r_w(\tau)$

3.2.3.3. NHR

NHR quantifica a relação entre a componente aperiódica (ruído) e a componente periódica (parte harmónica), sendo o inverso de HNR, contudo, a medida não é feita no domínio logarítmico, nem os valores são o inverso (Boersma, 2004). Na Figura 6 é possível observar a determinação do NHR.

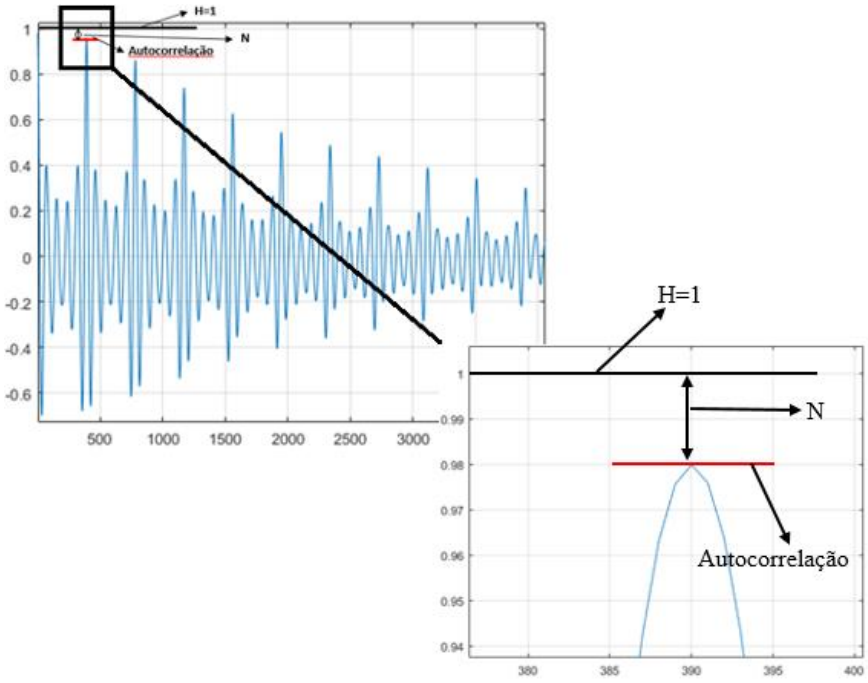

Figura 6 - Determinação do NHR

Tendo em conta a figura 6 o valor de NHR é o que está identificado na imagem como sendo o N. A determinação deste parâmetro é feita tendo em conta a Eq. 17

$$NHR = \frac{N}{H} = \frac{H - Autocorrelação}{H} = \frac{1 - Autocorrelação}{1} = 1 - Autocorrelação \qquad (17)$$

3.2.4. MFCC

Mel Frequency Cepstral Coefficients (MFCC), em Português Coeficientes Cepstrais na Frequência Mel, são parâmetros de curto termo baseados no espectro do sinal. Os MFCCs descrevem o perfil da magnitude do espectro de frequências e são baseados no ouvido humano para o qual a perceção das frequências não segue uma escala linear. Os MFCCs podem definir-se como o cepstro de uma janela de análise determinada a partir de uma DFT numa escala de frequência e magnitude que é característica da audição humana (Logan, 2000) (Costa, 2013) (Alves, 2016).

O cepstro é uma representação do sinal de voz onde um sinal de fonte glotal, de variação temporal lenta, e a resposta do trato vocal, de variação rápida, são desacoplados e transformados em dois componentes aditivos. Os parâmetros cepstrais têm como

particularidade conseguirem separar a excitação do trato vocal, ou seja, a fonte do filtro. No filtro (trato vocal), relativamente aos primeiros coeficientes cepstrais, existem mais particularidades que permitem a distinção entre oradores (Costa, 2013) (Cordeiro, 2016).

Os MFCCs derivam dos coeficientes cepstrais e surgem como a introdução de informação preceptiva, através da filtragem do espectro do sinal com um banco de filtros de escala Mel. Foi necessário criar uma escala Mel, através da qual, os parâmetros MFCC se regem, e onde são utilizados filtros lineares para frequências inferiores a 1000 Hz e logarítmico superior a 1 kHz (Alves, 2016) (Cordeiro, 2016). No cálculo destes parâmetros é necessário seguir uma série de passos como podemos ver na Figura 7.

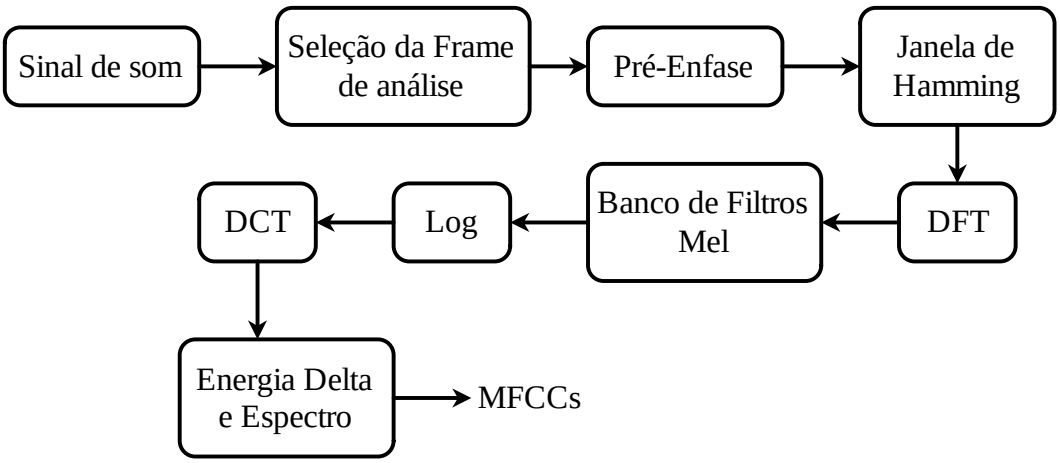

Figura 7 - Diagrama dos passos a seguir para extração dos parâmetros MFCCs

Inicialmente o sinal, aqui representado por *x[n]*, é filtrado, de modo, a realçar as frequências mais elevadas. Neste processo a energia do sinal nas altas frequências é aumentada (Muda, et al., 2010). Este processo designa-se de pré-enfase e está representado na Eq.18.

$$y[n] = x[n] - ax[n-1] \tag{18}$$

Nesta equação, por exemplo, *y[n]* é o sinal filtrado e o valor *a* (coeficiente de pré-enfase). Se este coeficiente for 0,97, significa que 97% de qualquer amostra é originada a partir

da amostra anterior (Alves, 2016). Ressalve-se que na função usada este valor pode ser definido pelo utilizador.

Seguidamente o sinal é dividido em N janelas, designadas por *frames*, com tamanhos que variam entre 20 e 40 ms. Aplica-se uma janela de *Hamming* de acordo com a Eq.17, de forma a reduzir o efeito de *leakage* e obter a magnitude do espectro (Alves, 2016) (Costa, 2013). Onde *w[n]* representa a janela de *Hamming*.

$$w[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \ \leq n \geq N-1 \tag{19}$$

Em seguida converte-se o domínio dos tempos no domínio das frequências utilizando a Transformada de Fourier (FFT). Para cada *frame* é necessário calcular o periodograma da potência espectral. Uma vez que, o espectro tem uma gama de valores bastante alargada e o sinal de voz não segue uma escala linear é necessário aplicar o banco de filtros de acordo com a escala Mel. Utilizam-se filtros passa-banda triangulares que calculam uma soma ponderada das componentes espectrais, a fim de a saída se aproximar à escala Mel com o objetivo de reduzir as características e realçar as frequências, uma vez que, de um modo percetivo, são mais importantes. São consideradas mais significativas as componentes de frequências mais baixas. Em cada filtro existe uma magnitude de 1 no centro e vai decrescendo de forma linear até 0 nas pontas. Temos como exemplo, um banco com 10 filtros o que significa que vamos ter 12 pontos igualmente espaçados, a que vão corresponder 12 coeficientes, sendo o mínimo o valor da frequência mínima do espectro e o máximo o valor da frequência máxima do espectro (Alves, 2016) (Costa, 2013).

Para converter a frequência em Hz para Mel utiliza-se a Eq.20 (Alves, 2016).

$$F(Mel) = 2595 \times \log_{10}\left[\frac{1+f}{700}\right] \tag{20}$$

A transformada discreta do cosseno, do Inglês discrete cosine transform (DCT), permite converter o espectro na base log Mel para o domínio dos tempos. Ao resultado da

conversão dá-se o nome de Coeficientes de Cepstro na Frequência Mel. Designa-se por vetor acústico o conjunto dos coeficientes (Alves, 2016).

Por último é necessário fazer o cálculo da energia e de um fator designado de delta. Este último passo pretende representar a dinâmica do sinal de *frame* para *frame*. Deste modo, aos 12 coeficientes de cepstro é adicionada a energia, perfazendo 13 coeficientes delta ou de velocidade, assim como, 39 coeficientes duplo delta ou de aceleração. A Energia num *frame* do sinal *x*, no tempo *t1* para o tempo *t2* é expressa pela Eq.19 (Muda, et al., 2010) (Alves, 2016).

$$Energia = \sum x^2[t] \tag{21}$$

Os coeficientes delta e duplo delta são conhecidos, também, como coeficientes diferenciais e de aceleração. Os coeficientes MFCC, apenas, representam num único *frame* a potência espectral, porém, o sinal, também, contém informação na sua dinâmica, ou seja, ao longo do tempo, quais são as trajetórias dos coeficientes MFCC. Calculando as trajetórias e juntando-as aos coeficientes MFCC a performance de um sistema de análise acústica pode aumentar. Assim sendo, teríamos 13 coeficientes MFCC, 13 delta e 13 duplo delta, fazendo um total de 39 coeficientes, como foi referido anteriormente. Sendo que, cada um dos 13 coeficientes delta representa a variação de *frame* para *frame* (Muda, et al., 2010) (Alves, 2016). Para o cálculo dos coeficientes delta utiliza-se a Eq.22.

$$d(t) = \frac{c(t+1)-c(t-1)}{2} \tag{22}$$

Onde *d(t)* representa o coeficiente delta da *frame t*, calculado em termos de coeficientes estáticos *c(t+1)* e *c(t-1)*. Os coeficientes duplo delta são calculados da mesma forma, porém, a partir dos coeficientes delta e não dos coeficientes estáticos (Alves, 2016).

Hoje em dia, os Coeficientes Cepstrais na Frequência Mel, são os parâmetros mais utilizados para todo o tipo de aplicações de reconhecimento de fala e orador (Cordeiro, 2016).

3.2.5. *Linear Prediction Coefficients* (LPC)

LPC (*Linear Prediction Coefficientes*) ou LP (*Linear Prediction*) é uma poderosa técnica de análise da fala. Inicialmente esta técnica era usada para estimar parâmetros básicos da fala, como, por exemplo, a frequência fundamental, formantes, larguras de banda e funções de área do trato vocal. É um método importante, uma vez que, é relativamente fácil e rápido estimar os parâmetros da fala e com uma elevada precisão (Teixeira, 1995).

Na análise por predição linear qualquer amostra do sinal de fala pode ser aproximado por uma combinação linear das amostras anteriores. A minimização da soma das diferenças quadradas (num intervalo finito) entre a amostra de fala atual e a predita linearmente leva a que, apenas, seja possível determinar um conjunto de coeficientes de predição, sendo que, estes coeficientes são o peso dos coeficientes usados na combinação linear (Teixeira, 1995).

A predição linear está relacionada com o modelo de fala, em que a fala é modelada como a saída de um sistema linear variante no tempo, excitado por impulsos quase periódicos, durante a fala vocalizada, ou ruído aleatório na fala não vocalizada. Este método é robusto, fiável e preciso para se estimarem os parâmetros que caracterizam o sistema linear variante no tempo (Teixeira, 1995).

