

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE ELETROTÉCNICA
CURSO DE ENGENHARIA ELÉTRICA

JULIANA FERNANDA FERREIRA SANTOS

**RECONHECIMENTO DE PADRÕES EM MEDIDAS ACÚSTICAS
PARA IDENTIFICAÇÃO DE PATOLOGIAS DA LARINGE**

TRABALHO DE CONCLUSÃO DE CURSO

**CURITIBA
2017**

JULIANA FERNANDA FERREIRA SANTOS

**RECONHECIMENTO DE PADRÕES EM MEDIDAS ACÚSTICAS
PARA IDENTIFICAÇÃO DE PATOLOGIAS DA LARINGE**

Trabalho de Conclusão de Curso de graduação, apresentado à disciplina de Trabalho de Conclusão de Curso I, do curso de Engenharia Elétrica, do Departamento Acadêmico de Eletrotécnica (DAELT) da Universidade Tecnológica Federal do Paraná (UTFPR), como requisito parcial para obtenção do título de Engenheira Eletricista.

Orientador: Prof. Dr. Marcelo de Oliveira Rosa

CURITIBA

2017

Juliana Fernanda Ferreira Santos

RECONHECIMENTO DE PADRÕES EM MEDIDAS ACÚSTICAS PARA IDENTIFICAÇÃO DE PATOLOGIAS DA LARINGE

Este Trabalho de Conclusão de Curso de Graduação foi julgado e aprovado como requisito parcial para a obtenção do Título de Engenheira Eletricista, do curso de Engenharia Elétrica do Departamento Acadêmico de Eletrotécnica (DAELT) da Universidade Tecnológica Federal do Paraná (UTFPR).

Curitiba, 23 de junho de 2017.

Prof. Emerson Rigoni, Dr.
Coordenador de Curso
Engenharia Elétrica

Profa. Annemarle Gehrke Castagna, Mestre
Responsável pelos Trabalhos de Conclusão de Curso
de Engenharia Elétrica do DAELT

ORIENTAÇÃO

Marcelo de Oliveira Rosa, Dr.
Universidade Tecnológica Federal do Paraná
Orientador

BANCA EXAMINADORA

Antonio Carlos Pinho, Dr.
Universidade Tecnológica Federal do Paraná

Ana Paula Dassi-Leite, Dra.
Universidade Estadual do Centro-Oeste

A folha de aprovação assinada encontra-se na Coordenação do Curso de Engenharia Elétrica.

À memória de Pureza de Oliveira Ferreira, que sempre me incentivou a estudar.

À Maria Rosalina, Sebastião e João Paulo, por me ajudarem a perseverar frente às dificuldades.

E ao Senhor Deus, por tudo que Ele significa para mim.

AGRADECIMENTOS

Agradeço ao meu professor orientador, Professor Marcelo de Oliveira Rosa, pela sua disposição, ensinamentos e motivação, pois foram determinantes para o desenvolvimento deste trabalho. Agradeço também aos professores da banca, Professor Antonio Carlos Pinho e Professora Ana Paula Dassie-Leite, pelos apontamentos e considerações que melhoraram a qualidade do trabalho.

Ainda, agradeço à todos aqueles que de algum modo contribuíram na construção deste trabalho, seja de maneira direta ou indireta, principalmente a meus familiares e amigos pelo apoio e incentivo.

RESUMO

SANTOS, Juliana F. Reconhecimento de Padrões em medidas acústicas para identificação de patologias da Laringe. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica), Universidade Tecnológica Federal do Paraná, Curitiba, 2016.

O diagnóstico de patologias da laringe tem sido realizado predominantemente de forma qualitativa, sendo obtido através de um exame invasivo e desconfortável ao paciente. Através de técnicas de processamento digital de sinais (aplicadas à vogal sustentada /a), medidas ditas acústicas como *jitter* (ou perturbação da frequência fundamental da voz), *shimmer* (perturbação da amplitude do sinal de voz ao longo do tempo de vocalização) e relação entre energia harmônica e não harmônica podem ser obtidas para se quantificar o funcionamento anormal da laringe a fim de se poder obter resultados mais precisos. A partir da avaliação estatística dessas medidas, empregando-se o uso de redes neurais artificiais, pode-se, então, definir um modelo de identificação das patologias que afetam as estruturas da laringe.

Palavras-chave: processamento digital de sinais. análise acústica da voz. diagnóstico de patologias da laringe. aplicação de redes neurais artificiais.

ABSTRACT

SANTOS, Juliana F. Patterns Recognition in Acoustic Measures for Identification of Laryngeal Pathologies. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica), Universidade Tecnológica Federal do Paraná, Curitiba, 2016.

The diagnosis of laryngeal pathologies has been performed predominantly in a qualitative way, being obtained through an invasive and uncomfortable examination to the patient. Through digital signal processing techniques (applied to the sustained vowel / a /), so-called acoustic measures such as jitter (or disturbance of the fundamental frequency of the voice), shimmer (disturbance of the amplitude of the voice signal throughout vocalization time) and the relation between harmonic and non-harmonic energy can be obtained to quantify the abnormal functioning of the larynx in order to obtain more precise results. From the statistical evaluation of these measures, using the artificial neural networks, can, then, define a model for the identification of the pathologies that affect the structures of the larynx.

Keywords: Digital signal processing. Acoustical analysis of voice. Diagnosis of laryngeal pathologies. Application of artificial neural networks.

LISTA DE FIGURAS

Figura 1: Ilustração da laringe: vista anterior e lateral.....	21
Figura 2: Ilustração das estruturas da laringe, quando fechada e aberta.....	22
Figura 3: Principais cartilagens da laringe: vista superior.....	22
Figura 4: Cavidades glóticas.....	24
Figura 5: Sinal de voz saudável, proveniente do fonema /a/.....	29
Figura 6: Sinal de voz patológica, proveniente do fonema /a/.....	29
Figura 7: Ilustração representando uma trajetória do espaço de fase.....	35
Figura 8: Espaço de fase para um sinal de voz patológica.....	37
Figura 9: Espaço de fase para um sinal de voz saudável.....	37
Figura 10: Representação de um neurônio.....	39
Figura 11: Modelo de neurônio artificial.....	40
Figura 12: configuração da matriz de confusão genérica.....	44
Figura 13: matriz de confusão obtida para terceiro teste utilizando os dados provenientes da combinação dos estimadores YIN e ACF, para a configuração de 50 neurônios e porcentagens de treinamento, teste e validação de 70%, 15% e 15%, respectivamente, para uma configuração equiparada.....	45

LISTA DE TABELAS

Tabela 1: resultado obtido utilizando-se todas os valores de Jitter Absoluto obtidos.....	47
Tabela 2: resultado obtido utilizando-se todos os valores de Jitter Local.....	47
Tabela 3: resultado obtido com uma configuração utilizando todos os valores resultante do uso de todos os estimadores.....	48
Tabela 4: estimadores que apresentaram um melhor desempenho.....	48
Tabela 5: resultado obtido utilizando-se todos os valores de Jitter Local, com quantidade de amostras equilibradas.....	49
Tabela 6: melhora na média obtida para Jitter Local, sendo a Média 1 a melhor média obtida utilizando-se todos os dados, e Média 2 o resultado obtido com uma equiparação das amostras saudáveis e patológicas, ambos os resultados foram discriminados anteriormente.....	49
Tabela 7: melhora na média obtida para configuração utilizando todas as medidas, com equiparação de amostras.....	50
Tabela 8: melhora na média obtida para configuração utilizando todas as medidas, com equiparação de amostras.....	50
Tabela 9: resultado obtido com a utilização de todas as medidas.....	51
Tabela 10: melhores resultados obtidos com todas as medidas de desvio padrão e na sequência a média.....	51
Tabela 11: melhor resultado obtido a partir de combinações, com YIN, ACF e PRAAT.....	52
Tabela 12: melhores resultados obtidos com a utilização de todas as medidas, com amostragem equiparada.....	52
Tabela 13: melhores resultados obtidos com todas as medidas de desvio padrão e na sequência a média, com número de amostras equilibrado.....	53
Tabela 14: melhores resultados obtidos para cada estimador.....	53
Tabela 15: melhor resultado obtido a partir de combinações do estimador YIN e ACF.....	54
Tabela 16: escala de seis níveis de concordância para o coeficiente Kappa.....	55
Tabela 17: amostras utilizadas.....	56
Tabela 18 : resultados obtidos para Jitter Local e Absoluto.....	57
Tabela 19: resultados obtidos utilizando-se todos os valores resultantes de todos os estimadores.....	58
Tabela 20: resultados obtidos para medidas provenientes dos estimadores YIN e PRAAT.....	59
Tabela 21: resultados obtidos para combinações de estimadores.....	59
Tabela 22: resultados obtidos utilizando-se todos os resultados de desvio padrão.....	60
Tabela 23: resultados obtidos utilizando-se todos os resultados das médias.....	60
Tabela 24: resultados obtidos utilizando-se todos os resultados obtidos para todos os estimadores.....	60
Tabela 25: resultado para os valores obtidos com o estimador ACF.....	61

Tabela 26: melhores resultados obtidos com a combinação de estimadores.....	61
---	----

LISTA DE QUADROS

Quadro 1: Classificação das patologias de acordo com sua origem.....	26
Quadro 2: Patologias associadas aos parâmetros perceptuais.....	27
Quadro 3: Número de casos por patologia.....	28
Quadro 4: fórmulas para cálculo de Acurácia, Precisão, Sensibilidade e Especificidade.....	46
Quadro 5: Resultado para cálculo de Acurácia, Precisão, Sensibilidade e Especificidade.....	46
Quadro 6: Matriz fictícia de terceira ordem para cálculo de <i>Kappa</i>	55
Quadro 7: Resolução da matriz fictícia de terceira ordem para cálculo de <i>Kappa</i>	55

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1 TEMA	14
1.1.1 Delimitação do tema	15
1.2 PROBLEMAS E PREMISSAS.....	15
1.3 OBJETIVOS	16
1.3.1 Objetivo Geral.....	16
1.3.2 Objetivos Específicos	17
1.4 JUSTIFICATIVA.....	17
1.5 ESTRUTURA DO TRABALHO	18
2 REFERENCIAL TEÓRICO	19
2.1 COMPOSIÇÃO DO APARATO VOCAL.....	19
2.2 FORMAÇÃO DA VOZ.....	23
2.3 PATOLOGIAS.....	25
2.4 ANÁLISE ACÚSTICA	28
2.4.1 Frequência Fundamental (<i>Pitch</i>)	30
2.4.2 Período Fundamental (T_0).....	30
2.4.3 <i>Jitter</i>	31
2.4.4 <i>Shimmer</i>	31
2.4.5 Quociente de Perturbação da Amplitude (APQ).....	32
2.4.6 Perturbação Média Relativa (RAP)	32
2.4.7 Relação Ruído-Harmônico (HNR)	33
2.5 MEDIDAS DE RECORRÊNCIA.....	33
2.5.1 Sistemas dinâmicos e espaço de fase	34
2.6 CLASSIFICADORES.....	37
2.6.1 Redes Neurais Artificiais (RNA's).....	40
3. RESULTADOS E DISCUSSÕES	42
3.1 RESULTADOS DA REDE NEURAL: DISCRIMINAÇÃO ENTRE VOZ SAUVÁVEL X PATOLÓGICA, PARA VALORES DE JITTER.....	43
3.2 RESULTADOS DA REDE NEURAL: DISCRIMINAÇÃO ENTRE VOZ SAUDÁVEL X PATOLÓGICA, PARA VALORES DE HNR.....	51
3.3 RESULTADOS DA REDE NEURAL: DISCRIMINAÇÃO ENTRE GRUPOS DE PATOLOGIAS.....	53
3.3.1 Grupo Carcinoma x Leucoplasia, e Nódulo x Cisto	55
3.3.2 Grupo Nódulo x Edema de Reinke x Leucoplasia	56
4. CONCLUSÃO.....	63
Referências	65

1. INTRODUÇÃO

O tratamento adequado de qualquer patologia exige um diagnóstico rápido e preciso, e para isso se faz necessário que o especialista tenha um grande conhecimento técnico e experiência. Tendo em vista que o método mais comum de diagnóstico é qualitativo, em que o médico faz uso do tato, visão e audição para identificar quaisquer anormalidades, o diagnóstico pode variar conforme a sensibilidade e experiência do médico especialista. (ROSA, 1998).

Atualmente, os métodos mais comuns de exame diagnóstico de patologias referentes à laringe se baseiam na observação das estruturas do trato vocal, sendo as vozes graduadas de acordo com a percepção auditiva do especialista, além disso, o exame é de um certo desconforto para o paciente, pois se trata de um método invasivo de diagnóstico.

Esta técnica, embora comum, apresenta três problemas: é extremamente dependente da percepção do especialista; não existe uma padronização na graduação dos índices, e devido à complexidade do trato vocal, é possível que diversas estruturas produzam um mesmo som, dificultando a identificação precisa da origem da anomalia. (ROSA, 1998).

Neste sentido, o desenvolvimento de novas tecnologias tem auxiliado os profissionais da saúde a realizarem diagnósticos mais precisos, pois através de informações e dados quantitativos pode-se chegar a uma conclusão mais objetiva. Além disso, ao se fornecer uma ferramenta de auxílio de diagnóstico ao profissional da saúde, há um ganho social intrínseco, pois possibilita uma maior rapidez tanto no diagnóstico quanto no encaminhamento à especialidade correta, propiciando um tratamento correto e rápido, que pode interferir diretamente na qualidade de vida do paciente, e em se tratamento de uma rede pública de saúde, uma maior agilidade no diagnóstico possibilita uma diminuição nas filas de atendimento.

O diagnóstico de patologias da laringe, através do processamento digital de sinais, caracteriza-se como um método não invasivo de diagnóstico (ROSA, 1998), uma vez que o reconhecimento da patologia se dá através do encontro, e posterior análise, das características determinantes presentes no sinal da voz de uma pessoa.

A voz é o resultado de fenômenos fisiológicos que determinam sua emissão acústica, podendo ser classificada basicamente como três tipos de sons: sons vocálicos, que são produzidos através da oscilação das pregas vocais; sons fricativos, produzidos através de uma

construção formada no trato vocal, e sons plosivos, produzidos pelo fechamento completo do trato vocal (MORI, 2005).

Para o diagnóstico de patologias da laringe são usados os sons vocálicos, visto que a maioria das alterações no comportamento vocal é proveniente de alterações nas pregas vocais, as quais estão sujeitas a diversas patologias como pólipos, edemas, úlceras, leucoplasias, entre outras, e sendo muitas vezes necessária uma intervenção rápida, seja por tratamento ou cirurgia.

Na análise dos sons vocálicos se faz necessário o estudo de uma grande variedade de parâmetros, ou medidas acústicas, (RAZERA, 2004), dos quais suas combinações são formadas de modo a fornecer informações técnicas que possibilitem uma melhora da margem de acertos no diagnóstico. Esses parâmetros são compostos basicamente pelos componentes que descrevem um sinal, tais como a amplitude, a frequência e a forma de onda, e sendo assim, qualquer alteração em um ou mais componentes podem ser indicativos de uma patologia.