3.2.5.1. Princípios Básicos da Análise por Predição Linear

A Figura 8 representa o modelo equivalente para análise por predição linear.

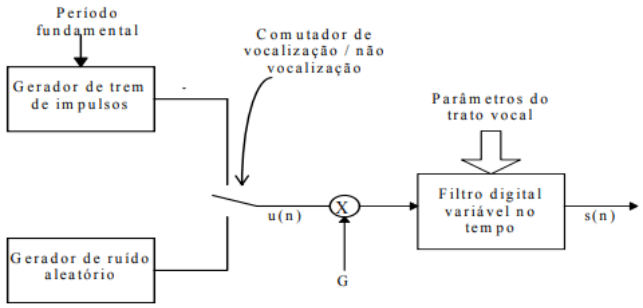

Figura 8 - Diagrama de blocos do modelo simplificado de produção de fala (Teixeira, 1995)

A representação de filtro digital variante no tempo é a combinação dos efeitos de radiação, trato vocal e excitação glotal e a função do sistema em estado estacionário está representada na Eq. 23 (Teixeira, 1995).

$$H(z) = \frac{S(z)}{U(Z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{23}$$

Para que este sistema seja excitado é necessário impulsos para a fala vocalizada ou por uma sequência de ruído aleatório para a fala não vocalizada. Assim sendo, os parâmetros deste modelo são: classificação do modo de excitação, período fundamental para a fala vocalizada, parâmetro ganho G, e os coeficientes $\{a_k\}$ do filtro digital. Estes parâmetros vão variando lentamente com o tempo (Teixeira, 1995).

Este é um modelo simplificado só com polos e tem como vantagem poder estimar-se de forma direta e computacionalmente eficiente através do método de análise por predição direta o parâmetro de ganho G, e os coeficientes do filtro $\{a_k\}$ (Teixeira, 1995).

Para este sistema as amostras de fala s*(n)* estão relacionadas com a excitação *u(n)* por uma equação às diferenças (Eq.24) (Teixeira, 1995) (Sathler, 2008).

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \tag{24}$$

Um dos modelos mais utilizados é o de um filtro de resposta impulsiva finita (FIR), onde um sistema de predição linear com os coeficientes de predição $\{\alpha_k\}$ tem a saída definida pela Eq.25 (Teixeira, 1995) (Sathler, 2008).

$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{25}$$

Onde p é a ordem de predição e $\{\alpha_k\}$ são os coeficientes de predição. Através da transformada-z do filtro de predição obtêm-se a Eq.26 (Teixeira, 1995) (Sathler, 2008).

$$P(z) = \sum_{k=1}^{p} \alpha_k z^{-k} \tag{26}$$

O erro de predição *e(n)* é definido através da Eq.27 (Teixeira, 1995) (Sathler, 2008).

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \tag{27}$$

Através da equação anterior é possível observar que a sequência do erro de predição é a saída de um sistema cuja função de transferência é dada pela Eq.28 (Teixeira, 1995) (Sathler, 2008).

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \tag{28}$$

Comparando as equações 24 e 27 consegue-se perceber que se o sinal de fala obedece exatamente à equação modelo (28), e se $\alpha_k = a_k$, então $e(n)=Gu(n)$. Ou seja, o filtro de predição linear, *A(z)*, será um filtro inverso para o sistema, *H(z)* da equação 23, isto é expresso através da Eq.29 (Teixeira, 1995) (Sathler, 2008).

$$H(z) = \frac{G}{A(z)} \tag{29}$$

O grande problema da análise por predição linear é a determinação do conjunto de coeficientes de predição $\{\alpha_k\}$ diretamente do sinal de fala de forma a obter uma boa estimativa das propriedades espectrais do sinal de fala com recurso ao uso da equação 29. Uma vez que os sinais de fala variam no tempo, os coeficientes de predição devem ser estimados em segmentos curtos do sinal de fala. A aproximação básica é encontrar o conjunto de coeficientes de predição que minimizem o erro quadrático médio de predição, num curto segmento da forma de onda do sinal de fala (Teixeira, 1995) (Sathler, 2008).

3.2.5.2. Métodos de Predição Linear

Para se desenvolver um procedimento de análise em intervalos curtos de tempo, os limites devem estar dentro de um intervalo finito. Existem três métodos que permitem o cálculo dos coeficientes de predição. Seguidamente segue um resumo destes métodos (Sathler, 2008).

Método	Características
Algoritmo de *Burg*	• Baseado na estrutura de *Lattice* • Minimiza a energia dos erros de predição futuros *fm(n)* • Minimiza os erros de predição passados *bm(n)* • Cálculo dos coeficientes *fm* e *bm* é recursivo: $k_m = \frac{2\sum_{n=m}^{N-1}[f_m - 1(n)b_m - 1(n-1)]}{\sum_{n=m}^{N-1}[f_m^2 - 1(n)b_m^2 - 1(n-1)]}$ $f_m(n) = f_{m-1}(n) - k_m b_{m-1}(n-1), n = m+1, \dots, N-1$ $b_m(n) = b_{m-1}(n-1) - k_m f_{m-1}(n), n = m, , \dots, N-1$
Método de Autocorrelação	• Minimiza os erros de predição • Solução eficiente pela recursão de *Levinson-Durbin* $k_m = \frac{R(m) - \sum_{k=1}^{m-1} a_k^{n-1} R(m-k)}{E_{m-1}}$ $a_k^m = k_m$ $a_k^m = a_k^{m-1} - k_m a_{m-k}^{m-1}, k = 1, , \dots, m-1$ $E_m = (1 - k_m^2)E_{m-1}$
Gradiente Adaptativo de *Lattice*	• Semelhante ao algoritmo de *Burg*, minimização da soma dos erros de predição passados e futuros. $D_m(n) = \lambda D_m(n-1) + (1-\lambda)[f_{m-1}^2(n) + b_{m-1}^2(n-1)]$ $2\mu_m = \frac{\alpha}{D_m(n)}$

3.2.6. *Line Spectral Frequency* (LSF)

Os LSF geralmente são usados para codificação de voz devido à sua grande eficiência de codificação e suas propriedades atraentes para interpolação (Alencar & Alcaim, 2008). Os parâmetros LSF definem-se como as frequências correspondentes às raízes de dois polinómios de ordem $p+1$, $P(z)$ e $Q(z)$, derivadas do filtro inverso de predição linear $A(z)$, de ordem p (sendo p a ordem de predição), onde, $P(z)$ corresponde ao trato vocal com a fonte glotal completamente fechada (coeficiente de reflexão $k_{p+1}=1$) e $Q(z)$ representa o trato vocal com a fonte glotal completamente aberta (coeficiente de reflexão $k_{p+1}=-1$) (Cordeiro, 2016).

Os parâmetros LSF podem ser obtidos através da conversão dos coeficientes de predição linear (LPC), em que, cada raiz de $A(z)$ é convertida num par de raízes complexas conjugadas no círculo unitário. Caso o espectro de entrada seja plano, os parâmetros LSF estão separados uniformemente entre 0 e $F_s/2$. Se uma raiz de $A(z)$ apresentar um valor do seu módulo perto da unidade, a largura de banda é estreita sendo muito provável que este pólo seja um formante, contudo, se uma raiz de $A(z)$ apresentar um valor de módulo baixo, será grande a largura de banda correspondente e a sua contribuição traduzir-se-á apenas na inclinação espectral, estando as raízes correspondentes de $P(z)$ e $Q(z)$ afastadas. Assim sendo é possível relacionar a distância entre duas raízes consecutivas, $P(z)$ e $Q(z)$, com a largura de banda dos formantes, sendo, também, esta uma característica do orador (Cordeiro, 2016).

4. DESENVOLVIMENTO

4.1. EXTRAÇÃO DOS PARÂMETROS

Neste trabalho pretendia-se extrair um conjunto de parâmetros, *jitter* relativo, *jitter* absoluto, *shimmer* relativo, *shimmer* absoluto, HNR, NHR e autocorrelação. Desenvolveu-se um algoritmo de forma a poder-se extrair o NHR, de uma outra forma, e um segundo conjunto de parâmetros que correspondem aos coeficientes cepstrais na frequência mel, os coeficientes LPC e os parâmetros LSF.

4.1.1. Algoritmo para Extração do HNR, NHR e autocorrelação

O algoritmo desenvolvido por (Gonçalves, 2015) permite a extração de 9 parâmetros, *jitter* absoluto, *jitter* relativo, *jitter* rap, *jitter* ppq5, *shimmer* absoluto, *shimmer* relativo, *shimmer* apq3, *shimmer* apq5 e HNR.

Contudo, este algoritmo na determinação do HNR não obtém os melhores resultados, daí a necessidade de se ter modificado este parâmetro e ainda adicionar o NHR e a autocorrelação.

Para calcularmos a autocorrelação foi implementada a equação 16 com uma janela de *hanning* de 6 períodos glotais e para o NHR a equação 17.

Este algoritmo foi usado, apenas (tal como justificado no capítulo "3.2 - Parâmetros Extraídos do Sinal Acústicos"), para extrair 7 destes parâmetros, *jitter* relativo, *jitter* absoluto, *shimmer* relativo, *shimmer* absoluto, HNR, NHR e autocorrelação, para três vogais e três tons diferentes a partir de sinais disponíveis na base de dados SVD. As vogais disponíveis são /a/, /i/ e /u/ e os tons baixo, normal e alto.

Para determinar o HNR, o NHR e a autocorrelação ao longo do sinal, é necessário segmentá-lo em *frames*, o comprimento destes vai ser analisado à frente. Inicialmente começa-se por definir o comprimento das *frames* em função de períodos glotais. De seguida determina-se a autocorrelação normalizada de uma janela (por exemplo *Hanning* ou *Hamming*) com o mesmo comprimento das *frames*. Cada *frame* do sinal vai ser multiplicado por uma janela com o mesmo comprimento da *frame*, esta janela também vai ser analisada à frente. Com o resultado da multiplicação é calculada a autocorrelação e dividida pela autocorrelação normalizada. Este valor é devolvido em vetor. Para

acharmos o valor da autocorrelação acha-se o pico máximo em metade do segmento. Por último, é calculado para cada valor do pico o HNR através da equação 11. Tendo estes valores para todas as *frames* faz-se a média do HNR e da autocorrelação.

Através da figura 9 é possível observar o fluxograma do algoritmo desenvolvido.

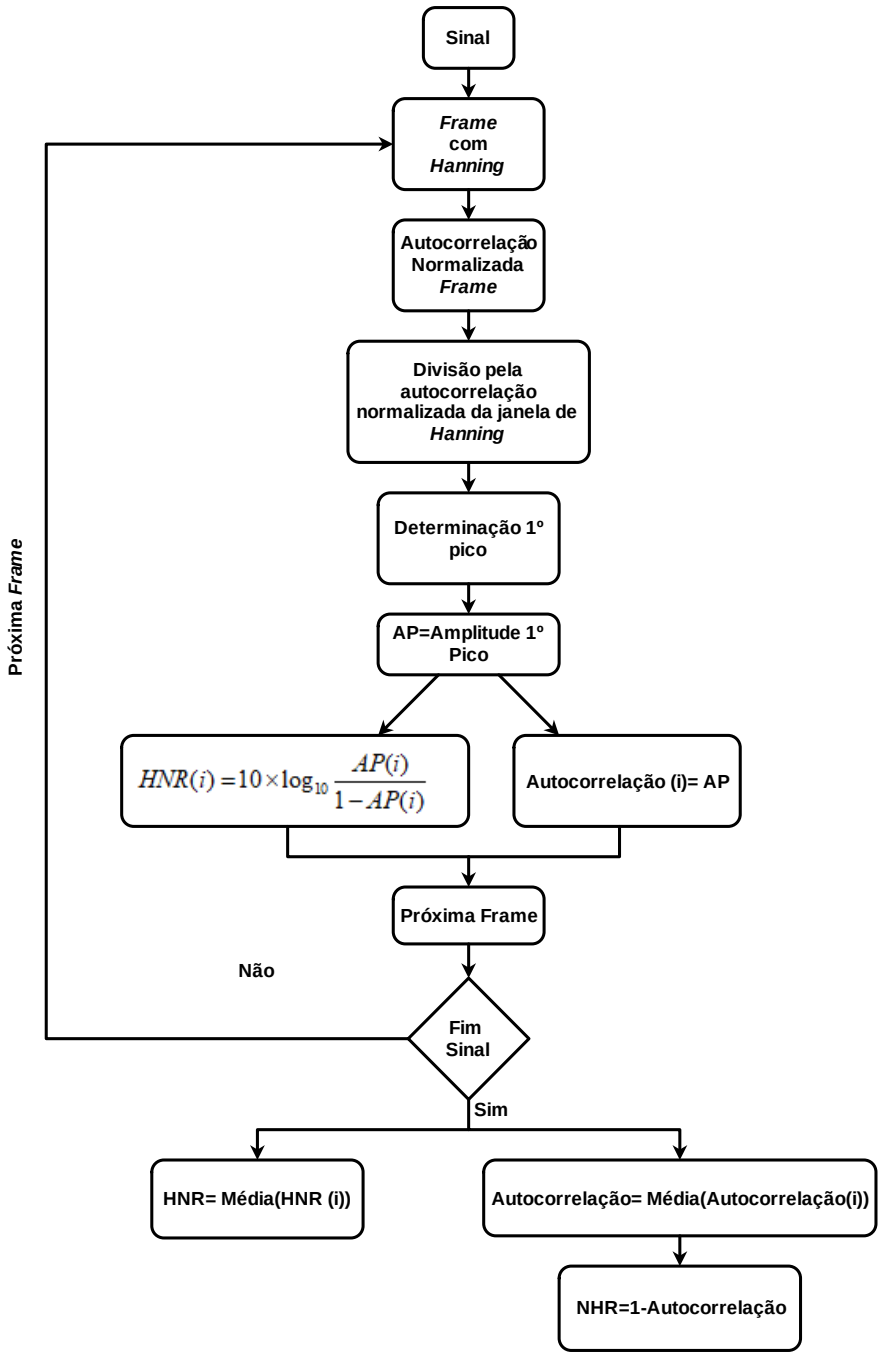

Figura 9 - Fluxograma do algoritmo para determinar o parâmetro HNR, NHR e autocorrelação

Desenvolveu-se uma segunda forma de determinar o NHR de acordo com (Daliyski, 1993) e (Ogawa, et al., 1986). Para este algoritmo começa-se por fazer a transformada de Fourier do sinal de voz *X(k)*. Tendo a transformada, foi necessário calcular os máximos da função (M). Para encontrar o primeiro máximo da função foi definida uma gama para a frequência fundamental de 75 Hz a 300 Hz para as mulheres e de 75 Hz a 180 Hz para os homens, e para o segundo pico, para as mulheres, procurava desde a frequência em que se encontra o primeiro pico até mais 300 Hz, e para os homens funciona exatamente da mesma maneira, porém, em vez de mais 300 Hz é mais 180 Hz. Na figura 10 é possível observar como se determinou o NHR.

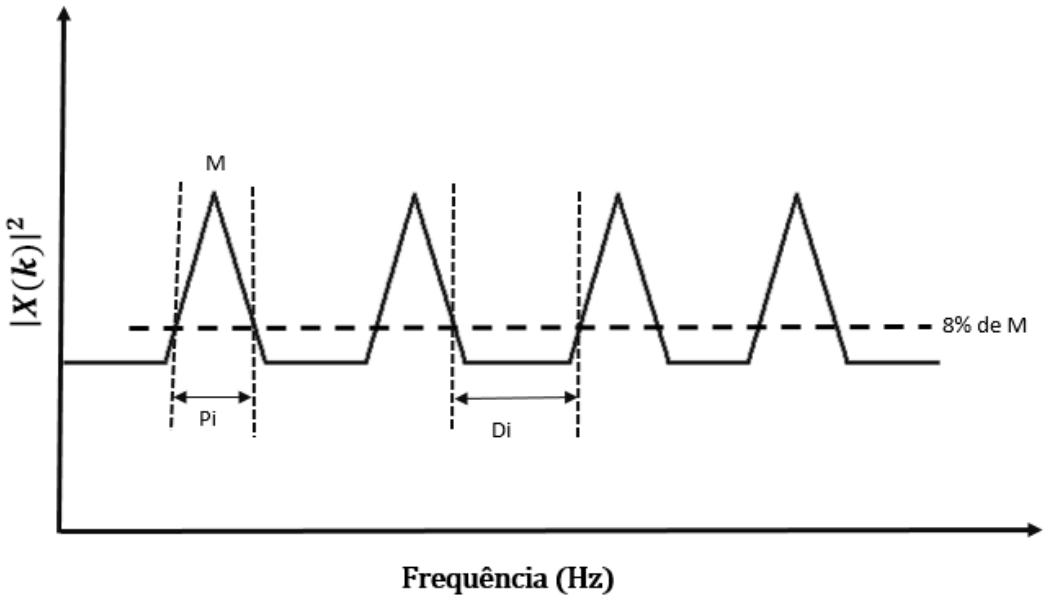

Figura 10 – Ilustração esquemática da determinação do NHR

Para definir o Pi (intervalo que define os picos – componentes harmónicas) fez-se 8% de cada máximo da função (0,08*M) e assim foi possível determinar os intervalos Pi. Ao valor de 8% chegou-se experimentalmente. Os intervalos Di (vales – componentes de ruído) eram definidos entre o fim de um Pi e o início do seguinte Pi. Definidos os picos e vales calcula-se a potência para cada pico e vale. De seguida é feita a soma da potência de todos os picos e de todos os vales e por último para determinarmos o NHR fazemos a divisão entre o somatório da potência dos vales e o somatório da potência dos picos.

Na determinação dos picos existiram dificuldades, que levaram a que este algoritmo não fosse utilizado na determinação do NHR, como é possível observar através da figura 11.

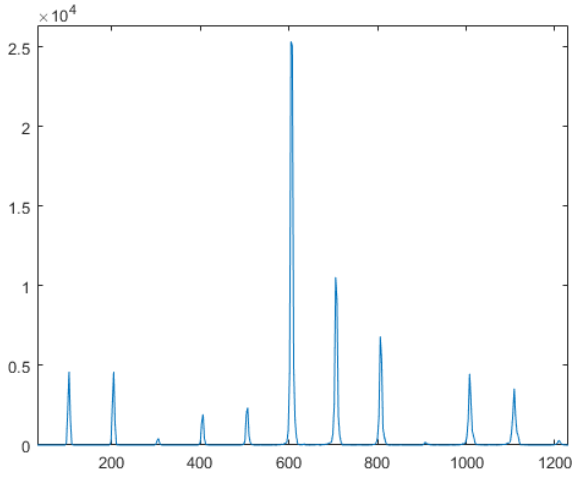

Figura 11 - Transformada de Fourier de um sujeito de controlo

Na figura 11 é possível observar que por volta dos 300 Hz existe um pico muito menor e logo de seguida existem picos de maior valor (400 Hz). Uma das dificuldades que surgiu, e que, também, não permite um bom resultado dos valores de NHR, através deste método, é determinar quando o algoritmo deve parar de procurar picos, uma vez que o pico por volta dos 300 Hz tem uma energia bastante reduzida, mas os picos seguintes são necessários. Esta situação volta a acontecer nos 900 Hz.

Para além deste problema para os maus resultados do algoritmo, a falha deste já tinha sido anteriormente confirmada por (Boersma, 2004), uma vez que, não permite obter valores inferiores a 0,1.

4.1.2. Determinação MFCCs, LPC e LSF

Antes de se passar à extração dos parâmetros, foi necessário selecionar apenas a parte do sinal onde ocorre fala. Para tal, aplicou-se a média deslizante com uma janela de *hanning* de modo a conseguirmos identificar os períodos de fala e de silêncio. Para esta janela foi utilizado um comprimento de 32 ms. Através da análise de vários sons foi possível perceber que, quando a média deslizante era superior a um determinado limite (10^{-6} para esta base de dados), ocorria fala, desse modo evitaram-se as zonas de silêncio iniciais e finais.

Para a extração dos coeficientes cepstrais na frequência mel recorreu-se à função disponível em https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab (MathWorks, 2011), que devolve os coeficientes cepstrais na frequência mel (MFCC). Nesta função inicialmente o sinal é pré-enfatizado usando um filtro FIR de primeira ordem através de um coeficiente de pré-enfase fornecido pelo utilizador. Este sinal é sujeito a uma análise de curto termo com tamanho de janela e intervalos definidos pelo utilizador através da aplicação da transformada de Fourier. De seguida calcula-se a potência espectral e aplica-se o banco de filtros triangulares uniformemente espaçados entre a frequência mínima e máxima na escala mel. Por último aplicasse um filtro sinusoidal. Como parâmetros de entrada utilizou-se o seguinte: duração da *frame* de análise (N=35), deslocamento da *frame* (varia em função da duração do sinal, de forma a dar um total de 50 janelas), coeficiente de pré-enfase (a=0,97), gama de frequências a considerar na análise [300, 3700], número de canais do banco de filtros (20), número de coeficientes cepstrais (13) e parâmetro de alisamento cepstral (c=22).

Para a extração dos coeficientes de predição linear utilizou-se como exemplo a documentação do MatLab (https://www.mathworks.com/help/signal/ref/lpc.html) (MathWorks, 2006). A função LPC devolve os coeficientes de predição linear. Esta função utiliza o método de autocorrelação da modelação autorregressiva para encontrar os coeficientes do filtro. Para esta função é necessário indicar a ordem do LPC. O número de coeficientes deve ter em conta a frequência de amostragem em kHz, uma vez que, deve ser mais dois a quatro polos que esta. Para este caso o valor mínimo ideal será 52 coeficientes, tendo em conta que a frequência de amostragem é de 50 kHz (Vieira, et al., 2016).

Tendo os coeficientes de predição linear apenas é necessário uma função que permite converter os LPC em LSF (*line spectral frequency*). Para tal recorreu-se à documentação do MatLab https://www.mathworks.com/help/signal/ref/poly2lsf.html#d120e128903 (MathWorks, 2006).

5. RESULTADOS E DISCUSSÃO

Nesta secção vão ser relatados os resultados de algumas análises feitas, assim como a discussão das mesmas.

Fez-se uma análise comparativa para os valores de HNR e autocorrelação, entre os valores obtidos pelo algoritmo e pelo Praat, de modo, a sabermos qual o melhor comprimento de janela e a melhor janela a utilizar.

Esta análise não foi feita para o NHR, porque, ao usar 1 menos a autocorrelação, o resultado fica *per si* justificado.

De forma a garantir a veracidade dos resultados, usaram-se os valores obtidos pelo algoritmo para 10 sujeitos de uma vogal e tom, comparando, esses mesmos 10 sujeitos, com os valores obtidos pela referência. Esta segunda análise foi feita para a HNR, autocorrelação e NHR.

Com o objetivo de obter valores de HNR e autocorrelação do algoritmo próximos dos obtidos pelo *software* Praat, utilizou-se este *software* como valor de referência, e testou-se a influência da janela e o seu comprimento. Experimentaram-se janelas de *hanning*, *hamming* e *blackman*, e para cada janela fez-se variar o comprimento da janela correspondente a 3, 6, 12 e 24 períodos glotais. A escolha destes comprimentos de janela baseia-se nos valores utilizados por (Boersma, 1993) que usou os mesmos comprimentos de janela. Gonçalves, 2015, experimentou 5, 10, 20 e 50 períodos glotais e selecionou como comprimento 10 períodos glotais. No entanto, os resultados com 10 períodos glotais que (Gonçalves, 2015) obteve não eram suficientemente próximos dos valores de referência do Praat, daí a necessidade de se refazer esta análise do comprimento de janela de acordo com a reformulação do HNR.