Uma vez identificados os padrões de sinais de voz, pode-se diferenciar sinais de vozes saudáveis e patológicas, e assim, através de distribuições de probabilidade, identificar, com o menor erro possível, a patologia que aflige o paciente.

1.1 TEMA

O diagnóstico de patologias da laringe tem sido realizado predominantemente de forma qualitativa, sendo obtido através de um exame invasivo e desconfortável ao paciente. Através de técnicas de processamento digital de sinais (aplicadas à vogal sustentada /a), medidas ditas acústicas como *jitter* (ou perturbação da frequência fundamental da voz), *shimmer* (perturbação da amplitude do sinal de voz ao longo do tempo de vocalização) e relação entre energia harmônica e não harmônica podem ser obtidas para se quantificar o funcionamento anormal da laringe a fim de se obter resultados mais precisos.

Neste trabalho se pretende viabilizar um método diagnóstico não invasivo e quantitativo, no qual através de técnicas de processamento digital de sinais, será possível analisar e extrair, de um banco de vozes existente, as medidas acústicas citadas com o intuito de se fazer o reconhecimento de padrões para a identificação das condições da laringe.

1.1.1 Delimitação do tema

Através de técnicas de processamento digital de sinais serão obtidas as medidas acústicas de sinais de fala oriundos da pronúncia sustentada da vogal /a/. Tais medidas serão usadas em Redes Neurais Artificiais (HAYKIN, 2001) e Modelos de Misturas Gaussianas (MCLACHLAN, 2000), que são reconhecedores de padrões, para relacionar as medidas acústicas com as doenças da laringe.

1.2 PROBLEMAS E PREMISSAS

Uma das principais dificuldades enfrentadas para o diagnóstico de possíveis patologias da laringe são os atuais métodos de exame, pois são métodos qualitativos de diagnóstico, ou seja, são métodos que dependem exclusivamente do conhecimento técnico e experiência do médico, fazendo com que haja uma dificuldade em se padronizar resultados e diagnósticos (ROSA, 1998).

Por se tratar de um método invasivo, muitos pacientes adiam a realização do procedimento enquanto podem ou até mesmo enquanto os sintomas permitem, fazendo, assim, que a identificação do problema seja tardia (MINITI, 2003). Além disso, a demora em conseguir agendar um exame de laringe pelo SUS contribui para que a patologia avance enquanto se espera a confirmação do diagnóstico. Isso resulta, muitas vezes, em um prognóstico não tão favorável quanto poderia ser com a identificação da patologia em fase inicial

Outro fator contribuinte na dificuldade do diagnóstico com os métodos utilizados atualmente é a necessidade de uma participação ativa do paciente durante o exame. A avaliação da laringe se dá pela inserção (por meio da boca e garganta) de um aparelho com câmera, para que o médico, por vídeo, possa visualizar os aspectos anatômicos da laringe, bem como sua coloração e funcionamento (ABC.MED.BR, 2013). Ainda com o aparelho inserido na garganta do paciente, através da sustentação de sons vocálicos e pronúncias de frases pré-determinadas, o médico é capaz de observar em tempo real a laringe em funcionamento. O entrave para este procedimento se dá no fato de que este causa

desconforto ao paciente, e mesmo fazendo-se uso de medicamentos para tentar minimizar o incômodo, algumas pessoas são incapazes de prosseguir com o exame.

Somando-se as dificuldades apresentadas, ainda existe a necessidade de que o médico tenha certo nível de experiência, haja vista que o diagnóstico se dá visualmente, demandando uma boa acuidade visual do especialista. Devido à dificuldade em se padronizar diagnósticos qualitativos como este, existe a possibilidade de diferentes médicos chegarem a conclusões distintas ou, ainda, um mesmo médico ter uma opinião diferente se consultado em outro momento.

Tendo em vista os problemas apresentados, inerentes aos processos de identificação de patologias da laringe, estudar-se-á um método não invasivo de diagnóstico, no qual a voz do paciente se torna o principal recurso para a obtenção de informações que possibilitem uma análise das condições da laringe, podendo assim detectar possíveis anomalias que resultem em uma determinação da existência ou não de uma patologia.

Sendo assim, através do estudo dos dados obtidos por meio da análise de vozes provenientes de um banco de dados (apenas aplicada aos fonemas sustentados /a), neste trabalho pretende-se fazer o reconhecimento de padrões de sinais de voz de maneira que seja possível a identificação de patologias através de um método quantitativo, possibilitando, assim, uma maior segurança para o médico especialista fornecer um diagnóstico.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Viabilizar um método diagnóstico de patologias da laringe através de um exame não invasivo, a partir do qual seja possível reconhecer padrões em sinais de voz, de maneira que se possa classificar a voz como patológica ou saudável, bem como identificar a patologia que aflige o paciente.

1.3.2 Objetivos Específicos

- Estudar métodos de cálculo de frequências fundamentais (necessárias para cálculos de algumas medidas acústicas, como *jitter*, e identificação de regiões espectrais contendo energia harmônica e não-harmônica, por exemplo);
- Identificar e caracterizar as medidas acústicas relevantes;
- Estudar métodos de reconhecimento de padrões, tais como Redes Neuras Artificiais;
- Definir classificadores mais adequados à proposta;
- Avaliar a eficiência do método,
- Analisar os resultados obtidos, de acordo com a porcentagem de erros e acertos esperados.

1.4 JUSTIFICATIVA

O resultado deste trabalho visa atender os seguintes aspectos: primeiramente fornecer um método diagnóstico não invasivo que complementaria os tradicionais exames para confirmar se possui ou não uma determinada patologia na laringe, o segundo aspecto fundamental é a rapidez e eficácia diagnóstica que possibilita que o paciente tenha um prognóstico mais favorável ao iniciar o tratamento, sob a supervisão de um profissional de saúde.

Outro ponto relevante para a pesquisa de métodos alternativos aos tradicionais é que, embora muito se estude sobre o tema, vê-se ainda a necessidade de contribuir com o estudo, seja por meio do uso de métodos de classificação mais recentes e de técnicas de processamento digital de sinais que produzam métricas acústicas (processo conhecido como redução de dados) mais significativas para quantificar as diferentes doenças da laringe, de modo a diminuir a taxa de erro (falsos positivos e falsos negativos) na identificação dessas patologias da laringe, dando mais segurança ao profissional de saúde para seu uso na rotina clínica.

A partir dessa integração entre engenharia e medicina, pode-se visar uma melhora na qualidade de vida do médico e do paciente, no qual ambos saem ganhando. O paciente, que

através de métodos eficientes e não agressivos, poderá obter o diagnóstico com maior agilidade, e conseqüentemente dar início antecipado ao tratamento. E assim, possibilitando ao médico uma maior tranquilidade, tanto para a determinação da patologia como para a indicação de tratamento.

Essa associação bem sucedida pode ser observada em diversas áreas da medicina, seja através de equipamentos ou introdução de novas técnicas de diagnóstico. Ao se adentrar em um ambiente hospitalar, é possível observar a presença de máquinas de raios-X, tomógrafos, entre outros aparelhos, que são essenciais para todos os tipos de diagnóstico e tratamento, fazendo com que a engenharia possibilite à medicina atuar de maneira mais rápida, eficaz e confiável.

1.5 ESTRUTURA DO TRABALHO

O presente trabalho, em sua primeira parte, é composto por dois capítulos: sendo o primeiro de introdução, no qual é feita uma apresentação do tema e a delimitação do problema, e o segundo contendo a fundamentação teórica do trabalho, no qual contém uma explicação sobre o trato vocal, a formação da voz e as patologias que farão parte deste estudo, assim como a fundamentação referente às medidas acústicas e aos classificadores que serão usados neste trabalho, sendo eles as Redes Neurais Artificiais.

Na segunda parte do trabalho, pretende-se implementar as medidas acústicas e os algoritmos de frequência fundamental (*Pitch*), obtendo todos os dados necessários para que se possa, então, classificá-los e determinar sua eficiência e percentual de acerto.

2. REFERENCIAL TEÓRICO

Nesta seção serão, primeiramente, apresentados os aspectos da anatomia e fisiologia da laringe, de modo que seja facilitada a compreensão dos elementos envolvidos na produção da fala, bem como indicar a maneira que as alterações nestes elementos, provenientes das diversas patologias, podem influenciar o sinal produzido.

Na sequencia serão apresentadas as principais medidas acústicas que serão utilizadas neste trabalho, e também, para análise das medidas obtidas dos sinais de voz, serão apresentadas os métodos de classificação, sendo estes as Redes Neurais Artificiais e os Modelos de Misturas Gaussianas.

2.1 COMPOSIÇÃO DO APARATO VOCAL

Para se entender os distúrbios de voz provenientes de patologias da laringe, primeiramente se, faz-se necessário entender a estrutura do aparelho fonador, bem como seu funcionamento.

O aparelho fonador é dividido basicamente em três subsistemas: os pulmões, a laringe e o trato vocal, que pode ser observado conforme figura 4. Embora esses três sistemas trabalhem em unidade para a produção de voz, cada qual possui uma função distinta e qualquer variação nestes conjuntos pode resultar em uma alteração do sinal de voz gerado. (MORI, 2005)

O primeiro subsistema é constituído pelo pulmão, juntamente com a traqueia. Os pulmões são órgãos de estruturas esponjosas, possuem formato de pirâmide, cuja base encontra-se apoiada sobre o diafragma. Sua principal função é a hematose, mas também tem participação na regulação de temperatura corporal e na produção de voz. (RAZERA, 2004) Na produção de voz sua participação se dá pelo escoamento do ar com determinada vazão, que é definida pela diferença de pressão entre o ar nos pulmões e a pressão atmosférica fora do corpo. O movimento do pulmão se dá de duas maneiras: pelo movimento do diafragma, que proporciona um aumento ou diminuição da altura da cavidade torácica, e pelo movimento das costelas, que alteram o diâmetro desta cavidade, sendo ambos os movimentos

responsáveis pelo volume de ar existente no pulmão. (MORI, 2005)

A traqueia, por sua vez, funciona como um duto transportador de ar para a laringe. Ela é constituída por tecido músculo-membranoso com anéis cartilagosos, que a confere um formato cilíndrico. (RAZERA, 2004)

O segundo subsistema é formado pela laringe, sendo esta a principal estrutura do sistema fonador. Situada acima da traqueia, a laringe é composta por músculos, cartilagens e por um epitélio de revestimento especial. (JUNQUEIRA, 1999) O conjunto destes elementos dá forma à laringe, de cuja forma, a qualidade vocal depende essencialmente.

A laringe atua como um componente vibrador, o qual é constituído basicamente pelas pregas vocais, que fazem protrução das paredes laterais para o centro da glote, são estriadas e posicionadas por vários músculos específicos da laringe. (MORI, 2005)

Com função de prover a sustentação da laringe, as cartilagens são porções fixas, sendo movimentadas pelos músculos e cartilagens conjuntamente. As principais cartilagens são: a cricóide, a tireóide, o par de aritenóides e a epiglote, como podem ser vistos nas figuras 1 e 2. Ainda existem duas outras, as corniculares, que apesar de quase não possuírem função na fonação, juntamente com as demais cartilagens formam uma espécie de esqueleto do sistema vibratório. (RAZERA, 2004)

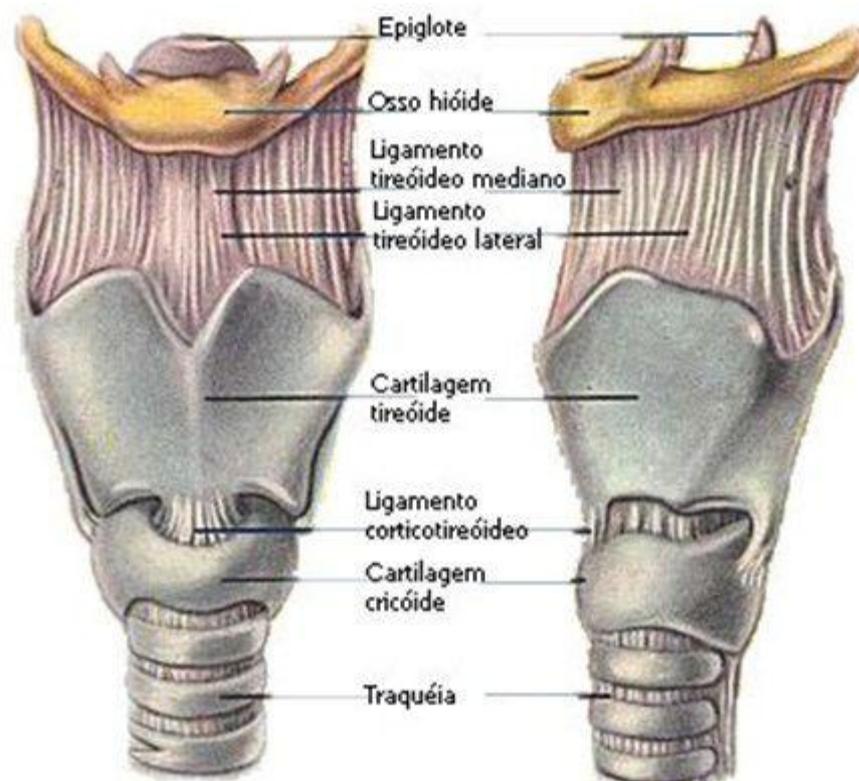


Figura 1: Ilustração da laringe: vista anterior e lateral.

Fonte: CESAR, 2003.