A sobreposição das janelas também foi testada, no entanto, os valores foram semelhantes aos que não tiveram uma sobreposição de janela e, portanto, decidiu-se usar o algoritmo sem sobrepor as janelas para não sobrecarregar desnecessariamente o peso computacional do algoritmo.

Esta análise foi feita utilizando uma amostra de 10 sujeitos de controlo (5 masculinos e 5 femininos) e 10 sujeitos pacientes (5 masculinos e 5 femininos), de 3 das doenças

utilizadas. Destes, 4 sujeitos tinham laringite crónica, dois de cada género, 4 sujeitos com disfonia, dois de cada género e 2 com paralisia das cordas vocais, um de cada género. Os sons são provenientes da base de dados SVD com uma frequência de amostragem (Fa) de 50 kHz. Os sujeitos de controlo têm idades compreendidas entre 40 e 65 anos, tendo em média 55,10 anos, e os sujeitos pacientes têm idade compreendida entre 16 e 77 anos, tendo em média 51,10 anos. Considera-se que esta diferença na média das idades de apenas 4 anos não interfira, significativamente, nos resultados da análise.

5.1. HNR

Para os 10 sujeitos de controlo e patológicos foi extraído o HNR com as três janelas e, para cada janela, utilizaram-se os quatro comprimentos de janela.

Tendo os valores dos parâmetros fez-se a média, ou seja, para a janela de *hanning* com 6 períodos glotais fez-se a média do HNR dos 10 sujeitos de controlo. O processo repetiu-se para todas as janelas e todos os comprimentos, tanto nos sujeitos de controlo como nos patológicos.

De modo a perceber-se qual a melhor janela e período glotal a utilizar, com os resultados das médias foi feita a subtração, em módulo, com o valor de referência. Com os valores da subtração fez-se uma nova média usando os valores obtidos para as 3 vogais e 3 tons, para cada janela e comprimento, em que, por exemplo, para a janela de *hamming* de 6 períodos glotais, fez-se a média de todos os valores com esta janela e comprimento. Na tabela 2 é possível observar os resultados desta análise.

Tabela 2 - Média das diferenças do HNR relativamente a cada janela e comprimento

Janela	**Comprimento da janela (n.º de Períodos Glotais)**	**Médias**	
		Pacientes	Controlo
Hamming	3	2,47	3,6
	6	0,42	0,5
	12	1,69	2,23
	24	3,1	3,88
Hanning	3	3,33	4,95
	6	**0,26**	**0,42**
	12	1,59	2,07
	24	2,94	3,68
Blackman	3	8,61	11,56
	6	0,37	0,4
	12	1,21	1,66
	24	2,56	3,21

Através dos dados da tabela 2 é possível observar que para os sujeitos pacientes a melhor janela é a de *hanning* com 6 períodos glotais e para os sujeitos de controlo é a janela de *hanning* ou a de *blackman*, ambas, com 6 períodos glotais. Uma vez que só pode ser escolhida uma das janelas, vai-se considerar a janela de *hanning* como sendo a melhor considerando os sujeitos de controlo e pacientes.

Analisadas as médias dos sujeitos, com base no algoritmo e na referência, faz-se de seguida uma análise dos valores individuais por sujeito.

Assim, selecionou-se, aleatoriamente, uma vogal para os sujeitos de controlo e pacientes. Para o algoritmo a janela escolhida foi a referida anteriormente, *hanning* com um comprimento de 6 períodos glotais.

As vogais utilizadas para o HNR são /i/ tom normal para controlo e pacientes.

Nas figuras 12 e 13 é possível observar os resultados de HNR para os 10 sujeitos de controlo e pacientes, respetivamente.

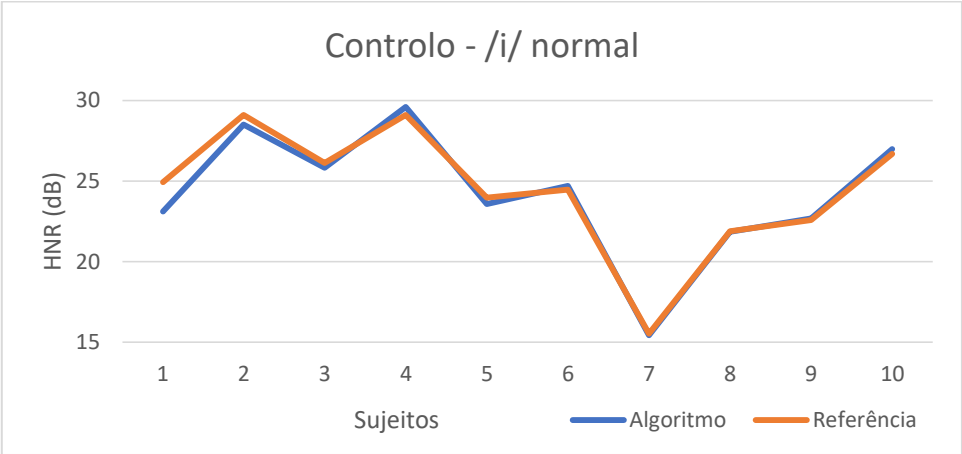

Figura 12 – Comparação dos valores de HNR para os 10 sujeitos de controlo

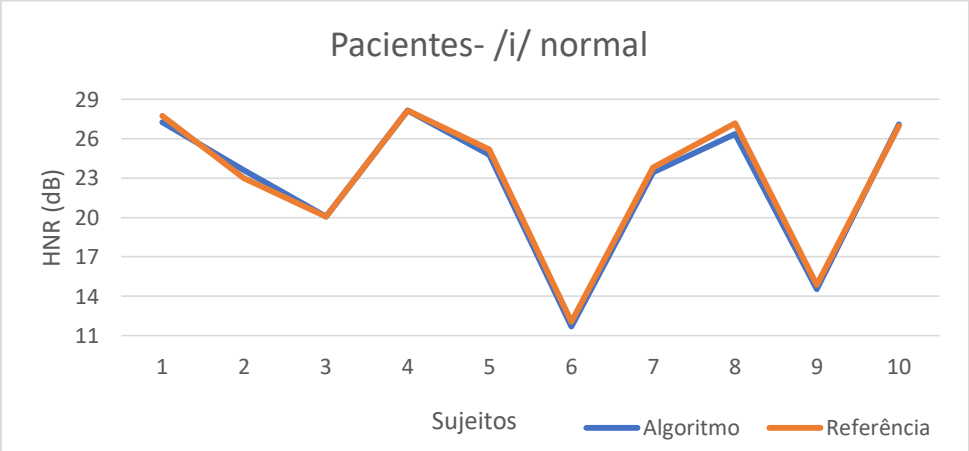

Figura 13 - Comparação dos valores de HNR para os 10 sujeitos pacientes

Através das figuras anteriores é possível observar que os resultados de HNR do algoritmo e de referência são bastante semelhantes. Nos sujeitos de controlo, o erro absoluto é inferior a 1,82 dB, e nos pacientes é inferior a 0,80 dB.

Assim sendo, conclui-se que a janela a ser usada deve ser a de *hanning* com um comprimento de 6 períodos glotais. Esta janela de *hanning* com este comprimento já foi confirmada nos trabalhos realizados anteriormente por (Boersma, 1993). (Gonçalves, 2015), nos seus trabalhos, confirmou apenas esta janela.

5.2. AUTOCORRELAÇÃO

A análise realizada para o HNR foi, também, feita para a autocorrelação, a fim de saber qual a melhor janela e comprimento a utilizar.

Deste modo, com os resultados das médias fez-se a subtração, em módulo, com o valor de referência, tal como está descrito na tabela 2. Na tabela 3 é possível observar os resultados desta análise.

Tabela 3 - Média das diferenças da autocorrelação relativamente a cada janela e comprimento

Janela	**Comprimento da janela (n.º de Períodos Glotais)**	**Médias**	
		Pacientes	Controlo
Hamming	3	0,005	0,001
	6	**0,004**	**0,001**
	12	0,007	0,004
	24	0,018	0,009
Hanning	3	0,006	0,002
	6	**0,004**	**0,001**
	12	0,007	0,003
	24	0,017	0,009
Blackman	3	0,013	0,009
	6	**0,004**	**0,001**
	12	0,006	0,002
	24	0,015	0,007

Através dos dados da tabela 3 é possível observar que os melhores resultados se obtêm com as 3 janelas para 6 períodos glotais

Em conclusão devemos usar 6 períodos glotais com umas das janelas de *hanning*, *hamming* ou *blackman*.

Analisadas as médias dos sujeitos, com base no algoritmo e na referência, fez-se uma outra onde se compara os mesmos sujeitos, mas de forma individual.

Assim, selecionou-se uma vogal para os sujeitos e para os de controlo. Para o algoritmo a janela escolhida foi a de *hanning* com um comprimento de 6 períodos glotais.

As vogais selecionadas para a autocorrelação são /i/ tom normal para controlo e pacientes. O uso destas vogais justifica-se pois foram as escolhidas também para o HNR.

Na figura 14 é possível observar os resultados da autocorrelação para os 10 sujeitos de controlo e na figura 15 os resultados para os 10 sujeitos pacientes.

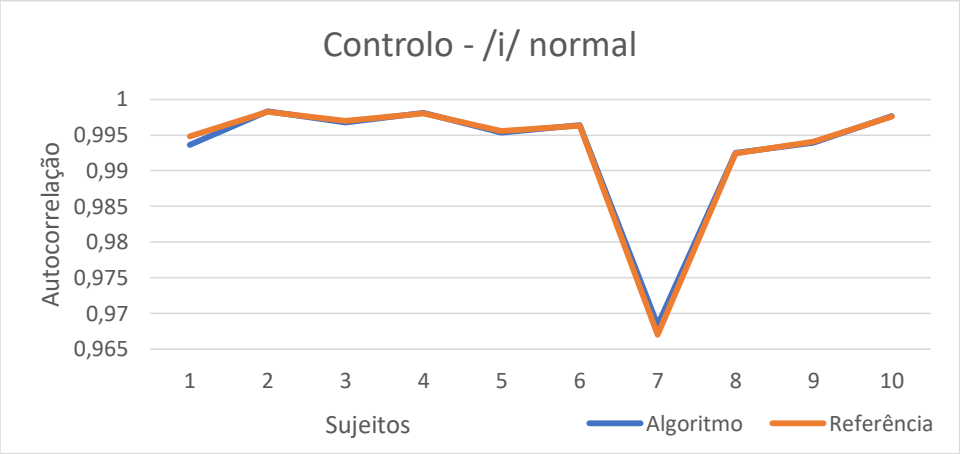

Figura 14 - Comparação dos valores da autocorrelação para os 10 sujeitos de controlo

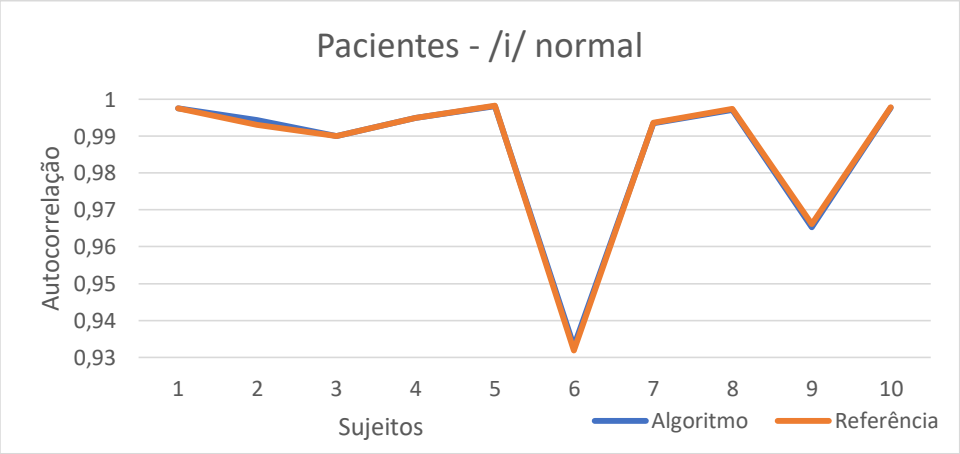

Figura 15 - Comparação dos valores da autocorrelação para os 10 sujeitos pacientes

Através das figuras 14 e 15 é possível observar que os resultados obtidos pelo algoritmo e pela referência para a autocorrelação são praticamente idênticos. Em ambos os casos o erro absoluto é inferior a 0,001.