A parte muscular é dividida de acordo com sua posição no sistema de vocalização, sendo caracterizados como dois tipos: extrínsecos e intrínsecos. Os músculos possuem o papel de fazer a rotação ou de tracionar as cartilagens, de modo que as pregas vocais possam ser estiradas, gerando sons agudos, ou afrouxadas, emitindo sons mais baixos. Ainda podem modificar a forma e a massa das bordas das pregas vocais. (MORI, 2005)

Os músculos extrínsecos são aqueles ligados as estruturas externas, e que permitem a elevação ou abaixamento da laringe, além de mantê-la fixa e possuir função auxiliar na produção de voz. A elevação e o abaixamento da laringe aparecem geralmente durante as inflexões da voz falada ou durante o canto, em emissões de notas agudas e graves, e ocorrendo, também, durante e após o processo de deglutição. Alguns distúrbios de voz são caracterizados por uma tendência a manter a posição da laringe acima do normal, provocando um aumento da abertura da glote e uma tensão muscular excessiva. (JUNQUEIRA, 1999)

Os músculos intrínsecos são aqueles ligados somente às estruturas internas da laringe, sendo seis pares e agindo diretamente sobre a fonação. Os cricoaritenóideos posteriores permitem a abdução (abertura) das pregas vocais, enquanto que os laterais produzem uma adução (fechamento) das pregas. As fibras aritenóideas transversas auxiliam os cricoaritenóideos laterais, além de comprimir as pregas vocais. Já os músculos aritenóideas oblíquos no fechamento da glote. Os tiroaritenóideos formam o volume muscular da prega vocal, uma vez que esta compreende a superfície interna da cartilagem aritenóidea e o ligamento vocal, que é um ligamento elástico, sendo recoberto por uma membrana esbranquiçada que vibra com a variação da pressão. A tensão neste músculo é controlada por nossa vontade, podendo ser contraído ou relaxado. (RAZERA, 2004)

Assim sendo, o espaço compreendido entre as bordas das pregas vocais (a fenda glótica), pode ser alargada ou restringida. Ao se fazer a passagem do ar pela glote, o ar provoca vibrações de intensidades diferentes, correspondendo, assim, aos diferentes sons emitidos. (RAZERA, 2004) A altura da vibração é determinada, principalmente, pelo grau de estiramento das pregas, além do grau de aproximação e da massa de suas bordas. (MORI, 2005)

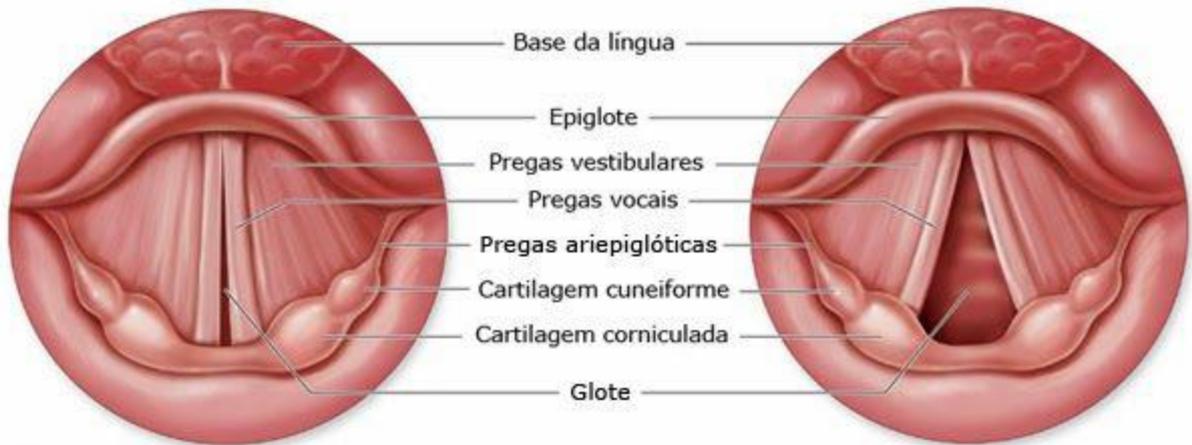


Figura 2: Ilustração das estruturas da laringe, quando fechada e aberta.

Fonte: MELLO, 2013.

No interior de cada prega vocal há o ligamento vocal. Ele é ligado anteriormente à cartilagem tireoidea, fazendo uma protrusão na parte anterior do pescoço, sendo esta estrutura denominada como epiglote. Posteriormente, ele é ligado às cartilagens aritenóideas. Ambas as cartilagens articulam-se com a cartilagem cricóide, como pode ser observado na figura 3.

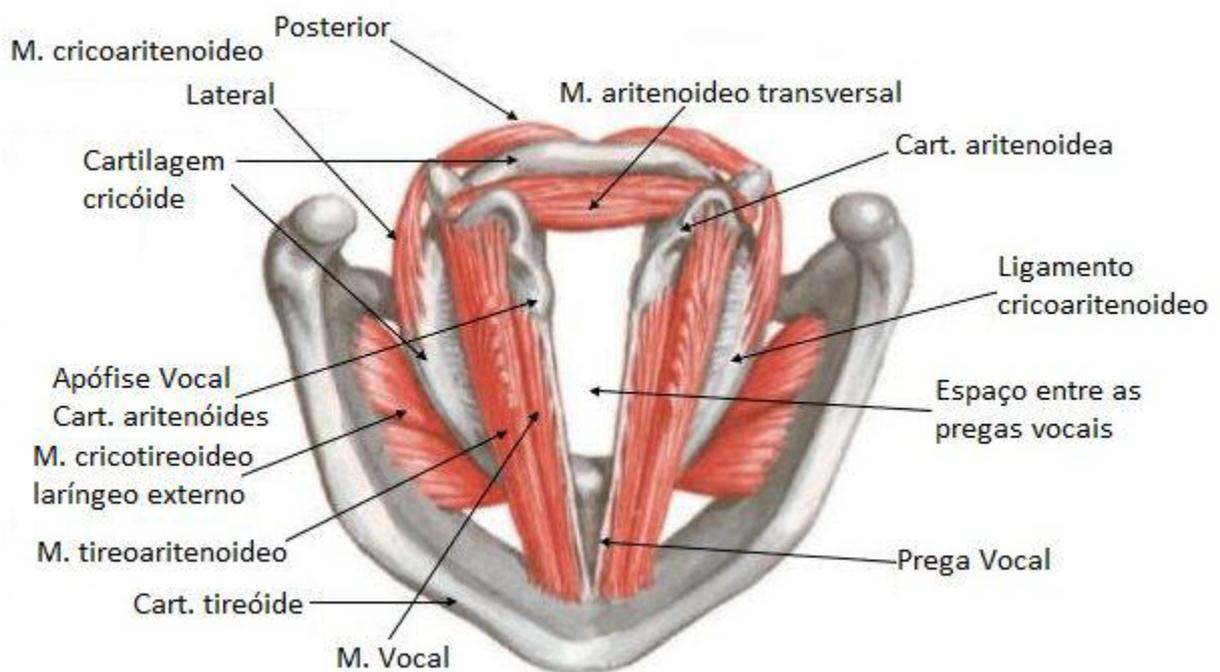


Figura 3: Principais cartilagens da laringe: vista superior.

Fonte :Adaptado de SOUZA, 2004.

O trato vocal corresponde ao terceiro subsistema, e é composto pelo nariz, boca e faringe, compreendendo, assim, a porção acima das pregas vocais. (FUKUYAMA, 2001)

De uma maneira simplificada, pode-se dizer que na produção da fala a fonte emissora do sinal é a vibração da laringe, na qual esta se comporta como um transdutor da energia proveniente do fluxo do ar em acústica, por meio da abertura e fechamento das pregas vocais, e o filtro, ou sistema ressonante, é o trato vocal. (FUKUYAMA, 2001)

O sistema ressonador pode ser representado como um conjunto de tubos de ressonância que enfatizam o sinal glotal, ou seja, que amplificam e atuam componentes espectrais desse sinal, cuja modulação é realizada através da articulação dos lábios, língua e do palato mole. Ao se alterar a forma do trato vocal pode-se alterar o som básico gerado pela laringe, criando, assim, uma enorme variedade de sinais, resultando, por fim, em sons audíveis e inteligíveis ao ouvido humano.

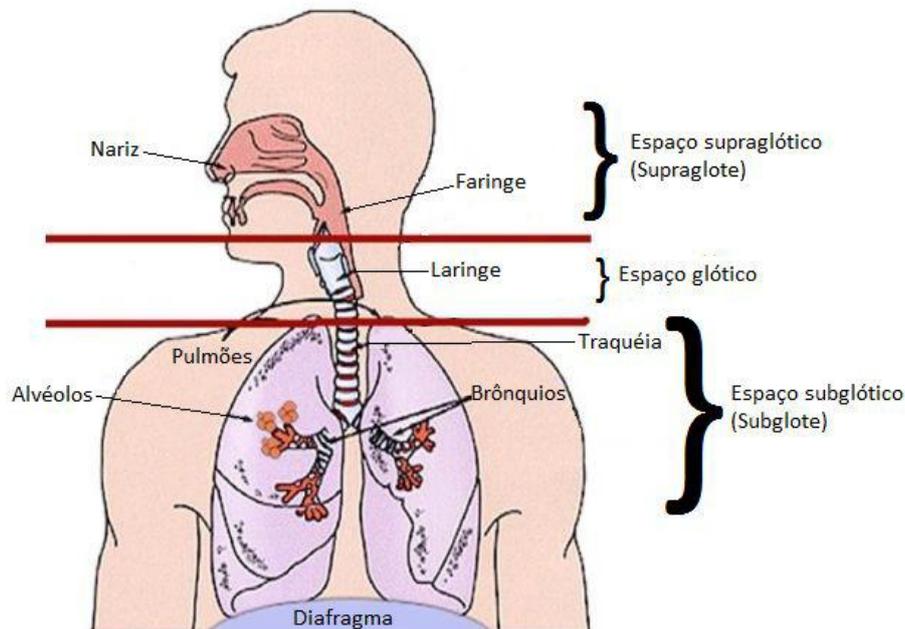
2.2 FORMAÇÃO DA VOZ

Uma maneira simplificada de compreender a formação da voz é como sendo o caminho percorrido pelo ar até ser transformado em voz.

No momento em que há somente a respiração normal e tranquila, sem vocalização, todas as estruturas relacionadas à voz permanecem relaxadas. (RAZERA, 2004) As pregas vocais se encontram separadas, com um formato triangular, de modo a facilitar a passagem do ar.

Para a formação da voz, há uma aproximação destas pregas, através da ação dos músculos intrínsecos, e essa aproximação acaba por gerar um aumento da pressão subglótica (na região abaixo das pregas vocais). Esse aumento de pressão força as pregas a se abrirem, proporcionando a passagem de fluxo do ar proveniente do pulmão. Neste momento, com um pequeno alívio da pressão, as pregas vocais adquirem energia elástica ao se abrirem, sendo que o acúmulo dessa energia propicia que as pregas vençam a pressão do ar ainda existente e se fechem novamente. Este ciclo de abertura e fechamento gera uma complexa vibração, produzindo um sinal quase periódico que se propaga para a cavidade supraglótica (região acima das pregas vocais), que irá atenuar ou amplificar determinadas frequências presentes neste sinal. A disposição das cavidades glóticas pode ser mais bem entendida com a observação da

figura 4.

**Figura 4: Cavidades glóticas.****Fonte: Adaptado de EAV, 2012.**

Relacionado diretamente ao número de ciclos vibratórios que ocorrem por segundo está a frequência fundamental da voz, ou *pitch*. Em geral, se aceita que haja uma faixa de normalidade, ou aceitabilidade, da frequência fundamental, sendo de 80 a 150 Hz para homens, de 150 a 250 Hz para mulheres, e acima de 250 Hz para crianças.

O som, então produzido pelo aparelho fonador, pode ser dividido, principalmente, em três categorias:

a) Sons vocálicos:

São produzidos por pulsos de ar quase periódicos, gerado na laringe pelo movimento cíclico das pregas vocais. São identificados principalmente pela emissão das vogais.

b) Sons fricativos:

São produtos da passagem turbulenta do ar através de alguma constrição formada no trato vocal. São caracterizados como /r/ e /s/.

c) Sons plosivos:

São produzidos pelo fechamento completo do trato vocal, com um aumento da pressão e uma liberação abrupta desta. São exemplos disso: /d/, /t/, /p/ e /b/.

Na produção dos sons vocálicos, há uma vibração das pregas vocais e ressonância do trato vocal completo, portanto, este tipo de som é o mais adequado para estudo e análise do

trato vocal, uma vez que o som produzido possui toda informação do sistema em questão. (RAZERA, 2004)

2.3 PATOLOGIAS

Em termos gerais pode-se dizer que uma “patologia” é qualquer alteração ou desvio do padrão comum e esperado, podendo estas alterações serem estruturais, bioquímicas e funcionais, nas células, tecidos e/ou órgãos, dando, assim, origem aos sinais e sintomas das doenças.

Partindo de uma forma mais abrangente, a patologia da laringe pode ser considerada como sendo qualquer alteração do sistema produtor de voz, cujo resultado não seja satisfatório ao paciente.

A maioria das patologias da laringe, ou disfonias, causam mudanças significativas no processo de produção da fala, uma vez que geralmente produzem modificações assimétricas nas pregas vocais, resultando em diferentes tipos de vibrações, de maneira que o sinal resultante proveniente da fala seja afetado pela patologia.

Essas patologias podem ocorrer em qualquer um dos três subsistemas apresentados anteriormente (pulmões, laringe e trato vocal), sendo classificados convencionalmente quanto à origem, ou seja, de acordo com a etiologia do sistema. Assim, pode-se considerar três tipos de disfonias: funcional, organo-funcional e orgânica.

As disfonias de origem funcional são causadas pelo uso abusivo da voz e são caracterizadas por alterações na emissão da voz. Também podem ser decorrentes de ineficiências vocais e alterações psicogênicas. Já as disfonias de origem orgânica são aquelas provenientes de anomalias físicas no sistema de produção da voz. As disfonias organo-funcionais também são causadas por problemas funcionais (mal-uso), mas apresentam lesões secundárias, lesões estas geralmente relacionadas a diagnósticos tardios de disfonias funcionais. Segundo alguns autores, há ainda as disfonias psicogênicas, embora relativamente raras de se encontrar, são aquelas causadas por motivos psicológicos, não tendo razão estrutural para se originar e estando relacionada a uma tensão muscular. O quadro 1 apresenta a classificação da patologia de acordo com sua origem.

ORIGEM		PATOLOGIAS ASSOCIADAS
ÔRGANICA	Estutural	Úlcera de contato Cistos Granuloma Hemorragia Hiperqueratose Laringite Nódulos Papiloma Pólipos Trauma
	Neurogênica	Paralisia Paresia Disfonia Espasmódica Tremor Causadas por patologias neurológicas
FUNCIONAL		Disfonia de tensão muscular Estiramento Anterior – Posterior Hiperfunções Constrição Faríngea
PSICOGÊNICA		Fonação ventricular
		Disfonia de conversão ou Afonía Puberfonia

Quadro 1: Classificação das patologias de acordo com sua origem.
Fonte: SODRÉ, 2016.

Acusticamente, as patologias são diferenciadas em relação a padrões perceptuais como Rouquidão, Aspreza e Soprosidade, e às escalas destas características. Vozes ásperas são aquelas causadas pelo estiramento das pregas vocais, causando variação na frequência fundamental e amplitude do sinal glotal. Vozes soprosas não apresentam fechamento total da glote, produzindo ruído espectral em todas as frequências, principalmente nas altas. Vozes roucas apresentam uma modificação (oscilação) da frequência fundamental das pregas vocais.

A voz classificada como normal pode ser entendida como uma voz agradável, melódica, audível e coerente, que possibilita uma boa comunicação e permite ao indivíduo ajustar-se a situações sociais e diferentes tipos de ambiente.

Existe ainda a diferenciação dos padrões perceptuais referentes a Astenia, Tensão e Instabilidade. A astenia corresponde à hipofunção das pregas vocais, aliada a pouca energia na geração da voz. A tensão pode ser associada a um esforço vocal gerado pelo aumento da adução glótica. E a instabilidade refere-se ao tremor das estruturas vocais.

Essas características são graduadas em uma escala de 0, que significa normal, a 3, que é a característica extrema, do ponto de vista da escala GRBAS, ou RASAT. (SODRÉ, 2016)

Sendo essa graduação e caracterização da voz do paciente atrelada ao grau de percepção e experiência do profissional da saúde, não há uma clareza quanto à natureza exata da queixa do paciente, uma vez que cada profissional possa vir a ter uma percepção diferente. De um modo geral, o quadro 2 apresenta uma associação entre parâmetros perceptuais e suas definições.

Parâmetros	Definição
<i>G – Grade</i>	Grau de alteração vocal – impressão global da voz, identificação do grau de alteração vocal, como um todo
<i>I – Instability</i>	Instabilidade, flutuação na F0 e/ou na qualidade vocal
<i>R – Roughness</i>	Irregularidade na vibração das pregas vocais, sensação de “rugosidade” (<i>raucité</i> em Francês) que corresponde a flutuações irregulares da F0 e/ou da amplitude do som glótico
<i>B – Breathiness</i>	Soprosidade, relativo a turbulência audível, escape de ar importante através da glote, sensação de ar na voz
<i>A – Asteny</i>	Astenia, fraqueza vocal, falta/perda de potência na voz relacionada com intensidade fraca, energia vocal reduzida, harmônicos pouco definidos e pouco elevados
<i>S – Strain</i>	Tensão, impressão de estado hiperfuncional, frequência anormalmente aguda (ruído nas frequências altas do espectro e/ou riqueza de harmônicos agudos marcados)

Quadro 2: Patologias associadas aos parâmetros perceptuais.
Fonte: (BEHLAU, 2001).

No quadro 3 são apresentadas as principais patologias que farão parte deste trabalho, bem como o número de vezes em que ela foi observada. Pode-se perceber que algumas patologias apresentam elevado índice de aparecimento, isso se dá ao fato de que uma mesma pessoa pode ser acometida por diversas patologias, e sendo algumas delas, como a hiperfunção, percebida na maioria dos casos.

PATOLOGIAS	NÚMERO DE CASOS
Cistos	10
Carcinoma	2
Disartria	1
Disfonia Espasmódica	24
Edema de Reinke	33
Eritema	20
Granuloma	15
Hiperkeratose	1
Hiperfunção	289
Laringite	4
Leucoplasia	30
Nódulos Vocais	23
Papiloma	4
Paralisia	79
Paresia	22
Pólipo	23
Presbifonia	14
Refluxo Gástrico	53
Tremor Vocal	18

Quadro 3: Número de casos por patologia.
Fonte: Autoria própria.

2.4 ANÁLISE ACÚSTICA

Uma forma de complementar de modo não invasivo a análise perceptivo-auditiva dos profissionais de saúde é através de técnicas de análise acústica. Essa análise baseia-se na ideia de que o sinal produzido é um reflexo das perturbações da laringe, podendo, então, ser usado para medir esses distúrbios, tendo em vista que as irregularidades se propagam pela faringe, cavidades nasais e boca, influenciando, assim, o sinal resultante. Isso pode ser exemplificado pelas figuras 5 e 6, que mostram a variação do sinal emitido para um mesmo fonema, sendo a

figura 5 o sinal proveniente de uma voz saudável, na qual se pode perceber a periodicidade da onda, e a figura 6 sendo o sinal proveniente de uma voz acometida por diversas patologias (laringite, hiperfunção e alteração de massa das pregas vocais), na qual, embora possa ser percebida certa periodicidade, notam-se muitas variações no sinal.

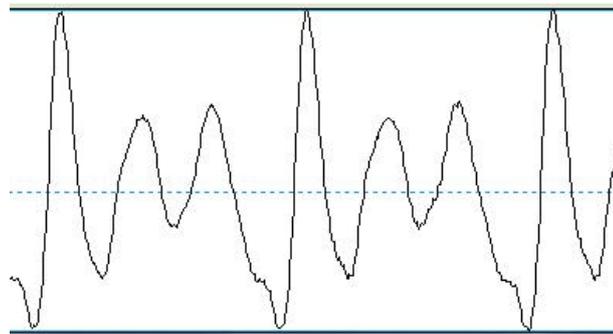


Figura 5: Sinal de voz saudável, proveniente do fonema /a/.

Fonte: o autor.

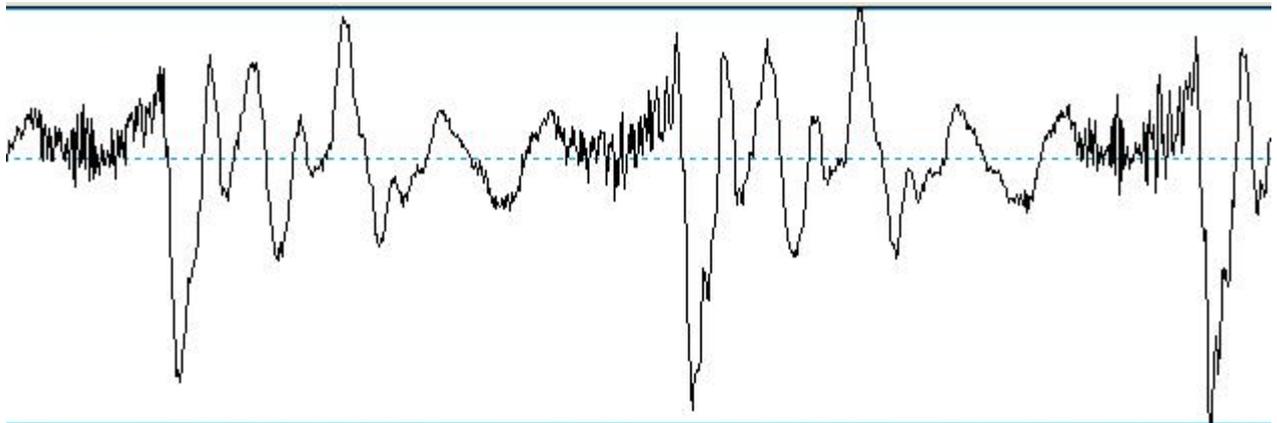


Figura 6: Sinal de voz patológica, proveniente do fonema /a/.

Fonte: o autor.

Essas técnicas baseiam-se na extração de parâmetros dos sinais de voz ou espectro de potência. Esses parâmetros são obtidos para sinais de pessoas com laringes saudáveis e patológicas, então, os valores são confrontados e testes estatísticos realizados para verificar a separação dos grupos.

Os sintomas acústicos mais comuns causados por patologias da laringe são mudança na frequência fundamental da voz, intensidade vocal e qualidade vocal. Sendo assim os parâmetros mais usados para a análise são *jitter*, *shimmer* e medidas de ruído, além do espectro de potência do sinal.

As medidas acústicas devem ser obtidas a partir de um fonema sustentado, como /a/,

/e/ ou /i/, por exemplo, para que seja possível a obtenção de uma frequência fundamental, cuja aquisição não seria viável a partir de uma frase, pois haveria sobreposição de várias frequências fundamentais provenientes dos diferentes sons.

Existem diversas formas de quantificar as medidas acústicas, a seguir serão apresentadas algumas dessas maneiras.

2.4.1 Frequência Fundamental (*Pitch*)

A frequência fundamental, também chamada de *Pitch*, é representado por f_0 , possui relação com a massa, a tensão (elasticidade) e o comprimento das pregas vocais, além da pressão subglótica e pela configuração do trato vocal. Corresponde ao número de ciclos vibratórios das pregas vocais por segundo, em que cada ciclo vibratório corresponde ao ciclo de abertura e fechamento das pregas vocais, e é medido em *Hertz* (Hz).

Sua obtenção se dá através da demarcação de cada ciclo glótico, onde se faz a comparação da forma de onda de dois seguimentos adjacentes e de mesmo comprimento. Dentre os métodos para sua demarcação estão o coeficiente de correlação entre os trechos analisados, o espaço de fase e *Time-Event Rate Detection*, entre outros. (GERHARD, 2003)

Existe uma faixa típica de valores de f_0 de acordo com a idade e o sexo de cada indivíduo. Para homens o valor é geralmente menor que 150 Hz, tendo como valores médios próximos de 100 Hz. Para mulheres, esses valores variam de 150 a 250 Hz, tendo os valores médios em torno de 200 Hz. Já para crianças, os valores médios se encontram acima de 250 Hz.

2.4.2 Período Fundamental (T_0)

Representando por T_0 , o período fundamental é o inverso da frequência fundamental e é geralmente medido em milissegundos (ms).

$$T_0 = \frac{1}{f_0} \quad (1)$$

2.4.3 Jitter

É uma medida de dispersão (perturbação) do período, ou seja, é uma medida que procura quantificar as variações ciclo a ciclo de cada período de um sinal de voz. De outra forma, o *jitter* pode ser definido como a média das variações de cada período de um sinal de voz.

Alterações do *jitter* estão relacionadas a presença de lesões de massa ou doenças neurológicas e psicológicas, sendo associado ao aspecto perceptivo de aspereza.

Comparando-se os valores obtidos para os sinais saudáveis, estes eram menores que para os sinais patológicos.

Essa medida pode ser obtida através das equações 2 e 3, que podem ser observadas abaixo, sendo N o número total de amostras; T_i o tempo no instante i , ou seja, o período da amostra f_0 extraída, e T_{i+1} o tempo no instante $i+1$. (SODRÉ, 2016)

A primeira, que representa o *Jitter Local*, que é a diferença média absoluta entre dois períodos consecutivos divididos pelo período médio, e a segunda, o *Jitter Absoluto*, sendo este a diferença média absoluta entre dois períodos consecutivos

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

$$Jitter(absolute) = \frac{1}{N-1} \sum_{i=1}^{N-1} (|T_i - T_{i+1}|) \quad (3)$$

2.4.4 Shimmer

É uma medida de dispersão da amplitude. De forma semelhante ao *jitter*, o *shimmer* é definido como a média das variações de amplitude de sucessivos períodos de *pitch* de um sinal. Ou seja, pode-se entender o *shimmer* como uma medida das variações ciclo-a-ciclo na intensidade da excitação glótica. Essa medida é dada em decibéis (dB).

As alterações de *shimmer* ocorrem, principalmente, por patologias que apresentam lesões de massa nas pregas vocais, como nódulos, pólipos, edemas e carcinomas. Assim, o aspecto perceptivo de rouquidão está relacionado com essa medida.

Essa medida pode ser observada através da equação 4, na qual A_i corresponde a amplitude pico-a-pico, e N corresponde ao número total de amostras, ou seja, o número total de períodos extraídos da amostra. (SODRÉ, 2016)

$$Shimmer (dB) = \frac{1}{N-1} * \sum_{i=1}^{N-1} 20 \log \left| \frac{A_{i+1}}{A_i} \right| \quad (4)$$

2.4.5 Quociente de Perturbação da Amplitude (APQ)

Essa medida procura avaliar a variação da amplitude pico a pico no interior da amostra da voz analisada com um fator de amortecimento de 11 períodos. Os valores obtidos são expressos em porcentagem.

Este parâmetro pode ser expresso através da equação 5. (SODRÉ, 2016)

$$Shimmer(apq 11) = \frac{\left(\frac{1}{N} - 10\right) \sum_{i=6}^{N-5} |A_i| - \left(\sum_{k=i-5}^{i+5} \frac{A_k}{11}\right)}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (5)$$

2.4.6 Perturbação Média Relativa (RAP)

Este parâmetro mede a flutuação de *pitch*, assim sendo, ele é a diferença média absoluta entre um período e a média dele e de seus dois vizinhos (anterior e posterior), dividido pelo período médio. Os valores obtidos dessa medida para vozes patológicas são diferentes dos valores obtidos para vozes saudáveis, especialmente para pacientes com neoplasmas e paralisia unilateral. Esse parâmetro é medido em Hz.

Essa medida pode ser obtida através da equação 6, sendo que T corresponde ao período obtido e N o número de amostras. (SODRÉ, 2016)

$$RAP = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| \left(\frac{T_{i-1} + T_i + T_{i+1}}{3} \right) - T_i \right|}{\frac{1}{N} \sum_{i=1}^{N-1} T_i} \quad (6)$$

2.4.7 Relação Ruído-Harmônico (HNR)

É uma medida que relaciona as componentes harmônicas com o ruído, isso é realizado pela observação da quantidade de estrutura harmônica do espectro que é substituída por ruído. Pode-se dizer, de um modo simples, que a relação ruído-harmônico é a divisão entre a medida da energia das componentes harmônicas de $f(t)$ pela energia das componentes de ruído de $f(t)$.

Em outras palavras, essa medida pode ser entendida como a relação entre um componente periódico, como a vibração das pregas vocais, e um componente aperiódico, como o ruído glótico. A relação entre esses dois componentes traduz a eficiência do processo de fonação, pois quanto mais íntegra a vibração, menor o ruído e mais eficiente é este processo. A rouquidão pode ser relacionada com esta medida. (SODRÉ, 2016)

Essa medida pode ser obtida através da equação 7.

$$HNR = \frac{N \int_0^T f_A^2(\tau) d\tau}{\sum_{i=1}^N \int_0^{T_i} [f_i(\tau) - f_A(\tau)]^2 d\tau} \quad (7)$$

2.5 MEDIDAS DE RECORRÊNCIA

As técnicas de análise de dados são cada vez mais utilizadas como ferramentas para o trabalho e compreensão de sistemas dinâmicos. Métodos tais como o cálculo da dimensão fractal, do expoente de *Lyapunov* e a entropia (COSTA, 2014), embora sejam amplamente utilizados, encontram dificuldades em sua execução, pois a maioria dessas ferramentas necessita de análise de longas séries temporais, e que acabam sendo afetadas, principalmente, pela presença de ruídos nos sinais e pela dificuldade de processamento de longas séries.

A recorrência é uma propriedade de muitos sistemas dinâmicos. Segundo o Teorema da Recorrência (VIEIRA, 2014), para um conjunto grande de sistemas dinâmicos, as trajetórias retornam infinitas vezes, arbitrariamente próximas, a quase todos os pontos iniciais, formando um conjunto infinito de instantes de retorno, ou seja, dado um ponto qualquer, tantas trajetórias que dali saírem, retornarão a um ponto próximo ao ponto de saída.

A análise da quantificação de recorrência é constituída de um método de análise subjetiva, a qual se dá através dos gráficos de recorrência, e de um método de análise objetiva, que é a quantificação da recorrência.

2.5.1 Sistemas dinâmicos e espaço de fase

Um sistema é um conjunto de componentes interconectados que apresentam certas relações de causa e efeito, e que atuam como um todo, com um determinado objetivo. Um sistema dinâmico é aquele em que as propriedades descritivas do sistema, seu estado, variam no tempo, podendo também variar espacialmente. Assim, o sistema dinâmico pode ser definido por um modelo matemático, podendo ser empregado em sistemas biológicos, físicos, econômicos, entre outros.

A quantidade de variáveis de estado corresponde à dimensão do sistema dinâmico, ou seja, ao número de graus de liberdade do sistema. A partir das variáveis do estado, pode-se investigar o comportamento da dinâmica dos sistemas, por meio do chamado espaço de fase.

Espaço de fase é um espaço abstrato que representa a evolução do sistema dinâmico, cujas dimensões são as variáveis de estado. Cada estado do sistema, que pode ser entendido como sendo cada solução do modelo matemático, representa um ponto neste espaço, e sendo a trajetória do espaço de fase formado pela sequência dos estados, como pode ser exemplificado pela figura 7.