Assim sendo, conclui-se que a janela a ser usada deve ser a de *hanning* com um comprimento de 6 períodos glotais.

5.3. NHR

Este método é determinado em função da autocorrelação (1-autocorrelação), e, deste modo, não se realizou a análise das médias de forma a sabermos qual a melhor janela e período glotal, uma vez que foi feito para a autocorrelação e fica justificado por si.

Contudo, realizou-se a análise entre sujeitos, de modo, a perceber-se como se comporta o algoritmo comparativamente ao valor de referência.

Para o NHR utilizou-se, também, a janela de *hanning* com 6 períodos glotais, uma vez que é o valor utilizado para a autocorrelação, e as pessoas e vogais são as mesmas que as utilizadas para o HNR e autocorrelação.

Nas figuras 16 e 17 é possível observar os resultados do NHR para os sujeitos de controlo e pacientes, respetivamente.

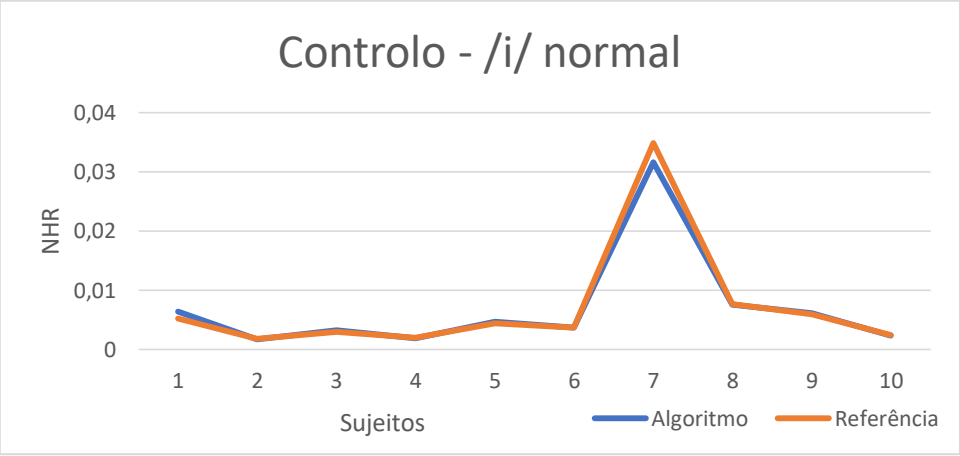

Figura 16 - Comparação dos valores de NHR para os 10 sujeitos de controlo

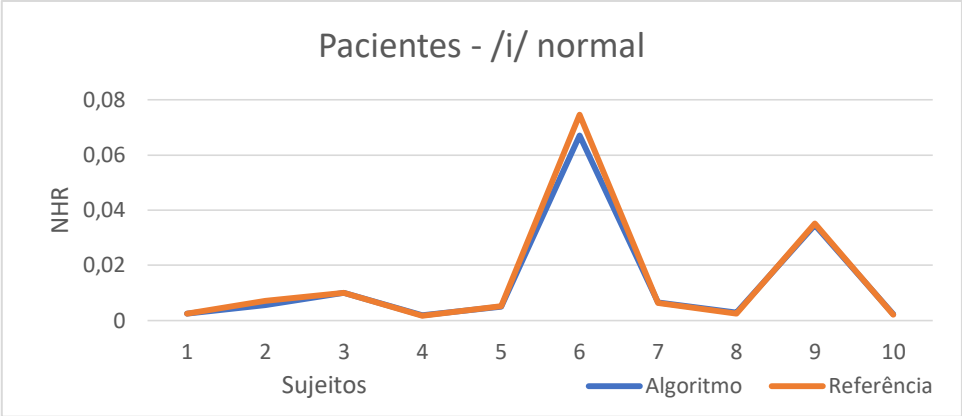

Figura 17 - Comparação dos valores de NHR para os 10 sujeitos pacientes

Através das imagens anteriores é possível observar que os resultados obtidos pelo algoritmo e pela referência para o NHR são praticamente idênticos tanto para a figura 16 como para a 17. Nos sujeitos de controlo, o erro absoluto é inferior a 0,003, e nos pacientes é inferior a 0,008.

Concluindo assim que este algoritmo obtém valores bastante próximos dos valores de referência, podendo ser usado para extrair este parâmetro com excelentes resultados.

5.4. VARIAÇÃO DA FREQUÊNCIA DE AMOSTRAGEM

Com a finalidade de testar como se comporta o algoritmo com a variação da frequência de amostragem, para o HNR e para a autocorrelação, fez-se a decimação dos mesmos sons e passou-se a ter uma frequência de amostragem de 25 kHz e 12,5 kHz.

O processo de decimação pretende reduzir a frequência de amostragem original de um sinal contínuo. Contudo é necessário ter em atenção a possível ocorrência de *aliasing*. De forma a garantir que isto não aconteça, uma vez que este processo é reversível, a operação de decimação é normalmente precedida de uma filtragem passa baixo digital e só depois é feita a decimação do sinal.

A decimação dos sinais foi feita no *software* MatLab e utilizou-se a função *decimate*.

Para esta análise considerou-se o mesmo método que para uma frequência de amostragem de 50 kHz, quando se pretendia determinar qual a melhor janela e período glotal a utilizar. Contudo, nesta apenas se utiliza a janela de *hanning* com 6 períodos glotais, uma vez que

esta janela e comprimento foram os que obtiveram melhores resultados para os dois parâmetros anteriormente.

5.4.1. HNR Com Variação da Frequência de Amostragem

De forma a verificar se a frequência de amostragem influência os valores obtidos pelo algoritmo e pela referência, após se realizar a análise das médias das vogais para o parâmetro HNR para uma frequência de amostragem de 12,5 kHz e 25kHz, fez-se a análise igual à feita na tabela 2. Na tabela 4 é possível observar o resultado da análise para as duas frequências de amostragem.

Tabela 4 - Média das diferenças do HNR em função da frequência de amostragem

Frequência de Amostragem (kHz)	**Média**	
	Pacientes	Controlo
25	0,58	1,16
12,5	2,03	3,45

Através dos dados da tabela 4 é possível observar que há um ligeiro aumento das diferenças do valor medido pelo algoritmo comparativamente aos valores obtidos com Fa de 50 kHz. Sendo que, este aumento é mais acentuado para Fa de 12,5 kHz.

De seguida foi feita uma análise mais aprofundada dos resultados apresentados nas tabelas 2 e 4 possibilitando a comparação da média e desvio padrão dos valores do algoritmo com os de referência, verificando-se, assim, se o HNR varia com a frequência de amostragem. Na tabela 5 apresenta-se o resultado para o HNR. Nesta análise, para cada frequência de amostragem está a média de todas as pessoas de controlo e pacientes (180 sinais utilizados) para a janela utilizada e períodos glotais.

Tabela 5 - HNR em função da frequência de amostragem

Frequência de Amostragem (kHz)	Algoritmo	Referência
50	23,89 (σ=2,56)	24,02 (σ=2,32)
25	22,98 (σ=2,59)	23,85 (σ=2,49)
12,5	21,85 (σ=2,55)	24,59 (σ=2,33)

Através da análise da tabela 5 é possível observar que o HNR, para os valores do algoritmo varia com a frequência de amostragem e tem variação máxima em 2,04 dB.

Uma vez que, os valores do algoritmo variam um pouco com a frequência de amostragem, fez-se uma análise comparativa entre as 3 frequências de amostragem para algumas vogais e tons, escolhidos aleatoriamente, tanto para pacientes como para controlo, de modo a tentar perceber se a diferença é em algum som específico ou se ocorre em vários sons. Para esta análise foram usadas as mesmas 10 pessoas que anteriormente. As figuras 18 e 19 correspondem aos pacientes e as figuras 20 e 21 aos de controlo.

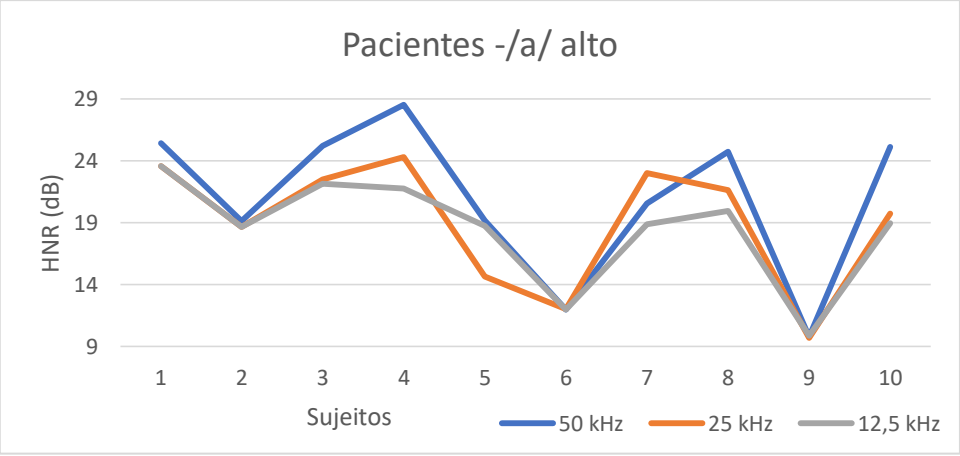

Figura 18 - Análise comparativa usando as 3 frequências de amostragem para pacientes com a vogal /a/ tom alto

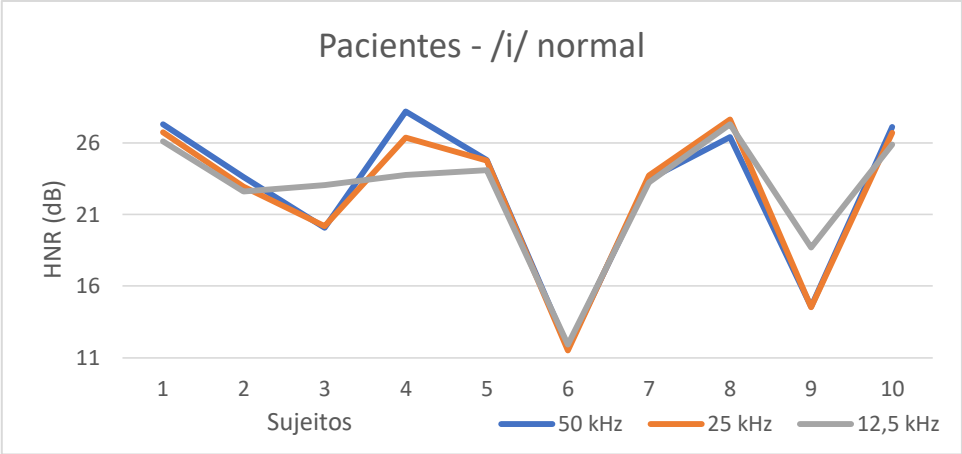

Figura 19 - Análise comparativa usando as 3 frequências de amostragem para pacientes com a vogal /i/ tom normal