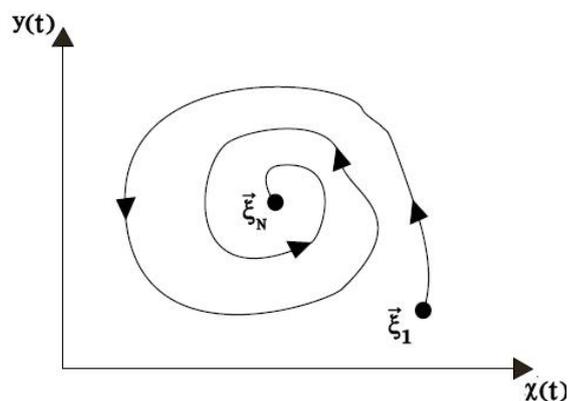


Figura 7: Ilustração representando uma trajetória do espaço de fase.

Fonte: VIEIRA, 2014.

Por vezes, algumas variáveis do sistema podem não ser observáveis, e no caso de séries temporais, tais como os sinais de voz, a evolução no tempo é percebida através de apenas uma variável de estado.

Assim, para a análise do comportamento dos sistemas dinâmicos deve ser utilizadas técnicas de reconstrução do espaço de fase. A reconstrução do espaço de fase é um método de análise dinâmica não linear, que mostra padrões visuais da dinâmica vocal que representam consideravelmente componentes caóticas e determinísticas presentes na voz.

Combinada com métodos estatísticos, como o expoente de Lyapunov ou a entropia (PINHO, 2016), pode-se determinar algumas medidas quantitativas que possam caracterizar o sinal.

Um método utilizado para essa reconstrução é baseado no Teorema de imersão de *Takens* (VIEIRA, 2014) (PINHO, 2016), no qual são construídos n vetores \mathcal{E}_i m -dimensionais

a partir da série temporal $x(t)$ e de suas versões defasadas no tempo, representada por $m-1$, sendo m a dimensão de imersão e τ o tempo de defasagem, sendo $i = 1, 2, \dots, T$, conforme equação a seguir:

$$\varepsilon_i = \{x(t_i), x(t_i + \tau), \dots, x(t_i + (m-1)\tau)\} \quad (8)$$

Sendo que a quantidade N de estados, a partir do teorema de *Takens*, pode ser obtida através da equação:

$$N = T - (m-1)\tau \quad (9)$$

A determinação da dimensão de imersão pode ser obtida, de maneira mais simplificada, através de duas formas: a primeira é a observação do comportamento do sistema quando se aumenta gradativamente a dimensão de imersão, e a outra é através do método dos falsos vizinhos, no qual é alterado gradativamente o valor de m e constata do quais pontos se distanciam.

A obtenção do tempo de defasagem τ é dada através do método da informação média mútua, no qual se pode garantir a reconstrução de vetores com menor nível de informação redundante, mas ainda correlacionados. A informação mútua média entre $x(t)$ e sua defasagem $x(t+\tau)$ é obtida a partir de um histograma de b intervalos de classes, criado para estimar a distribuição de probabilidade dos dados do sinal $x(t)$, conforme equação abaixo:

$$I(\tau) = \sum_b^{i=1} \sum_b^{j=1} P_{i,j}(x(t), x(t+\tau)) \log_2 \left[\frac{P_{i,j}(x(t), x(t+\tau))}{P_i(x(t))P_j(x(t+\tau))} \right] \quad (10)$$

em que $P_i(x(t))$ é a probabilidade de que o sinal $x(t)$ assuma um valor dentro do i -ésimo intervalo do histograma, $P_j(x(t+\tau))$ é a probabilidade de que $x(t+\tau)$ esteja no j -ésimo intervalo e $P_{i,j}(x(t), x(t+\tau))$ é a probabilidade de que, simultaneamente, $x(t)$ esteja no i -ésimo intervalo e $x(t+\tau)$ no j -ésimo intervalo. Assim, o tempo de defasagem τ , então, é o valor para o qual a função de informação mútua média atinge seu primeiro mínimo local (VIEIRA, 2014), que representa a autocorrelação da série medindo a dependência do sinal consigo mesmo.

A diferença do espaço de fase proveniente de uma voz patológica e uma voz saudável podem ser percebidas pelas figuras 8 e 9, respectivamente, em que a primeira apresenta uma trajetória mais difícil de ser identificada, sendo esta voz a mesma que originou o sinal mostrado na figura 6, enquanto que a segunda a trajetória é perceptível mais facilmente.

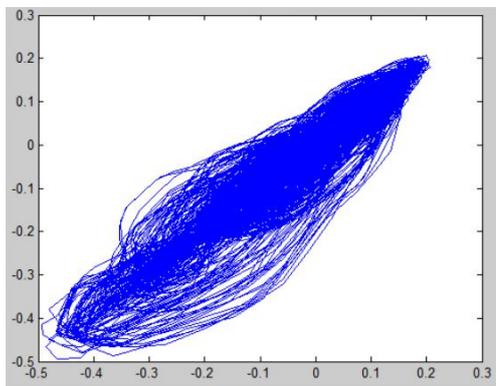


Figura 8: Espaço de fase para um sinal de voz patológica.

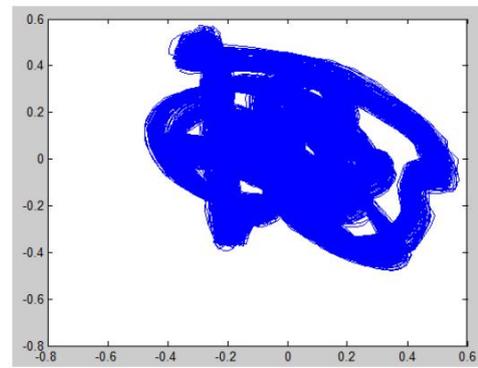


Figura 9: Espaço de fase para um sinal de voz saudável.

2.6 CLASSIFICADORES

O reconhecimento de padrões é o campo da ciência que tem por objetivo a classificação de objetos, em um determinado número de categorias ou classes, a partir da observação de suas características, ou seja, é uma maneira mais simples de representar um conjunto de dados através de suas características mais relevantes, possibilitando, assim, sua divisão em classes.

Deste modo, pode-se entender que os classificadores são utilizados para classificar, ou descrever padrões ou objetos, baseados em um conjunto de suas propriedades ou características.

Essencialmente, existem dois casos de reconhecimento de padrões:

- 1) Classificação Supervisionada,
- 2) Classificação não-supervisionada.

Na classificação supervisionada são selecionadas as amostras para cada uma das classes que se deseja classificar, sendo o padrão e as classes já conhecidas previamente, sendo este conjunto chamado de “conjunto treinamento”. Já na classificação não-supervisionada, o padrão e o número total de classes não são conhecidos, o conjunto de dados é dividido em grupos de acordo com características específicas, características essas que são extraídas das amostras por meio das medidas.

As características extraídas do sinal formam um espaço multidimensional, o espaço de características, onde cada medida é representada por um vetor de característica como sendo um ponto neste espaço. Assim, a tarefa de classificação pode ser entendida geometricamente como o reconhecimento de agrupamentos no espaço de características.

A escolha de características adequadas à classificação é feita a partir da observação dos objetos conhecidos e de suas respectivas classes, ou seja, o conjunto treinamento. O grupo de características deve, por si só, descrever bem os objetos, agrupando os objetos similares e separando os objetos distintos. Ele precisa ser robusto para tolerar as diferenças dentro de cada classe e ser sensível o bastante para discriminar essas classes. Assim, as características são escolhidas de modo a facilitar a tarefa do classificador.

O classificador é a função que calcula as medidas de similaridade entre objetos no espaço de características, podendo ser divididos em dois tipos: os estatísticos (paramétricos e

não-paramétricos) e os conexionistas (redes neurais).

No classificador estatístico os objetos do conjunto treinamento são utilizados para estimar certos parâmetros estatísticos de cada classe, e esses parâmetros especificam o classificador. Já o classificador conexionista é mais complexo, aprendendo por experiência.

Neste trabalho, o classificador utilizado serão as Redes Neurais Artificiais, sendo esta detalhada a seguir.

2.6.1 Redes Neurais Artificiais (RNA's)

Uma Rede Neural Artificial é um sistema de processamento de informações que apresentam algumas características em comum com as redes neurais biológicas, que formam o sistema nervoso dos seres vivos, aqui representado pela figura 10, que mostra um neurônio humano. Estas características em comum foram elaboradas através do estudo dos neurônios e então seu funcionamento foi reproduzido similarmente através de um conjunto de algoritmos.

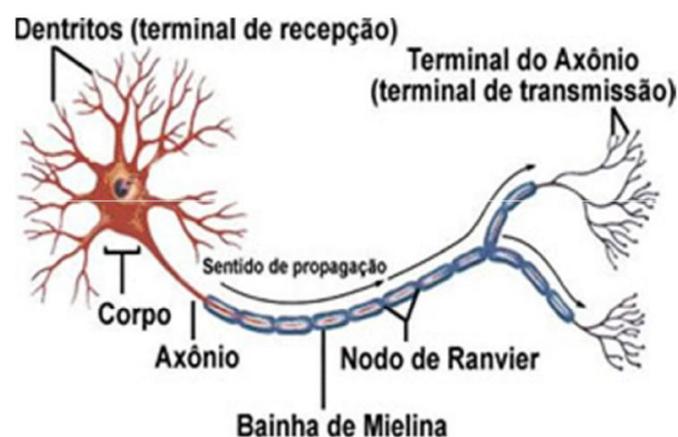


Figura 10: Representação de um neurônio.

Fonte:Infoescola.

Assim sendo, pode-se dizer que o funcionamento de uma RNA tem duas semelhanças principais com a rede de neurônios cerebral humana: é capaz de estruturar uma base de conhecimentos baseada em um processo de aprendizagem do ambiente no qual está inserida,

e modelar o conceito de pesos sinápticos, para assim ser capaz de armazenar o conhecimento adquirido.

Uma RNA pode ser definida como um processador maciçamente paralelo e distribuído, formado de simples unidades capazes de armazenar conhecimento baseado em experiência. (HAYKIN, 2001)

A partir de generalizações do modelo de um neurônio artificial, inicialmente proposto por McCulloch e Pitts, podem-se identificar três elementos básicos: (HAYKIN, 2001)

- 1) Um conjunto de sinapses caracterizadas por seu respectivo peso;
- 2) Um somatório dos sinais de entrada ponderados pelos pesos das respectivas sinapses.
- 3) Uma função de ativação para limitar a amplitude da saída do neurônio.

Cada neurônio, considerado como uma unidade de processamento e representado pela figura 11, calcula a soma ponderada de suas entradas, cujos terminais de entrada são identificados por x_1, x_2, \dots, x_n , representando de maneira análoga os dendritos, mais um polarizador identificado como *bias* e aplica o resultado em uma função de ativação que poderá produzir ou não uma saída, identificada por y_k . As sinapses são representadas pelos pesos $w_{k1}, w_{k2}, \dots, w_{kn}$, e da mesma maneira que os neurônios biológicos, estes pesos podem assumir valores positivos ou negativos, assim, portanto, os sinais de entrada são dados em função de $x_i w_i$.

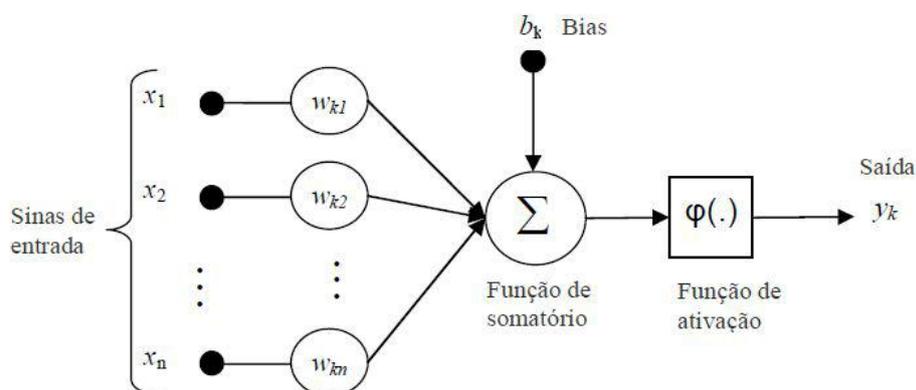


Figura 11: Modelo de neurônio artificial.

Fonte: SANTANA, 2012.

Este funcionamento pode ser descrito, matematicamente, pela equação:

$$y_k = \varphi \sum_{j=1}^n [w_{kj} x_j + b_k] \quad (11)$$

onde n é o número de entradas do neurônio k , x_j é a entrada, w_{kj} é o peso correspondente a entrada j do neurônio k , b_k corresponde ao parâmetro polarizador φ é a função de ativação e y_k é a saída do neurônio k .

A função de ativação, identificada por φ , é a função responsável pela definição da saída do neurônio em função de seu nível interno de ativação. As funções de ativação mais comumente usadas são duas:

1) Função limiar: cuja saída pode ser observada na equação abaixo:

$$f(u) = \begin{cases} 1, & \text{para } u > 0 \\ 0, & \text{para } u \leq 0 \end{cases} \quad (12)$$

2) Função linear: sua saída é representada pela equação abaixo, sendo α um número real que define a saída linear:

$$f(u) = \alpha u + \beta \quad (13)$$

Neste trabalho, primeiramente, será utilizada a função linear de ativação, em que uma voz saudável será classificada como 0, e uma voz patológica como 1. Na continuidade do trabalho, serão testadas outras funções de ativação e comparados os resultados obtidos.

A forma como os neurônios são conectados define diferentes tipos de topologias de rede. De maneira geral, podem-se identificar três classes de arquitetura de redes neurais:

- 1) Redes *Feedforward* de camada única: nesta classe os neurônios são organizados em uma única camada que produzem diretamente a saída da rede. O termo *feedforward* indica o sentido único de propagação dos sinais, não existindo ciclos na camada.
- 2) Redes *Feedforward* de multicamadas: existem ou mais camadas escondidas, isto é, esta parte da rede não é diretamente conectada a saída da rede
- 3) Redes recorrentes: esta classe é caracterizada pela presença de pelo menos um laço de realimentação.

O processo de aprendizagem de uma RNA se dá através do ajuste dos pesos da rede neural a partir de estímulos fornecidos por um conjunto de dados. (SANTANA, 2012) Esse

ajuste pode ser feito, principalmente, pelas relações de Regra Delta e pelo cálculo do Gradiente Local, sendo que neste trabalho o ajuste se dará através da utilização do gradiente local, descrito matematicamente pela equação abaixo: (SODRÉ, 2016)

$$\delta_j(n) = e_j(n)\varphi'(v_j(n)) \quad (14)$$

em que $e_j(n)$ é o erro entre a entrada e a saída do neurônio j e a saída desejada da iteração n , e $\varphi'(v_j(n))$ é a derivada da função de ativação do neurônio j em relação a saída linear do neurônio na iteração n .

O algoritmo que será utilizado para o treinamento da rede neural deste estudo, será um algoritmo por *backpropagation* (ou retropropagação), mais precisamente o algoritmo de Levenberg-Marquardt, bastante utilizado na visão computacional sempre que deseja-se ajustar um modelo a um conjunto de dados experimentais (FRANÇA, 2009), que é um algoritmo de aprendizagem supervisionada, no qual se utiliza pares de entrada/saída desejada para ajustar os pesos da RNA por meio de um mecanismo de correção do erro. (SANTANA, 2012). Este treinamento se dá em duas fases: a fase de propagação, em que ocorre a propagação do sinal de entrada ao longo da rede, sendo os pesos sinápticos mantidos fixos, e a fase de retropropagação, na qual o sinal de erro é produzido através da comparação da saída obtida com a saída desejada, fazendo-se, então, o ajuste dos pesos das conexões da rede.

3. RESULTADOS E DISCUSSÕES

Nesta seção serão apresentados os resultados obtidos neste trabalho. Primeiramente, de modo que seja facilitada a compreensão, dar-se-á uma breve explicação da maneira que os resultados foram alcançados.

A partir dos sinais provenientes do banco de dados os valores de *Pitch* puderam ser extraídos de cada sinal por meio de quatro métodos diferentes, sendo estes os seguintes métodos: YIN, ACF, PRAAT e RAPT.

O YIN é um método estimador cujo objetivo é tentar corrigir alguns problemas da autocorrelação, em que são identificados picos através da variação dos atrasos, sendo que alguns picos sub-harmônicos podem ser erroneamente identificados como candidatos a frequência fundamental. (DALTER, 2016)

Já o ACF (*Autocorrelation Function*) baseia-se na medida de similaridade entre amostras de um sinal, através da soma do produto destas, espaçadas por determinado atraso (CARANHA, 2013), podendo este estimador obter os resultados através do método da Convolução Discreta ou através da FFT (*Fast Fourier Transform*).

Por sua vez, o estimador PRAAT é um algoritmo robusto que utiliza como base a função de autocorrelação normalizada para prováveis frequências fundamentais, e depois refina os resultados através de uma função que minimiza os erros e pondera os resultados. (MACHADO, 2013)

E por último, o RAPT (*Robust Algorithm for Pitch Tracking*) é um algoritmo robusto baseado na normalização da função de autocorrelação cruzada. (BRANDÃO *et al*, 2007)

Na sequência, a partir dos resultados obtidos para *Pitch*, foram obtidas a medida acústica *Jitter* de duas maneiras: a primeira que representa o *Jitter* Local, que é a diferença média absoluta entre dois períodos consecutivos divididos pelo período médio, o *Jitter* Absoluto, sendo este a diferença média absoluta entre dois períodos consecutivos.

A partir do processamento de todos os dados necessários, sendo a primeira amostragem composta por 53 sinais saudáveis e 657 sinais portadores de alguma patologia, configurando um total de 710 amostras, e a segunda composta por 53 sinais saudáveis e 56

sinais patológicos, sendo um total de 109 amostras, deu-se então o início do treinamento das Redes Neurais Artificiais.

3.1 RESULTADOS DA REDE NEURAL: DISCRIMINAÇÃO ENTRE VOZ SAUVÁVEL X PATOLÓGICA, PARA VALORES DE JITTER

Através da ferramenta de RNA proveniente do software Matlab, diversas abordagens e configurações puderam ser testadas. Primeiramente os valores *default* foram mantidos, sendo estes a quantidade de neurônios igual a 10 e as porcentagens de treinamento, validação e teste sendo 70%, 15% e 15% respectivamente, e então os valores foram sendo alterados de modo que se obtivesse o melhor resultado.

Os resultados obtidos foram avaliados através da matriz de confusão proveniente de cada teste, sendo cada configuração testada três vezes e então o resultado final considerado como sendo a média dos demais valores obtidos, e então efetuada a comparação dos resultados obtidos.

De maneira simplificada, a matriz de confusão pode ser entendida como uma matriz composta por valores positivos e negativos, assim como falsos positivos e falsos negativos, podendo sua configuração ser observada na figura abaixo:

	Positiva	Negativa
Positiva	Verdadeiros Positivos (TP)	Falsos Positivos (FP)
Negativa	Falso Negativos (FN)	Verdadeiros Negativos (TN)

Figura 12: configuração da matriz de confusão genérica.

Fonte: Autoria própria.

A matriz de confusão resultante obtida no Matlab é o resultado do treinamento, da validação e do teste das RNA, para assim compor a matriz de confusão final, que pode ser vista a seguir, utilizando como exemplo a matriz de confusão obtida numa configuração utilizando combinações das medidas do YIN e do ACF, conforme figura 13, e cujos dados resultantes podem ser vistos na tabela 18.



Figura 13: matriz de confusão obtida para terceiro teste utilizando os dados provenientes da combinação dos estimadores YIN e ACF, para a configuração de 50 neurônios e porcentagens de treinamento, teste e validação de 70%, 15% e 15%, respectivamente, para uma configuração equiparada.

Fonte: Autoria própria.

Os melhores resultados serão apresentados na forma de tabela, contendo os valores obtidos de Acurácia, Precisão, Sensibilidade e Especificidade, bem como a média dos resultados apresentados.

A sensibilidade visa medir a proporção de verdadeiros positivos classificados corretamente, já a especificidade mede a quantidade de verdadeiros negativos classificados adequadamente. A precisão mede a proporção de verdadeiros positivos classificados e acurácia identifica a proporção de classificações corretas para o total de elementos classificados. (GONZALEZ, 2013) As expressões que foram utilizadas para o cálculo dessas variáveis podem ser vistas no quadro a seguir, tendo em vista a configuração da matriz de confusão genérica apresentada na figura 12.

	Acurácia	Precisão	Sensibilidade	Especificidade
Fórmulas	$\frac{TP+TN}{TP+TN+FP+FN}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$\frac{TN}{TN+FP}$

Quadro 4: fórmulas para cálculo de Acurácia, Precisão, Sensibilidade e Especificidade.

Fonte: Autoria própria.

Essa abordagem visa evitar que a alta taxa de verdadeiros positivos esconda uma alta taxa de falsos positivos, dada em função da disparidade de amostras (alta quantidade de amostras patológicas comparada ao número total de amostragem).

De maneira que o entendimento seja facilitado, um exemplo fictício de matriz de confusão pode ser utilizado, supondo um total de 50 casos, sendo 30 patológicos e 20 saudáveis. Dentre os casos patológicos, a matriz de confusão obtida identificou como sendo 20 patológicos e 10 saudáveis, e dentre os casos saudáveis foram identificados 5 saudáveis e 15 como sendo patológicos, resultando em uma matriz de confusão conforme o quadro a seguir.

	Acurácia	Precisão	Sensibilidade	Especificidade
Fórmulas	$\frac{TP+TN}{TP+TN+FP+FN}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$\frac{TN}{TN+FP}$
Exemplo	$\frac{20+5}{20+5+15+10}$	$\frac{20}{20+15}$	$\frac{20}{20+10}$	$\frac{5}{5+15}$
Resultado	0,5	0,5714	0,666	0,25

Quadro 5: Resultado para cálculo de Acurácia, Precisão, Sensibilidade e Especificidade.

Fonte: Autoria própria.

Para este exemplo fictício obteve-se uma média dos resultados igual a 0,4968, ou seja, embora este caso tenha apresentado uma boa sensibilidade, sua especificidade foi baixa, resultando em uma taxa de acertos de cerca 49%.

Devido as diferentes configurações propostas, serão mostradas aquelas que obtiveram os melhores resultados, sendo considerados satisfatórios aqueles que apresentaram taxa de acerto superior a 60%.

A primeira configuração que teve um bom desempenho foi caracterizada apenas pelas medidas de *Jitter* Absoluto, mas contando com os resultados obtidos com o uso de todos os estimadores de *pitch*, sendo este desempenho observado na tabela a seguir:

Tabela 1: resultado obtido utilizando-se todas os valores de *Jitter* Absoluto obtidos.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,9239	0,9252	0,9984	0	0,7119	10 Neurônios (70/15/15)
0,9248	0,92537	0,9994	0	0,7124	20 Neurônios (70/15/15)
0,9253	0,9253	1	0	0,7126	40 Neurônios (70/15/15)
0,9239	0,9252	0,9984	0	0,7119	10 Neurônios (60/20/20)
0,9253	0,9253	1	0	0,7126	20 Neurônios (60/20/20)
0,9253	0,9253	1	0	0,7126	40 Neurônios (60/20/20)

Fonte: Autoria própria.

Nesta configuração a quantidade de neurônios foi variada de 0 a 100, bem como a porcentagem entre treinamento, validação e testes alterada, sendo os melhores resultados obtidos para a configuração cuja porcentagem era de 60%, 20% e 20% respectivamente, sendo que a alteração do número de neurônios não mostrou melhora significativa no aprendizado da rede.

A próxima configuração que obteve um resultado satisfatório foi utilizando-se das medidas de *Jitter* Local, sendo também os resultados obtidos, mostrados a seguir, provenientes do uso dos valores resultantes de todos os estimadores.

Tabela 2: resultado obtido utilizando-se todos os valores de *Jitter* Local.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,9258	0,9265	0,9989	0,0188	0,7175	10 Neurônios (70/15/15)
0,9248	0,9253	0,9994	0	0,7124	20 Neurônios (70/15/15)
0,9253	0,9253	1	0	0,7126	40 Neurônios (70/15/15)
0,9253	0,9265	0,9984	0,0188	0,7173	10 Neurônios (70/15/15)
0,9253	0,9257	0,9994	0,0062	0,7142	20 Neurônios (60/20/20)
0,9253	0,9253	1	0	0,7126	40 Neurônios (60/20/20)

Fonte: Autoria própria.

Com a utilização destas medidas, os resultados podem ser observados como sendo um pouco mais sensíveis a alteração dos parâmetros, sendo a melhor configuração, caracterizada com o retângulo em vermelho, obtida com o número de 10 neurônios, e as porcentagens de treinamento, validação e testes respectivamente como sendo 70%, 15% e 15%.

Os resultados da configuração fazendo o uso de todas as medidas obtidas para todos os estimadores pode ser vista na tabela abaixo.

Tabela 3: resultado obtido com uma configuração utilizando todos os valores resultante do uso de todos os estimadores.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,9253	0,9261	0,9989	0,0125	0,7157	10 Neurônios (70/15/15)
0,9253	0,9297	0,9944	0,0691	0,7298	20 Neurônios (70/15/15)
0,9286	0,9328	0,9944	0,1132	0,7422	40 Neurônios (70/15/15)
0,9253	0,928	0,9949	0,0628	0,7278	10 Neurônios (60/20/20)
0,9272	0,9287	0,9979	0,0503	0,726	20 Neurônios (60/20/20)
0,9262	0,9319	0,9928	0,1006	0,7379	40 Neurônios (60/20/20)

Fonte: Autoria própria.

A alteração dos parâmetros para esta configuração pode ser percebida como sendo mais expressiva no resultado obtido, sendo o melhor resultado caracterizado com o retângulo em vermelho, obtida com o número de 40 neurônios, e as porcentagens de treinamento, validação e testes respectivamente como sendo 70%, 15% e 15%. De modo geral nesta configuração, quanto menor o número de neurônios pior era o desempenho da rede, e em contrapartida a melhora do resultado se deu até o aumento para o valor de 40 neurônios, sendo que o aumento acima não mostrou melhoras significativas. A alteração das porcentagens não mostrou influência significativa no resultado.

A seguir, as configurações testadas visavam comparar o desempenho para os valores obtidos com cada estimador separadamente, sendo os melhores resultados mostrados na tabela abaixo.

Tabela 4: estimadores que apresentaram um melhor desempenho.

	Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
Praat	0,9286	0,9353	0,9913	0,1509	0,7515	50 Neurônios (70/15/15)
Rapt	0,9258	0,9294	0,9954	0,0628	0,7283	10 Neurônios (70/15/15)
Rapt	0,9277	0,9303	0,9964	0,0754	0,7324	50 Neurônios (70/15/15)

Fonte: Autoria própria.

Os estimadores YIN e ACF, quando testados isoladamente, não obtiveram resultados satisfatórios, não sendo, então, computados seus valores obtidos.

Na configuração utilizando apenas os valores obtidos através do PRAAT o melhor desempenho se deu para o número de 50 neurônios, com porcentagens de 70%, 15% e 15% para treinamento, validação e teste, respectivamente. Com o RAPT, a melhor configuração foi caracterizada da mesma forma.

Levando-se em conta que o tamanho e a representatividade do conjunto de dados de treinamento podem influenciar a capacidade de generalização de uma RNA (GONZAGA, 2017) , para certificação de que a composição do grupo de dados da amostragem não tenha mascarado os resultados obtidos devido à alta quantidade de amostras patológicas, todas as configurações foram testadas novamente agora com uma amostragem equiparada, contando com 53 casos normais e 56 patológicos, sendo estes escolhidos ao acaso. Os resultados que obtiveram uma melhora significativa podem ser vistos a seguir.

A primeira configuração que apresentou melhora significativa do resultado foi utilizando-se das medidas de *Jitter* Local, cujos parâmetros utilizados foram de 10 neurônios e as porcentagens de treinamento, validação e teste sendo 70%, 15% e 15% respectivamente. O resultado obtido pode ser visto na tabela abaixo.

Tabela 5: resultado obtido utilizando-se todos os valores de *Jitter* Local, com quantidade de amostras equilibradas.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,7523	0,8	0,6904	0,8176	0,765	10 Neurônios (70/15/15)

Fonte: Autoria própria.

Tendo em vista que os parâmetros que obtiveram melhor resultado foram iguais para ambos os casos, sendo 10 neurônios e as porcentagens de treinamento, validação e teste sendo 70%, 15% e 15%, pode-se perceber que uma melhor distribuição do conjunto de dados pode resultar numa melhora no aprendizado das RNA e por consequência pode-se obter um resultado com uma taxa de acertos melhor, como pode ser visto na tabela a seguir.

Tabela 6: melhora na média obtida para *Jitter* Local, sendo a Média 1 a melhor média obtida utilizando-se todos os dados, e Média 2 o resultado obtido com uma equiparação das amostras saudáveis e patológicas, ambos os resultados foram discriminados anteriormente.

Média 1	Média 2
0,7175	0,765

Fonte: Autoria própria.

Utilizando-se novamente de todas as medidas obtidas, a configuração que apresentou melhor desempenho foi caracterizada com o número de 50 neurônios e as porcentagens de treinamento, validação e teste sendo 75%, 10% e 15% respectivamente, estando destacada na tabela abaixo com um retângulo em vermelho.

Tabela 7: melhora na média obtida para configuração utilizando todas as medidas, com equiparação de amostras.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,7553	0,7588	0,7678	0,7421	0,756	10 Neurônios (70/15/15)
0,7889	0,8113	0,7678	0,8113	0,7948	50 Neurônios (70/15/15)
0,792	0,8333	0,744	0,8427	0,803	50 Neurônios (75/10/15)

Fonte: Autoria própria.

Com a equiparação do conjunto de amostragem pode-se perceber uma melhora significativa da média obtida, determinando a importância de um conjunto de dados formado de maneira que se tenha uma melhor representatividade dos grupos utilizados.

Tabela 8: melhora na média obtida para configuração utilizando todas as medidas, com equiparação de amostras.

Média 1	Média 2
0,7296	0,756
0,7379	0,7948
0,7422	0,803

Fonte: Autoria própria.

3.2 RESULTADOS DA REDE NEURAL: DISCRIMINAÇÃO ENTRE VOZ SAUDÁVEL X PATOLÓGICA, PARA VALORES DE HNR

De forma semelhante, foram executados os procedimentos para obtenção das medidas de HNR para cada amostra proveniente de cada estimador, e então calculados os valores de desvio padrão e média, que seguindo as mesmas premissas, foram classificados sob diversas abordagens e configurações da RNA.