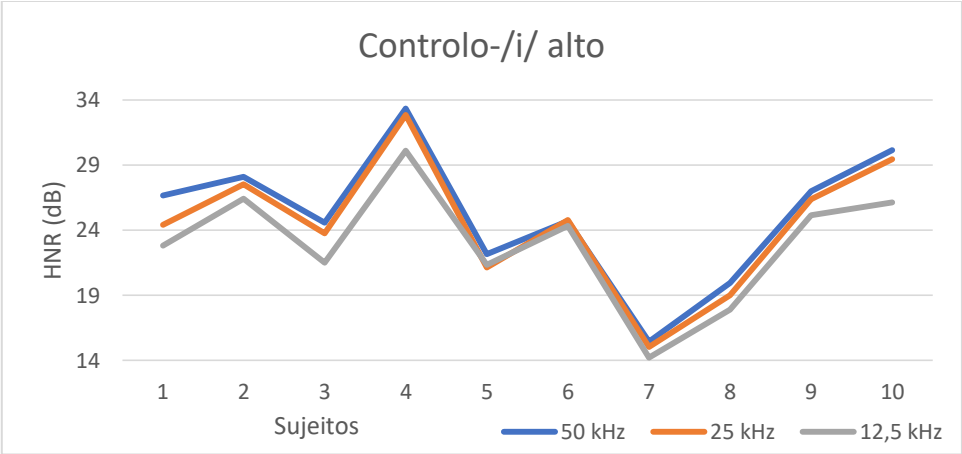

Figura 20 - Análise comparativa usando as 3 frequências de amostragem para controlo com a vogal /i/ tom alto

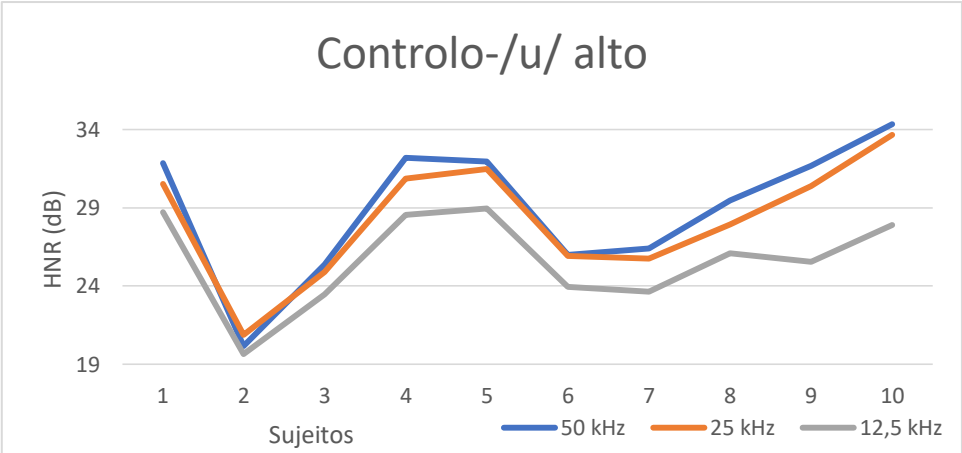

Figura 21 - Análise comparativa usando as 3 frequências de amostragem para controlo com a vogal /u/ tom alto

Através das figuras anteriores, da figura 18 à 21, é possível observar que há algumas variações na medição do HNR. As variações ocorrem principalmente na medição com uma Fa de 12,5 kHz em todas as vogais e tons apresentados, porém, para a vogal /a/ tom alto as variações ocorrem, também, para uma Fa de 25 kHz.

Uma vez que existe esta variação em função da frequência de amostragem, fez-se uma segunda análise de forma a garantir que o algoritmo possa ser utilizado para extrair o HNR e ser utilizado para a distinção entre sujeitos patológicos e saudáveis. Para tal fez-se a média de todos os sujeitos (90 sons para controlo e 90 para patológicos) em função da frequência de amostragem. Esta análise pode ser vista na tabela 6.

Tabela 6 - HNR em função da Frequência de Amostragem para os grupos de controlo e pacientes

Frequência de Amostragem (kHz)	**Pacientes**	**Controlo**
50	21,91	25,60
25	21,41	24,55
12,5	20,72	22,98

Através da tabela 6 é possível observar que, mesmo havendo variação na frequência de amostragem, ocorre variação entre os sujeitos patológicos e os de controlo.

Conclui-se assim que este algoritmo pode ser utilizado, cautelosamente, para determinar este parâmetro, pois, quando se baixa a frequência de amostragem, está a baixar-se a frequência útil do sinal, perdendo-se componentes de ruído, o que pode influenciar esta medida. A diferença menor entre sujeitos patológicos e de controlo é de 2,26 dB, o que é aceitável para distinguir estes sujeitos.

5.4.2. Autocorrelação Com Variação da Frequência de Amostragem

De forma a verificar se os valores obtidos pelo algoritmo e pela referência, para a autocorrelação, são influenciados pela frequência de amostragem, após se realizar a análise das médias das vogais para este parâmetro, para uma frequência de amostragem

de 12,5 kHz e 25kHz, fez-se a análise igual à feita na tabela 2. Na tabela 7 é possível observar o resultado da análise para este parâmetro para as duas frequências de amostragem.

Tabela 7 - Média das diferenças da autocorrelação em função da frequência de amostragem

Frequência de Amostragem (kHz)	**Média**	
	Pacientes	Controlo
25	0,006	0,001
12,5	0,006	0,004

Na tabela 7 é possível observar que os resultados são semelhantes aos obtidos para a autocorrelação com Fa de 50 kHz, havendo uma variação insignificante nos sujeitos pacientes e nos de controlo para uma Fa de 12,5 kHz.

De seguida foi feita uma análise mais aprofundada dos resultados apresentados nas tabelas 3 e 7 possibilitando a comparação da média dos valores do algoritmo com os de referência, verificando-se, assim, se a autocorrelação varia com a frequência de amostragem. Na tabela 8 apresenta-se o resultado para a autocorrelação. Nesta análise, para cada frequência de amostragem está a média de todas as pessoas de controlo e pacientes (180 sinais utilizados) para a janela utilizada e períodos glotais.

Tabela 8 - Autocorrelação em função da frequência de amostragem

Frequência de Amostragem (kHz)	**Algoritmo**	**Referência**
50	0,981 (σ=0,009)	0,984 (σ=0,007)
25	0,980 (σ=0,010)	0,984 (σ=0,007)
12,5	0,980 (σ=0,010)	0,984 (σ=0,007)

Analisando a tabela 8 é possível observar que a autocorrelação tem uma variação insignificante, uma vez que, o erro absoluto entre as Fa é de 0,001.

Com isto, conclui-se que a frequência de amostragem não influência o valor da autocorrelação.

5.5. DISCUSSÃO

Foram feitas análises para o HNR, autocorrelação e NHR de modo a perceber se este algoritmo obtém valores destes parâmetros próximos dos valores de referência.

Uma das análises foi perceber qual a janela e o seu comprimento para que o valor dos parâmetros fosse o mais próximo possível do valor de referência.

Para o HNR os sujeitos pacientes obtiveram como melhor janela a de *hanning* com 6 períodos glotais e os de controlo a janela de *hanning* ou a de *blackman*, ambas, com 6 períodos glotais.

Contudo, para as restantes análises para este parâmetro, só se poderia escolher uma janela com um comprimento glotal, e a escolhida como sendo a melhor foi a de *hanning* com 6 períodos glotais considerando os sujeitos de controlo e pacientes.

Na análise, comparativa entre sujeitos, é possível perceber que os resultados de HNR do algoritmo e de referência são bastante semelhantes, uma vez que os gráficos estão quase sempre sobrepostos. Nos sujeitos de controlo, o erro absoluto é inferior a 1,82 dB, e nos pacientes é inferior a 0,80 dB.

Deste modo, conclui-se que a janela a ser usada para o HNR deve ser a de *hanning* com um comprimento de 6 períodos glotais. Esta janela com este comprimento já foi confirmada nos trabalhos realizados anteriormente por (Boersma, 1993). (Gonçalves, 2015), nos seus trabalhos, confirmou apenas esta janela.

Para a autocorrelação também se pretendia saber qual a melhor janela e comprimento a utilizar de forma a obter, através do algoritmo, o valor para este parâmetro o mais próximo possível do valor de referência.

Assim sendo, através da análise feita, percebe-se que para este parâmetro as 3 janelas (*hamming*, *hanning* e *blackman*) com 6 períodos glotais obtêm os melhores resultados.

Contudo, para as restantes análises e extração dos parâmetros, vai ser considerada a janela de *hanning* com 6 períodos glotais, pois através do fluxograma (figura 9) é possível perceber que se o HNR e a autocorrelação utilizarem a mesma janela e comprimento haverá uma libertação da carga computacional.

Na análise comparativa entre sujeitos é possível observar que os resultados obtidos pelo algoritmo e pela referência para a autocorrelação são praticamente idênticos, tanto para pacientes como sujeitos de controlo. Em ambos os casos o erro absoluto é inferior a 0,001.

Para o NHR não se analisou a janela nem o comprimento desta, uma vez que esta análise é feita para a autocorrelação, ficando assim justificada a escolha da janela e comprimento.

Contudo, a análise comparativa entre sujeitos foi feita, e os resultados obtidos pelo algoritmo e pela referência são praticamente idênticos. Nos sujeitos de controlo, o erro absoluto é inferior a 0,003, e nos pacientes é inferior a 0,008.

Fez-se uma outra análise de modo perceber-se se os valores de HNR e da autocorrelação são influenciados pela variação da frequência de amostragem.

Para os mesmos ficheiros de som fez-se a decimação e obteve-se uma frequência de amostragem 12,5 kHz e de 25 kHz.

Para o HNR os valores do algoritmo variam com a frequência de amostragem e o erro absoluto máximo é de 2,04 dB.

Fez-se uma outra análise para perceber se a variação ocorria em algum som especifico ou se ocorria em vários e chegou-se à conclusão que as variações ocorrem principalmente na medição com uma Fa de 12,5 kHz em todas as vogais e tons apresentados, porém, para a vogal /a/ tom alto as variações ocorrem, também, para uma Fa de 25 kHz.

Numa outra análise entre os valores de controlo e patológicos pode concluir-se que este algoritmo pode ser utilizado cautelosamente, para determinar o HNR quando se utilizam frequências de amostragem baixas, pois quando se baixa a frequência de amostragem, está a baixar-se a frequência útil do sinal, perdendo-se componentes de ruído, o que pode influenciar esta medida. O erro absoluto menor entre sujeitos patológicos e de controlo é de 2,26 dB, o que é um valor aceitável para distinguir estes sujeitos.

Para a autocorrelação a variação da frequência de amostragem é insignificante, uma vez que o erro absoluto entre as Fa é de 0,001 comparando com o algoritmo, e 0,004

comparando o algoritmo com o valor de referência. Concluindo assim que que a frequência de amostragem não influência o valor da Autocorrelação.

.

6. BASE DE DADOS CURADA

A base de dados curada foi construída tendo em conta os sons disponíveis na base de dados SVD.

Nesta base de dados curada foram incluídas 19 doenças e sujeitos de controlo, descritas anteriormente neste trabalho. Tem um número total de amostras de 707 sujeitos patológicos (distribuídas por todas as doenças, no subcapítulo 3.1.1 é possível verificar o número de sujeitos correspondentes a cada doença) e 194 sujeitos de controlo, sendo que, para cada sujeito, existe uma frase e três vogais /a/, /i/ e /u/, cada vogal com três tons: alto, baixo e normal.

Na base de dados curada consta uma página inicial onde contem o id de cada sujeito, a idade, género, grupo de caracterização que corresponde ao sujeito e ainda a abreviatura que é utilizada para esse grupo. Na figura 22 é possível observar como está construída esta primeira parte.

ID N.	Age	Gender	Group Characterization	Abbreviation
1714	51	F	Carcinoma of the Vocal Cords	CVC
110	62	M	Carcinoma of the Vocal Cords	CVC
500	44	M	Carcinoma of the Vocal Cords	CVC
1048	59	M	Carcinoma of the Vocal Cords	CVC
1239	57	M	Carcinoma of the Vocal Cords	CVC
1245	57	M	Carcinoma of the Vocal Cords	CVC
1332	64	M	Carcinoma of the Vocal Cords	CVC

Figura 22 – Primeira página da base de dados curada

Para cada grupo existem 3 páginas. Na figura 23 é possível observar que, por exemplo, para os sujeitos de controlo constam 3 páginas com parâmetros distintos, que serão descritos de seguida.