Primeiramente, foram testados todos os valores obtidos de desvio padrão e média, provenientes dos quatro estimadores. Os resultados podem ser vistos na tabela a seguir.

Tabela 9: resultado obtido com a utilização de todas as medidas.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,3934	0,9449	0,3658	0,73584	0,61	10 neuronios (70/15/15)
0,7239	0,9346	0,7544	0,3459	0,6897	20 neuronios (70/15/15)
0,2267	0,9736	0,1689	0,9433	0,5781	30 neuronios (70/15/15)
0,3934	0,9449	0,3658	0,7358	0,61	10 neuronios (60/20/20)
0,6981	0,9346	0,7544	0	0,596	20 neuronios (60/20/20)
0,2267	0,9736	0,1689	0,9433	0,5781	30 neuronios (60/20/20)

Fonte: Autoria própria.

O melhor resultado foi obtido para os parâmetros com um número de 20 neurônios, e as porcentagens de treinamento, validação e teste sendo 70%, 15% e 15% respectivamente, já o pior resultado foi obtido duas vezes, sendo ambas com o número de neurônios igual a 30, e as porcentagens de 70%, 15% e 15%, e 60%, 20% e 20%.

A seguir, foram testadas somente as medidas de desvio padrão, e depois somente as médias, sendo os melhores valores obtidos mostrados na tabela abaixo

Tabela 10: melhores resultados obtidos com todas as medidas de desvio padrão e na sequência a média.

	Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
Desvios Padrão	0,968	0,9708	0,9959	0,6226	0,8898	50 neuronios (70/15/15)
Médias	0,9323	0,9385	0,9918	0,1949	0,7644	10 neuronios (70/15/15)

Fonte: Autoria própria.

Para o desvio padrão a configuração utilizada foi de 50 neurônios, e as porcentagens de treinamento, validação e teste sendo 70%, 15% e 15% respectivamente, e para a média os parâmetros utilizados foram de 10 neurônios, e as porcentagens de treinamento, validação e teste sendo 70%, 15% e 15% respectivamente.

Na sequência foram testadas combinações com as medidas obtidas pelos estimadores, sendo o melhor resultado proveniente da combinação das medidas de desvio padrão e média dos estimadores YIN, ACF e PRAAT, cujos resultados serão mostrados a seguir. A alteração dos parâmetros, tanto do número de neurônios quanto das porcentagens utilizadas não representou grande melhora no resultados.

Tabela 11: melhor resultado obtido a partir de combinações, com YIN, ACF e PRAAT.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,4685	0,9383	0,4556	0,6289	0,6228	10 neuronios (70/15/15)
0,4845	0,9056	0,4723	0,6352	0,6244	20 neuronios (70/15/15)

Fonte: Autoria própria.

Os resultados provenientes dos testes para cada estimador separadamente não foram satisfatórios, sendo estes com cerca de menos de 50% de acerto, e portanto, não sendo analisados estes valores.

Com base na questão da importância do tamanho e representatividade do conjunto de dados, os mesmos testes foram feitos agora com número equilibrado de vozes saudáveis e patológicas, sendo estas, novamente, escolhidas ao acaso. Os melhores resultados podem ser vistos a seguir.

Para a configuração utilizando todas as medidas, pode-se perceber uma grande melhora no resultado obtido em comparação com o anterior, valores estes não sofrendo grande influencia na alteração do número de neurônios e porcentagens utilizadas.

Tabela 12: melhores resultados obtidos com a utilização de todas as medidas, com amostragem equiparada.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,9197	0,9426	0,8969	0,9433	0,9256	10 neuronios (70/15/15)
0,9074	0,9411	0,8727	0,9433	0,91617	50 neuronios (75/10/15)

Fonte: Autoria própria.

A seguir, foram testadas novamente somente as medidas de desvio padrão, e depois somente as médias, sendo ambos os resultados obtidos para os parâmetros de 50 neurônios e porcentagens de 75%, 10% e 15%, cujo resultado é mostrado abaixo, embora a alteração destes também não tenham causado mudanças significativas nos resultados.

Tabela 13: melhores resultados obtidos com todas as medidas de desvio padrão e na sequência a média, com número de amostras equilibrado.

	Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
Desvios Padrão	0,8796	0,9256	0,8303	0,9308	0,8916	50 neuronios (75/10/15)
Médias	0,8518	0,9034	0,7939	0,9119	0,8652	50 neuronios (75/10/15)

Fonte: Autoria própria.

A seguir, os testes realizados foram configurados utilizando as medidas provenientes de cada estimador separadamente, e na sequência diferentes combinações utilizando os valores obtidos de cada estimador.

Os melhores resultados utilizando cada estimador isoladamente se deram para os valores obtidos pelo YIN e pelo PRAAT, que podem ser vistos na tabela a seguir, sendo o primeiro com os parâmetros de 50 neurônios e porcentagens de 70%, 15% e 15%, e o segundo com 10 neurônios e porcentagens de 70%, 15% e 15%. Alterações desses valores não resultaram em mudanças significativas nos resultados encontrados.

Tabela 14: melhores resultados obtidos para cada estimador.

	Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
YIN	0,8364	0,9516	0,715	0,9622	0,8663	50 neuronios (70/15/15)
PRAAT	0,8672	0,891	0,8422	0,893	0,8734	10 neuronios (70/15/15)

Fonte: Autoria própria.

Já para as combinações testadas, o melhor desempenho foi resultado da combinação de medidas provenientes do YIN e do ACF, com parâmetros sendo 50 neurônios e porcentagens de 70%, 15% e 15%, sendo os valores encontrados mostrados na tabela abaixo. Dentre todos os resultados obtidos provenientes da classificação pela RNA, esta combinação foi a que apresentou melhor desempenho, atingindo uma taxa de acerto de cerca de 95%.

Tabela 15: melhor resultado obtido a partir de combinações do estimador YIN e ACF.

Acurácia	Precisão	Sensibilidade	Especificidade	Média	Configuração
0,9567	0,9629	0,9454	0,9685	0,9584	50 neuronios (70/15/15)

Fonte: Autoria própria.

3.3 RESULTADOS DA REDE NEURAL: DISCRIMINAÇÃO ENTRE GRUPOS DE PATOLOGIAS

De maneira análoga, a classificação de patologias se deu de duas maneiras, sendo a primeira utilizando os valores referentes ao *Jitter*, e a segunda os valores de HNR. As porcentagens de treinamento, validação e teste, bem como o número de neurônios foram sendo alterados de maneira que pudessem ser encontrados os valores que apresentassem um melhor resultado. Cada configuração foi testada três vezes e seu resultado final considerado como sendo a média dos valores obtidos em cada teste.

A análise dos resultados obtidos se deu através da matriz de confusão proveniente de cada teste, sendo esta matriz de terceira classe (matriz 3x3), portando os valores resultantes foram estudados através do coeficiente *Kappa*, que é uma medida de concordância que define um limiar que permite avaliar o desempenho do sistema. O índice *Kappa* pode ser definido pela expressão 15, em que $P(A)$ é a proporção em que os anotadores concordam e $P(E)$ é a proporção de vezes que é esperado dos anotadores concordarem ao acaso ou aleatoriamente. (MATOS *et al*, 2009)

$$k = \frac{P(A) - P(E)}{1 - P(E)} \quad (15)$$

Os valores de *Kappa* podem ser interpretados através de uma escala de seis níveis, como mostrado na tabela abaixo:

Tabela 16: escala de seis níveis de concordância para o coeficiente *Kappa*.

Concordância	Escala
Ruim	Abaixo de 0
Leve	0,00 a 0,20
Sofrível	0,21 a 0,40
Moderada	0,41 a 0,60
Boa	0,61 a 0,80
Quase perfeita	Acima de 0,81

Fonte: MATOS *et al*, 2009.

Considerando a matriz de confusão fictícia a seguir, foi realizado o cálculo de *Kappa* para que o entendimento seja facilitado.

	1	2	3	Total
1	25	2	5	32
2	3	26	5	34
3	1	3	30	34
Total	29	31	40	100

Quadro 6: Matriz fictícia de terceira ordem para cálculo de *Kappa*.

Fonte: MATOS *et al*, 2009.

P(A)	P(E)	<i>Kappa</i>
$\frac{25+26+30}{100}$	$\frac{32+29+34+31+34+40}{100^2}$	$\frac{0,81-0,33}{1-0,33}=0,72$

Quadro 7: Resolução da matriz fictícia de terceira ordem para cálculo de *Kappa*.

Fonte: MATOS *et al*, 2009.

Primeiramente, foram estabelecidos os grupos de patologias que seriam investigados e que fariam parte deste estudo, sendo os grupos testados inicialmente apenas para verificar se o

número de amostras seria suficiente para gerar resultados satisfatórios. As amostras foram compostas conforme a tabela abaixo:

Tabela 17: amostras utilizadas.

Patologia	Número de casos
Carcinoma	2
Cistos	7
Edema de Reinke	24
Granuloma	4
Leucoplasia	27
Nódulo	20
Pólipo	20

Fonte: Autoria própria.

3.3.1 Grupo Carcinoma x Leucoplasia, e Nódulo x Cisto

O primeiro teste foi realizado utilizando-se das amostras de leucoplasias e carcinomas, numa tentativa de verificar o que estas patologias possuem em comum e possíveis pontos que as diferenciam. Devido ao escasso número de casos de carcinomas disponíveis, pode-se perceber que, embora mesmo com as diversas configurações abordadas na RNA, a representatividade do conjunto de dados interferiu nos valores obtidos, mascarando o resultado alcançado.

De forma semelhante, ao utilizar as amostras de nódulos e cistos, as mesmas dificuldades foram encontradas dada a quantidade de amostras de cistos disponíveis. Numa tentativa de melhorar a representatividade do conjunto de dados, 7 casos de nódulos foram escolhidos ao acaso e, então, juntamente com o 7 casos de cistos disponíveis, foi formado o grupo de amostras e então treinada a RNA. Diversas abordagens foram testadas e em todas os resultados obtidos eram similares, levando-se a crer que o conjunto de dados não foi suficiente para que a RNA pudesse ter um desempenho satisfatório.

Portanto, os resultados obtidos para estes conjuntos de amostras acima mencionados não foram analisados.

3.3.2 Grupo Nódulo x Edema de Reinke x Leucoplasia

A seguir, foram comparadas as amostras de patologias que causam alteração no número de ciclos glóticos: nódulos, edema de Reinke e leucoplasias. Os resultados obtidos a partir dos valores de *Jitter* foram discriminados abaixo, de acordo com a configuração utilizada.

Primeiramente foram utilizados os valores obtidos a partir de todos os estimadores para *Jitter* Local, e então para *Jitter* Absoluto, sendo o pior e o melhor resultado mostrado na tabela a seguir.

Tabela 18 : resultados obtidos para Jitter Local e Absoluto.

	<i>Kappa</i>	Configuração	Concordância
<i>Jitter</i> Local	0,25455	30 neurônios (75/10/15)	Sofrível
<i>Jitter</i> Absoluto	0,2397	10 neurônios (80/10/10)	Sofrível

Fonte: Autoria própria.

Para o *Jitter* Local a melhor configuração, que resultou em um melhor desempenho, foi composta por 30 neurônios e as porcentagens de treinamento, validação e teste com valores de 75%, 10% e 15%, respectivamente. O aumento do número de neurônios, bem como a alteração das porcentagens para quaisquer outros valores resultou em uma piora no desempenho. Para o *Jitter* Absoluto a alteração dos parâmetros não evidenciou melhora significativa do resultado, sendo a melhor configuração dada pelo número de 10 neurônios e porcentagens de 80%, 10% e 10%. Em ambos os casos, a utilização dos valores obtidos somente para *Jitter* não resultou em uma taxa de acertos satisfatória, apresentando uma concordância considerada sofrível, de acordo com a tabela 18.

A seguir, foram utilizados todos os valores provenientes de todos os estimadores. Com a alteração dos parâmetros pode-se perceber que houve uma melhora da taxa de acertos com o aumento do número de neurônios até o valor de 50, e então houve uma tendência de

estabilização do resultado, já a alteração dos valores referentes à porcentagem não representou uma nenhuma mudança significativa, tanto para melhor quanto para pior. O melhor resultado obtido foi alcançado com uma configuração contendo 50 neurônios e porcentagem de treinamento, validação e teste de 75%, 10% e 15%, respectivamente. Os melhores resultados obtidos podem ser vistos na tabela abaixo, e de acordo com a tabela 19, representam um resultado de concordância moderado.

Tabela 19: resultados obtidos utilizando-se todos os valores resultantes de todos os estimadores.

	<i>Kappa</i>	Concordância
10 Neurônios (75/10/15)	0,5043	Moderada
30 Neurônios (75/10/15)	0,5348	Moderada
50 Neurônios (75/10/15)	0,5573	Moderada

Fonte: Autoria própria.

Na sequência, foram analisados os resultados provenientes do uso dos valores obtidos para cada estimador, com o objetivo de identificar aquele que apresentou melhor desempenho em relação ao outro.

As configurações que fizeram uso somente das medidas isoladas provenientes dos estimadores ACF e RAPT não apresentaram um resultado satisfatório, sendo que qualquer alteração nos parâmetros configurou uma piora do desempenho. Portanto, os resultados para tais estimadores não foram analisados.

Para os estimadores YIN e PRAAT pode-se perceber que, em ambos, o aumento do número de neurônios resultou em uma melhora, e a partir de um dado valor, caracterizado por 50 neurônios, houve uma estabilização do resultado. Já a alteração dos valores de porcentagens resultava em uma piora no desempenho. O melhor resultado, para cada estimador, pode ser visto na tabela abaixo, e em ambos, para esses resultados foram utilizados 50 neurônios e porcentagens de 70%, 15% e 15%, apresentando um índice de concordância moderado.

Tabela 20: resultados obtidos para medidas provenientes dos estimadores YIN e PRAAT.

	<i>Kappa</i>	Configuração	Concordância
YIN	0,4124	50 neurônios (70/15/15)	Moderada
PRAAT	0,5371	50 neurônios (70/15/15)	Moderada

Fonte: Autoria própria.

A partir da combinação dos estimadores, primeiro 2 a 2, e depois 3 a 3, percebeu-se, em todos os casos, a melhora do desempenho com o aumento do número de neurônios para o valor até 50 e uma piora nos resultados com quaisquer alteração das porcentagens. A combinação dos estimadores 3 a 3 não apresentou resultados satisfatórios, sendo não analisados neste trabalho os valores encontrados. Já para a combinação 2 a 2, os melhores resultados podem ser vistos a seguir, obtidos com 50 neurônios e porcentagens de 70%, 15% e 15%, e apresentando uma concordância moderada.

Tabela 21: resultados obtidos para combinações de estimadores.

	<i>Kappa</i>	Configuração	Concordância
YIN e ACF	0,5764	50 neurônios (70/15/15)	Moderada
ACF e PRAAT	0,4414	50 neurônios (70/15/15)	Moderada

Fonte: Autoria própria.

Em todas as combinações, foi observado que a inserção dos valores provenientes do estimador RAPT causavam uma piora nos resultados.