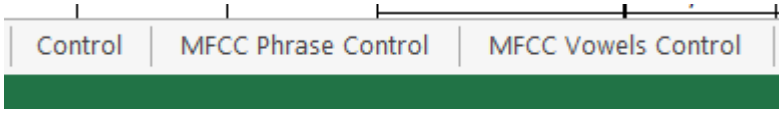

Figura 23 – Distribuição dos parâmetros por grupos de teste

Na primeira página de cada grupo constam os parâmetros: *jitter* absoluto, *jitter* relativo, *shimmer* absoluto, *shimmer* relativo, HNR, NHR e autocorrelação. Para cada sujeito está identificado ainda a idade e o género do sujeito, bem como a vogal e o tom a que o parâmetro corresponde. Na figura 24 é possível observar um exemplo desta página.

Patient			Recorded Sound		Jitter		Shimmer		Harmonicity		
ID N.	Age	Gender	Vowel	Tone	Absolute	Relative (%)	Relative (%)	Absolute(dB)	HNR (dB)	NHR	Autocorrelation
1	20	F	a	High	15,792	0,589	1,370	0,120	31,135	0,001	0,999
				Low	13,314	0,246	1,036	0,091	27,652	0,002	0,998
				Normal	27,519	0,755	2,509	0,219	23,208	0,007	0,993
			i	High	11,171	0,419	0,372	0,033	34,232	0,001	0,999
				Low	18,765	0,346	2,268	0,198	17,776	0,020	0,980
				Normal	11,004	0,304	0,587	0,051	26,066	0,003	0,997
			u	High	11,348	0,432	0,611	0,054	36,249	0,000	1,000
				Low	22,683	0,422	1,840	0,160	28,280	0,002	0,998
				Normal	76,443	2,124	2,380	0,207	28,979	0,002	0,998
10	22	F	a	High	34,420	0,782	4,063	0,355	23,707	0,005	0,995
				Low	18,526	0,364	6,492	0,568	20,923	0,010	0,990
				Normal	20,484	0,412	4,323	0,384	20,956	0,010	0,990
			i	High	22,542	0,516	3,958	0,341	22,605	0,006	0,994
				Low	27,888	0,546	5,060	0,441	20,911	0,009	0,991
				Normal	38,382	0,777	6,064	0,528	19,446	0,012	0,988
			u	High	221,899	5,145	30,991	3,291	11,875	0,080	0,920
				Low	44,843	0,888	20,282	1,894	12,452	0,069	0,931
				Normal	70,638	1,437	32,806	3,439	8,907	0,139	0,861

Figura 24 – Primeira página correspondente a cada grupo

Na segunda página constam os MFCCs das frases gravadas. Nesta constam 13 coeficientes cepstrais ao longo da frase, perfazendo um total de 50 segmentos da frase. Na figura 25 é possível observar parte da distribuição deste parâmetro.

Patient			Number of cepstral coefficients												
ID N.	Age	Gender		1	2	3	4	5	6	7	8	9	10	11	12	13
1	20	F	1	40,629	50,285	58,260	63,675	41,252	58,353	63,249	62,148	59,934	59,439	66,060	67,152	67,268
			2	-0,672	9,484	4,680	-3,285	-2,223	-11,515	-4,380	-1,345	-3,054	-2,947	7,390	6,063	6,553
			3	0,393	8,230	1,826	-4,654	-0,011	2,150	0,169	3,506	4,801	2,086	2,000	-2,043	-0,624
			4	-6,836	-1,818	-10,142	3,021	-1,581	1,944	2,727	3,157	-5,813	-3,158	-11,373	-19,286	-19,049
			5	-2,941	-5,460	-9,148	4,340	3,676	-0,609	-0,817	-3,957	-8,561	-3,041	-6,162	-0,060	-2,319
			6	-7,544	-0,525	2,056	0,667	-0,785	-5,166	-4,726	-5,668	0,254	7,639	2,170	-0,455	1,143
			7	-1,584	-1,276	-0,419	-10,280	-3,542	4,758	-12,052	-13,217	-8,392	-15,090	-5,179	1,514	-0,285
			8	-2,809	6,426	-5,786	-5,515	-4,471	3,795	0,087	7,684	7,285	-2,200	-2,454	-5,356	-8,342
			9	3,873	12,630	-0,923	9,335	4,524	-5,939	-16,727	-3,228	-12,773	-10,075	-2,709	3,444	1,985
			10	3,497	-0,425	-11,409	-9,383	-0,013	3,586	-8,101	-9,065	-14,634	-8,903	-2,047	3,627	4,210
			11	4,385	3,783	-5,179	-4,754	3,721	-2,481	-5,617	-6,734	-14,264	-12,203	2,787	0,970	-1,988
			12	0,471	-2,334	-17,603	-5,302	-4,489	0,186	-2,307	-6,639	-5,656	-17,876	2,609	7,318	3,660
			13	-1,544	-0,326	-14,060	-2,354	0,463	1,256	11,778	-2,941	-12,571	-20,269	-3,364	-0,048	3,103

Figura 25 - Segunda página correspondente a cada grupo

Por último, na terceira página constam os MFCCs para as vogais. Este parâmetro foi extraído também com 13 coeficientes cepstrais para cada vogal e tom, porém, para as vogais o som não foi segmentado, como aconteceu para a frase, extraindo-se um valor por coeficiente para o som todo. Na figura 26 pode ver-se como está organizado este parâmetro.

Patient			Recorded Sound		Number of cepstral coefficients												
ID N.	Age	Gender	Vowel	Tone	1	2	3	4	5	6	7	8	9	10	11	12	13
1	20	F	a	High	67,965	1,349	-1,319	-16,046	5,959	-2,746	2,348	10,535	14,577	31,756	20,841	13,264	-9,660
				Low	70,221	6,675	-3,596	-15,870	-2,339	-2,356	-3,699	-0,234	4,346	17,483	-2,171	-1,152	-2,138
				Normal	68,907	0,177	-6,134	-18,689	-2,976	-3,239	-7,146	-3,722	-0,534	-0,666	-8,800	-9,654	-8,978
			i	High	65,039	-4,670	14,529	9,859	2,031	15,785	8,172	9,832	19,350	15,203	9,480	8,682	-4,401
				Low	58,543	-12,696	20,092	2,520	-0,820	-1,971	4,439	-11,153	9,117	7,770	0,258	5,430	-3,198
				Normal	61,472	-10,027	14,225	-2,513	-4,384	7,333	-7,956	-2,145	0,990	-5,933	-0,142	3,845	4,935
			u	High	60,170	5,780	12,087	4,268	3,904	8,561	10,458	24,627	19,265	15,055	5,075	-7,813	-7,507
				Low	54,649	9,062	4,293	2,642	0,409	-1,514	2,009	-6,877	10,762	0,152	-0,115	4,106	0,939
				Normal	56,375	5,787	-0,133	-8,789	-6,062	1,358	-1,196	0,642	4,806	-1,760	-1,035	-2,048	4,233

Figura 26 - Terceira página correspondente a cada grupo

7. CONCLUSÕES E TRABALHOS FUTUROS

7.1. CONCLUSÕES

Na realização deste trabalho foi indispensável a aprendizagem de um conjunto de conceitos relacionados com o sinal de fala. A parte inicial teve como objetivo a pesquisa de parâmetros que possam ser utilizados para que posteriormente se possa fazer a distinção entre sujeitos saudáveis e patológicos, tais como *jitter, shimmer,* HNR, NHR, autocorrelação e MFCCs.

De seguida foi feita uma pesquisa a nível anatómico e dos sintomas que caracterizam as doenças utilizadas para a criação da base de dados curada.

Neste trabalho utilizou-se a base de dados SVD, uma vez que, esta apresenta sinais de voz de vogais sustentadas e de sinais de fala de uma frase que são as condições essenciais para o tipo de análise pretendida e a isto alia-se o facto de ser de acesso livre.

Implementaram-se os algoritmos para determinar o HNR, a Autocorrelação e o NHR e otimizaram-se as escolhas das janelas e do seu comprimento. O algoritmo utiliza um número inteiro de períodos glotais para determinar o *frame* a analisar, deste modo, foi necessário fazer uma análise comparativa entre os valores do algoritmo e os valores de referência.

Nesta experimentaram-se as janelas de *hanning, hamming* e *blackman* com comprimentos de 3, 6, 12 e 24 períodos glotais.

Para a extração dos parâmetros HNR, NHR e autocorrelação foi selecionada a janela de *hanning* com 6 períodos glotais como sendo a melhor janela. A escolha da janela e do seu comprimento está de acordo com os resultados de (Boersma, 1993) e de (Gonçalves, 2015), pois, este último já utilizou a janela de *hanning* com um comprimento de 10 períodos glotais.

Para percebermos como se comporta o algoritmo comparativamente com a referência, fez-se uma outra análise onde se compara, para uma vogal, os 10 sujeitos utilizados. Deste

modo, percebeu-se que os resultados são bastante bons e muito próximos dos de referência. Para o HNR, para os sujeitos de controlo, o erro absoluto é inferior a 1,82 dB, e nos pacientes é inferior a 0,80 dB, para a autocorrelação em ambos os casos o erro absoluto é inferior a 0,001 e para o NHR nos sujeitos de controlo, o erro é inferior a 0,003, e nos pacientes a 0,008.

Testou-se se o HNR é influenciado pela variação da frequência de amostragem, e chegou-se à conclusão que este parâmetro, com este algoritmo, tem uma pequena variação, sendo esta máxima em 2,04 dB entre os valores para uma frequência de amostragem de 50 kHz e 12,5 kHz. Contudo, este algoritmo pode ser utilizado, cautelosamente, para determinar este parâmetro quando se utilizam frequências baixas, pois, o erro absoluto menor entre sujeitos patológicos e de controlo, ocorre com uma frequência de amostragem de 12,5 kHz, e é de 2,26 dB, o que é aceitável para distinguir estes sujeitos. Para a autocorrelação foi feita a mesma análise e percebeu-se que a variação é insignificante, uma vez que o erro absoluto máximo entre as frequências é de 0,001.

Neste trabalho foi desenvolvido um outro algoritmo de forma a determinar o NHR por outro método, contudo, este algoritmo não obteve bons resultados quando comparados com os valores de referência. No desenvolvimento deste algoritmo surgiram algumas dificuldades, contudo, (Boersma, 2004) já tinha confirmado que este algoritmo não consegue obter valores inferiores a 0,1.

Foi ainda desenvolvido um outro algoritmo de forma a determinar os MFCCs, os LPCs e os LSF. Este utiliza fala contínua, onde era selecionada apenas a parte onde ocorre fala e, para tal, aplicou-se a media deslizante com uma janela de *hanning* de modo a conseguirmos identificar os períodos de fala, quando a média deslizante fosse superior a um determinado limite (10^{-6} para esta base de dados) ocorria fala, conseguindo-se assim evitar as zonas iniciais e finais de silêncio. As zonas de silêncio em fala contínua no meio da frase não eram evitadas, uma vez que, o algoritmo excluía desde o momento inicial até ao primeiro momento de fala, e, desde o momento final até ao último momento de fala.

Com o algoritmo desenvolvido neste trabalho é possível extrair os parâmetros referidos para acrescentar na base de dados curada contruída por (Fernandes, et al., 2018) e assim torná-la mais completa do ponto de vista dos parâmetros disponíveis e do número de doenças que contêm.

7.2. TRABALHOS FUTUROS

Como trabalhos futuros gostaria de sugerir que fossem analisados, através de classificadores inteligentes, o melhor número de coeficientes LPC e LSF a utilizar, de modo, a obter a melhor taxa de precisão para a distinção entre sujeitos doentes e saudáveis e também a distinção entre doenças. Esta análise pode, também, ser feita para os MFCCs e assim saber qual o melhor número de coeficientes mel cepstrais a usar de modo a atingir uma maior taxa de precisão.

Seria também interessante fazer uma análise através de *clusters,* de modo a conseguir perceber quais as doenças que mais se assemelham entre si a nível do sinal de fala e ainda quais os melhores parâmetros para conseguirmos distinguir estes grupos de doenças entre si.

Uma outra proposta seria fazer a mesma pesquisa utilizando outra base de dados, de forma, a perceber se os resultados se mantêm e ainda fazer uma análise comparando bases de dados de línguas diferentes e tentar perceber se a língua interfere nos resultados.

BIBLIOGRAFIA

Alencar, V. & Alcaim, A., 2008. Atributos Eficientes em Reconhecimento Automático de voz Distribuído. *Revista Controle e Automação,* Abril, Maio e Junho, Volume 19 nº2, pp. 147-154.

Almeida, N. C. d., 2010. *Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor de Suporte,* Rio Grande do Norte: s.n.

Alves, N., 2016. *Diagnóstico Inteligente de Patologias da Laringe,* Bragança: s.n.

Anon., 2015. *El fibroma de las cuerdas vocales, el fibroma de la laringe – la causa, los síntomas, la diagnosis y el tratamiento.* [Online] Available at: http://terapiaherbal.com/el-fibroma-de-las-cuerdas-vocales-el-fibroma-de-la-laringe-la-causa-los-s%EF%BF%BD-ntomas-la-diagnosis-y-el-tratamiento/ [Acedido em 13 Abril 2018].

Baena, A., 2013. *Disfonias.* [Online] Available at: http://www.sulms.saudeatual.com.br/especialidades/fonoaudiologia/disfonias [Acedido em 12 Abril 2018].

Baravieira, P. B., 2016. *Aplicação de uma Rede Neuronal Artificial para a Avaliação da Rugosidade e Soprosidade Vocal,* s.l.: s.n.

Bergamini, M., Englert, M., Ribeiro, L. & Azevedo, R., 2015. Estudo de caso: disfonia psicogênica. *CEFAC,* Janeiro/Fevereiro.Volume 17.

Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sample sound. *IFA Proceeding,* Volume 17, pp. 97-110.

Boersma, P., 2004. Stemmen meten met Praat. *Universiteit van Amsterdam.*

Boersma, P. & Weenink, D., s.d. *Praat: doing phonetics by computer.* Amsterdão: Phonetic Sciences, University of Amsterdam.

Castellon, L., s.d. *Trastornos de la voz.* [Online] Available at: http://www.logopedas-castellon.com/patologias/trastornos-de-la-voz_disfonias_disfuncionales.html
[Acedido em 12 Abril 2018].

Cordeiro, H., 2016. *Reconhecimento de Patologias da Voz usando Técnicas de Processamento da Fala ,* Lisboa: s.n.

Costa, C. R. d. O., 2013. *Reconhecimento Robusto de Vogais Isoladas,* Porto: s.n.

Cuf, s.d. *Disfonia Espasmódica.* [Online] Available at: https://www.saudecuf.pt/mais-saude/doencas-a-z/disfonia-espasmodica [Acedido em 12 Abril 2018].

Daliyski, D., 1993. *Acoustic model and evalution of pathological voice production.* Berlim, 3ª Conference on Speech Communication and Technology EUROSPEECH'93.

Dieguez, F. et al., 2010. Granuloma Laríngeo: Relato de caso. *Revista Cientifica da FMC,* Volume 5.

Estibeiro, H. & Trindade, C., s.d. *Tumores da Laringe e Faringe.* Lisboa: Instituto Português de Oncologia de Lisboa Francisco Gentil, E.P.E..

Fabron, E. et al., 2012. Tratamento Médico e Fonoaudiológico da Disfonia Espasmódica: Uma Revisão Bibiográfica. *CEFAC.*

Fernandes, J., Teixeira, F., Odete, P. & Teixeira, J. P., 2018. *Cured Dabtabase of Speech Parameters for Chronic Laryngitis Pathology.* Milão, IBIMA.

Fonseca, E. S. & Pereira, J. C., 2009. Normal Versus Pathological Voice Signals. *IEEE Engineering in Medicine and Biology Magazine*.

Forero, L., Kohler, M., Vellaasco, M. & Cataldo, E., 2016. Analysis and Classification of Voice Pathologies Using Glttal Signal Parameters. *Journal of Voice*.

Gonçalves, A. A., 2015. *Patologias da Laringe com Análise Acústica Vocal,* Intituto Politécnico de Bragança: Tese de Mestrado.

Guimarães, I., 2004. Os Problemas de Voz nos Professores: Prevalência, Causas, Efeitos e Formas de Prevenção. Volume 22.

Henríquez, P. et al., 2009. Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics. *IEEE Transactions on Audio, Speech, and Language Processing*, Agosto.

Logan, B., 2000. Mel Frequency Cepstral Coefficients for Music Modeling. *Cambrige Research Laboratory*.

Lopes, J. et al., 2008. *A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa.* S. Mamede de Infesta, SEEGNAL.

Martins, R. & Dias, N., s.d. *Complicações das vias aéreas relacionadas à intubação endotraqueal*. Botucatu: Faculdade de Medicina de Botucatu - Unesp.

Martins, R. et al., 2009. Edema de Reinke: estudo da imunoexpressão da fibronectina, da lamina e do colágeno IV em 60 casos por meio de técnicas imunoistoquímicas. *SCIELO,* Novembro/Dezembro.Volume 75.

MathWorks, 2006. *LPC.* [Online] Available at: https://www.mathworks.com/help/signal/ref/lpc.html [Acedido em 5 Maio 2018].

MathWorks, 2006. *poly2lsf.* [Online] Available at: https://www.mathworks.com/help/signal/ref/poly2lsf.html#d120e128903 [Acedido em 5 Maio 2018].

MathWorks, 2011. *HTK MFCC MATLAB.* [Online] Available at: https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab
[Acedido em 7 Novembro 2017].

Matuck, G. R., 2005. *Processamento de Sinais De Voz Padrões Comportamentais por Redes Neurais Artificiais,* São José dos Campos: s.n.

Medipédia, 2012. *MEDIPÉDIA.* [Online] Available at: https://www.medipedia.pt/home/home.php?module=artigoEnc&id=200 [Acedido em 26 Novembro 2017].

Merdau, D., s.d. *Pólipo nas cordas vocais.* [Online] Available at: https://www.infoescola.com/saude/polipos-nas-cordas-vocais/ [Acedido em 14 Abril 2018].

Muda, L., Begam, M. & Elamvazuthi, I., 2010. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. *Journal of Computing*, Março, Volume 2.

Muhammad, G., Alsulaiman, M., Mahmood, A. & Ali, Z., 2011. Automatic Voice Disorder Classification Using Vowel Formants. *IEEE Int. Conf. Multimed. Expo,*.

Murphy, P. J., 1999. Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis. *The Journal of the Acoustical Society of America*.

Navarra, C. U. d., 2015. *Disfonia hiperfuncional.* [Online] Available at: https://www.cun.es/diccionario-medico/terminos/disfonia-hiperfuncional [Acedido em 12 Abril 2018].

Navarra, C. U. d., 2015. *Disfonía hipofuncional.* [Online] Available at: https://www.cun.es/diccionario-medico/terminos/disfonia-hipofuncional [Acedido em 12 Abril 2018].

Ogawa, H., Mashima, K. & Ebihara, S., 1986. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *The Journal of the Acoustical Society of America*, Volume 80, pp. 1329-1334.

Oncoguia, E., 2018. *A Laringe e a Hipofaringe.* [Online] Available at: http://www.oncoguia.org.br/conteudo/a-laringe/671/134/ [Acedido em 14 Abril 2018].

Oncoguia, E., 2018. *Câncer de Laringe e Hipofaringe.* [Online] Available at: http://www.oncoguia.org.br/conteudo/sobre-o-cancer/672/134/ [Acedido em 13 Abril 2018].

Panek, D., Skalski, A., Gajda, J. & Tadeusiewicz, R., 2015. Acoustic Analysis Assement in Speech Pathology Detection. *Int. J. Appl. Math. Comput. Sci*.

Rapoport, A. et al., 2011. Revisão do paradigma terapêutico do câncer da hipofaringe. *Rev. Bras. Cir. Cabeça Pescoço,* Volume 40.

Santiago, R., s.d. *Disfonia Psicogenica.* [Online] Available at: https://www.academia.edu/18570060/Disfonia_Psicogenica_concluido [Acedido em 13 Abril 2018].

Sathler, C., 2008. *Linear Prediction: Audio Aplications.* Universidade Federal do Paraná: Departamento de Engenharia Elétrica - Campus Centro Politécnico - Processamento Digital de Sinais I.

ServiceS, U. S. D. o. H. &. H., 2011. *National Institute on Deafness and Other Communication Disorders (NIDCD).* [Online] Available at: https://www.nidcd.nih.gov/health/vocal-fold-paralysis [Acedido em 26 Novembro 2017].

Shama, K. & Cholayya, A. K. N., 2007. Study of harmonics-to-noise ratio and critical-band energy spectrum os speech as acoustic indicators of laryngeal and voice pathology. *EURASIP Journal on Applied Signal Processing*, 1 Janeiro.

Silva, D., Oliveira, L. C. & Andrea, M., 2009. Jitter Estimation Algorithms for Detection of Pathological Voices. *EURASIP Journal on Advances in Signal Processing*.

Silva, M. J. G. d., 2014. *Cancro da Laringe.* [Online] Available at: http://www.atlasdasaude.pt/publico/content/cancro-da-laringe [Acedido em 12 Abril 2018].

Teixeira, J. P., 1995. *Modelização Paramétrica de Sinais para Aplicação em Sistemas de Conversão Texto-Fala,* Faculdade de Engenharia - Universidade do Porto: Dissertação de Mestrado.

Teixeira, J. P. & Fernandes, P., 2015. Acoustic Analysis of Vocal Dysphonia. *Procedia Computer Science - ELSEVIER*.

Teixeira, J. P. & Fernandes, P. O., 2014. *Jitter, Shimmer and HNR classification within gender, tones and vowels in healty voices.* s.l., Procedia Technology.

Teixeira, J. P., Ferreira, D. B. & Carneiro, S. M., 2011. Análise Acústica Vocal - Determinação do Jitter e Shimmer para Diagnóstico de Patologias da Fala.

Teixeira, J. P. & Gonçalves, A., 2014. *Accuracy os Jitter and Shimmer Measurements.* s.l., Procedia Tecnology.

Teixeira, J. P., Oliveira, C. & Lopes, C., 2013. *Vocal Acoustic Analysis - Jitter, Shimmer and HNR Parameters.* s.l., Procedia Technology.

Teixeira, J. P., Teixeira, F., Fernandes, J. & Fernandes, P. O., 2018. Acoustic Analysis of Chronic Laryngitis - Satistical Analysis of Sustained Speech Parameters.

Valiullina, S., s.d. *Cisto na garganta - sintomas, tratamento e complicações.* [Online] Available at: http://pt.surgeon-live.com/content_kista-v-gorle-simptomy-lechenie-i-oslozhnenie.htm
[Acedido em 12 Abril 2018].

Vieira, V. et al., 2016. *Avaliação de Desempenho na Classificação de Patologias Laríngeas por Análise LPC de Sinais de Voz e Redes Neuronais MLP*. s.l., Congresso Brasileiro de Inteligência Computacional.

Virtual, P., 2010. *Granuloma.* [Online] Available at: http://patologiavirtual123.blogspot.pt/2010/04/granuloma-definicao-granuloma-e-uma.html
[Acedido em 13 Abril 2018].

Yumoto, E. & Gould, W. J., 1982. Harmonics-to-noite ratio as an index of the degree of hoarseness. *The Journal of the Acoustical Society of America*, Junho.