Já para as amostras provenientes dos valores de HNR, para patologias que causam alteração no número de ciclos glóticos, os resultados são discriminados a seguir, de acordo com a configuração utilizada.

Utilizando todas as medidas de desvio padrão, obtidas para todos os estimadores, obteve-se o resultado mostrado na tabela abaixo. Nesta configuração pode-se perceber que o aumento do número de neurônios ocasionou uma melhora no desempenho da rede até o número de 75 neurônios, sendo para valores acima observada uma estabilização na melhora. A alteração dos valores de porcentagem também causou uma alteração dos resultados, sendo o

melhor resultado obtido com os valores de porcentagem de treinamento, validação e teste de 70%, 10% e 20%, respectivamente, e o pior com 70%, 15% e 15%. Já para os números de neurônios, o melhor resultado foi obtido com 75 neurônios e o pior com 10. O melhor resultado apresentou um nível de concordância considerado bom.

Tabela 22: resultados obtidos utilizando-se todas os resultados de desvio padrão.

	<i>Kappa</i>	Configuração	Concordância
Desvio	0,2667	10 neurônios (70/15/15)	Sofrível
Padrão	0,618	75 neurônios (70/10/20)	Bom

Fonte: Autoria própria.

Para uma configuração utilizando-se todas as médias obtidas a partir dos valores de HNR percebeu-se que a alteração das porcentagens não influenciava no desempenho da rede. Em contrapartida, a alteração do número de neurônios, como já observado anteriormente, causava uma melhora no resultado obtido até certo ponto e depois, então, estabilizava. Os parâmetros que apresentaram melhor resultado foram 50 neurônios e valores de porcentagem de 70%, 15% e 15%, sendo este resultado mostrado na tabela a seguir, e caracterizado como sendo moderado quanto ao nível de concordância.

Tabela 23: resultados obtidos utilizando-se todas os resultados das médias.

	<i>Kappa</i>	Configuração	Concordância
Médias	0,584	50 neurônios (70/15/15)	Moderado

Fonte: Autoria própria.

Ao utilizar todas os resultados de desvio padrão e médias, obtidas para todos os estimadores, obteve-se um nível de concordância considerado bom, como mostrado na tabela a seguir. A tendência observada anteriormente também pode ser vista nesta configuração, sendo o melhor resultado caracterizado pelos parâmetros de 50 neurônios e valores de porcentagem de treinamento, validação e teste de 70%, 15% e 15%, respectivamente.

Tabela 24: resultados obtidos utilizando-se todos os resultados obtidos para todos os estimadores.

	<i>Kappa</i>	Configuração	Concordância
Todas as medidas	0,6685	50 neurônios (70/15/15)	Boa

Fonte: Autoria própria.

Partindo para uma análise isolada dos resultados obtidos para cada estimador, os valores provenientes dos estimadores YIN, PRAAT e RAPT não apresentaram resultados satisfatórios, sendo caracterizados por um nível de concordância leve, e portanto seus resultados não foram aqui discriminados.

Já o estimador ACF apresentou nível de concordância moderado, sendo o melhor desempenho, mostrado na tabela abaixo, caracterizado pelos valores de 50 neurônios e porcentagem de treinamento, validação e teste de 70%, 15% e 15%.

Tabela 25: resultado para os valores obtidos com o estimador ACF.

	<i>Kappa</i>	Configuração	Concordância
ACF	0,578	50 neurônios (70/15/15)	Moderado

Fonte: Autoria própria.

Com a combinação dos estimadores, novamente primeiro 2 a 2, e depois 3 a 3, observou-se que as primeiras combinações apresentaram melhores resultados, sendo então os resultados provenientes das combinações 3 a 3 não discriminados neste estudo.

Na combinação dos estimadores a tendência de melhora e estabilização foi preservada, sendo os melhores resultados obtidos mostrados a seguir.

Tabela 26: melhores resultados obtidos com a combinação de estimadores.

	<i>Kappa</i>	Configuração	Concordância
YIN e ACF	0,4922	50 neurônios (70/15/15)	Moderado
RAPT e PRAAT	0,6589	50 neurônios (70/15/15)	Boa
ACF e PRAAT	0,705	50 neurônios (70/15/15)	Boa

Fonte: Autoria própria.

A combinação entre YIN e ACF foi caracterizada por um nível de concordância moderado, sendo que qualquer alteração dos parâmetros default do sistema causavam uma piora nos resultados.

Já as combinações entre PRAAT e RAPT, e ACF e RAPT apresentaram um nível de concordância considerado bom, em ambos o melhor resultado foi obtido com um número de 50 neurônios e porcentagem de treinamento, validação e teste de 70%, 15% e 15%, respectivamente.

Dentre os estimadores, para valores de HNR, percebeu-se que a inclusão dos resultados provenientes do estimador YIN causavam uma piora no resultado. Em contrapartida, os valores provenientes do ACF causavam uma melhora, portanto mesmo a combinação de um estimador que apresentou resultados não satisfatórios, ao incluir-se os resultados de um bom estimador, pode-se chegar num nível moderado de concordância.

4. CONCLUSÃO

Através deste estudo, buscou-se minimizar o desconforto gerado pelos procedimentos invasivos de diagnóstico de patologias da laringe por meio de uma ferramenta que pudesse reconhecer padrões provenientes de doenças e que interfiram na voz de uma pessoa.

Esse reconhecimento de padrões se deu de duas formas, sendo a primeira relacionada à identificação de vozes patológicas e de vozes saudáveis. Com o intuito de prover uma assistência ao profissional da saúde para que uma primeira triagem possa ocorrer, essa identificação visa dar possibilidade para uma maior rapidez no atendimento do paciente, proporcionando um encaminhamento correto ao médico especialista e, caso seja necessário, uma intervenção de modo que a qualidade de vida do paciente seja preservada da melhor maneira possível.

Os resultados para esta discriminação demonstram que os valores provenientes de HNR, em comparação com os valores obtidos para *Jitter*, apresentam um melhor desempenho na RNA, sendo este melhor desempenho caracterizado pela combinação dos estimadores YIN e ACF, o qual atingiu uma taxa de acerto de 95%.

A segunda maneira em que o reconhecimento de padrões foi aplicado deu-se na discriminação entre grupos de patologias. Essa diferenciação mostrou-se mais complicada de se realizar, dado o desempenho apresentado pela RNA, levando-se a crer que alguns grupos de patologias geram sinais muito semelhantes ao ponto de qualquer configuração da RNA utilizada não conseguir fazer a determinação de maneira satisfatória.

Para esta classificação também foi observada que os valores provenientes de *Jitter* podem ser insuficientes para uma boa discriminação entre os grupos de patologias. Os valores de HNR, embora em seu melhor resultado tenha apresentado um nível de concordância bom, com cerca de apenas 70% de taxa de acertos, também configura uma melhor alternativa de uso em comparação ao *Jitter*.

Em ambos os casos as medidas utilizadas se mostram não totalmente satisfatórias para uma boa classificação, pois, se tratando de identificação de patologias e do risco inerente à vida do paciente, quanto maior a taxa de acertos do classificador maior a confiabilidade para que o profissional da saúde possa tomar a decisão mais acertada.

A partir dos resultados obtidos neste estudo, alguns pontos podem ser ressaltados:

- A possibilidade, de mesmo para a discriminação de vozes saudáveis e patológicas, de se utilizar categorias que englobem as patologias de acordo com suas características acústicas mais relevantes, de modo que a comparação entre saudável e patológica possa ser vista mais especificadamente de acordo com as categorias.
- A maneira que alguns resultados se mostraram indicam que as medidas possam estar, de alguma maneira, sobrepostas, dificultando assim a correta classificação das patologias. E portanto, as medidas aqui utilizadas se mostraram insuficientes para uma boa classificação.
- Algumas estimadores resultaram em valores “NaN” (“Not a Number”), e mesmo depois de recalculados, alguns valores não puderam ser corrigidos. Isso pode ter acarretado num pior desempenho do estimador, dada a quantidade de valores “NaN” nele encontrados.

Para trabalhos futuros podem ser enumerados 2 tópicos principais, que podem resultar numa melhora do desempenho da RNA:

- A possibilidade de uma filtragem no sinal, para que este tenha sua qualidade melhorada, e de maneira que possa ser vista sua influência no resultado final. O filtro de Kalman pode ser uma alternativa para este estudo.
- A divisão das patologias em categorias de acordo com suas características mais relevantes, e sua comparação com as vozes saudáveis e patológicas.

Referências

- ABC.MED.BR, 2013. **Laringoscopia: o que é? Como é realizada? Quem deve e quem não deve fazer?**. Disponível em: <<http://www.abc.med.br/p/exames-e-procedimentos/359364/laringoscopia-o-que-e-como-e-realizada-quem-deve-e-quem-nao-deve-fazer.htm>>. Acesso em: 8 mai. 2016
- BENITE, M. **Aplicação de modelos de redes neurais na elaboração e análise de cenários macroeconômicos**. Dissertação (Mestrado em Engenharia de Produção) Escola de Engenharia de São Carlos Universidade de São Paulo, São Paulo, 2003.
- BRANDÃO, A. CATALDO, E. LETA, F. **Um Novo Método Usando Autocorrelação para Extração da Frequência Fundamental em Sinais de Voz**. Artigo. Universidade Federal Fluminense, Niterói, 2007
- CARANHA, A. **Sistema de Pesquisa de músicas através de solfejo com foco em músicas brasileiras**. 2013. 175f. Dissertação (Mestrado em Engenharia Elétrica). Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2013.
- CESAR, D. **Figura1** Disponível em <<http://anatomiaonline.com/respiratorio/laringe/laringe.htm>>. Acesso em: 18 set 2016.
- CIELO, C.; FINGER, L.; *Lesões organofuncionais do tipo nódulos, pólipos e edema de Reinke*. Rev. CEFAC, São Paulo ,2010.
- COSTA, D. G.. **Uso de Gráfico de Recorrência e Redes Complexas para caracterização de Sistemas Complexos de Baixa Dimensão**. Dissertação (Mestrado em Computação Aplicada) INPE. São José dos Campos. 2014
- DALTER.V. **Algoritmos para estimação de Frequência Fundamental**. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação). Universidade Estadual de Londrina. Londrina, 2016.
- EAV. Engenharia Audiovisual. **Figura 4**. Disponível em:< <http://www.eav.eng.br/tech/fisiologia/Constituidoaparelhofonadorhumano.html#Topic11>> Acesso em: 2 out 2016.
- FILHO, E. K. **Um estudo sobre o método da Mistura de Gaussianas para a formação de grupos de dados**. Dissertação (Mestrado em Ciências). Instituto Tecnológico de Aeronáutica, São José dos Campos, 2009.

FUKUYAMA, E. **Análise acústica da voz captada na faringe próximo à fonte glótica através de microfone acoplado ao fibrolaringoscópio.** Rev. Bras. Otorrinolaringol.vol.67,no.6, São Paulo,Nov. 2001

GERHARD, D. **Pitch Extraction and Fundamental Frequency: History and Current Techniques.** Artigo. Department of Computer Science. University of Regina. Regina. 2003.

GONZAGA, A. **Métodos de Avaliação de Classificadores.** Aula 9.

GONZALEZ, J. **Sistemas Inteligentes.** <<http://slideplayer.com.br/slide/8068543/>>
Acesso em: 20 maio 2017.

HAYKIN, Simon. **Redes neurais: princípios e prática.** trad. Paulo Martins Engel. - 2.ed. - Porto Alegre: Bookman, 2001.

INFOESCOLA.**Figura 10** Disponível em: <http://www.infoescola.com/biologia/tecido-nervoso/>. Acesso em: 6 nov 2016.

Introdução ao Estudo de Sistemas Dinâmicos 1 Disponível em: https://www.ime.usp.br/~oda/contents/01Matem%20E1tica/01Sistemas%20Din%20micos/01_Introd.pdf> Acesso em: 2 out 2016.

JUNQUEIRA, Roberta R. A. **Alterações mínimas da laringe: Um diagnóstico diferencial.** 1999. Monografia (Especialização em Voz). Centro de Especialização em Fonoaudiologia Clínica - CEFAC, Rio de Janeiro, 1999.

LOPES, F. **Introdução ao Reconhecimento de Padrões e aplicações em problemas de Bioinformática** Curso de Verão - Bioinformática - USP, 2012 Disponível em :<
https://www.ime.usp.br/posbioinfo/cv2012/reconhecimentoPadroes_FabricioLopes.pdf>
Acesso em : 25 set 2016.

MACHADO, A. **Conversão de Voz Inter-Linguística.** 2013. 221f Tese (Doutorado em Ciências). Universidade de São Paulo, São Paulo, 2013.

MATOS, P. **Relatório Técnico “Métricas de Avaliação.”** Disponível em: <
<http://gbd.dc.ufscar.br>>. Acesso em 20 maio 2017.

MCLACHLAN, G.; PEEL, D. **Finite Mixture Models.** Hoboken, New Jersey: John Wiley & Sons, Inc. 2000.

MELLO, A. **Figura 2**. Disponível em [http://vozinstrumento.blogspot.com.br/p /conhece-ndo-voz.html](http://vozinstrumento.blogspot.com.br/p/conhece-ndo-voz.html). Acesso em: 6 nov 2016.

MINITI, A. BENTO, R. BUTUGAN, O. **Tratado de Otonorrinolaringologia da Sociedade Brasileira de Otonorrinolaringologia**. 1 ed. Roca. 2003

MONTEIRO, J. **Artificial Neural Networks a Method of Behavioral Representation**. Artigo. 5º Congresso Internacional de Gestão da Tecnologia e Sistema de Informação, Universidade Paulista, São Paulo, 2008.

MORI, Ricardo N. **Perfis Acústicos para Caracterização de Patologias da Laringe**. 2005. 82f. Tese (Doutorado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, 2005.

OLIVEIRA, C.; VIEIRA, F.; SOUZA, M.;BORGES, M.; **Aplicação de Misturas Gaussianas na Análise e Modelagem de Tráfego VoIP** – Artigo - Escola de Engenharia Elétrica e de Computação, Goiânia, 2005.

OPPENHEIM, Alan V.; WILLSKY, Alan S. **Sinais e sistemas**. 2. ed. São Paulo, SP: Pearson, 2010.

OPPENHEIM, Alan V.; SCHAFER, Ronald W. **Processamento em tempo discreto de sinais**. 3. ed. São Paulo, SP: Pearson Education do Brasil, 2013.

PINHO, P.; COURAS, M.;DANTAS, E.; COSTA, S. **Classificação de patologias laríngeas por meio de características do espaço de fase reconstruído**. XXXIV Simpósio Brasileiro de Telecomunicações, Santarém, 2016.

PINTO, A.; YNOGUTI, C. **Reconhecimento de locutor usando misturas de gaussianas**. Artigo. Inatel- Instituto Nacional de Telecomunicações.

PISTORI, H. **Introdução à aprendizagem estatística com Modelos de Misturas de Gaussianas**. Pesquisa, desenvolvimento e inovação em Visão Computacional, Universidade Católica Dom Bosco, Campo Grande.

RAZERA, Daniel E. **Determinadores de Pitch**. 2004. 97f. Dissertação (Mestrado em Engenharia Elétrica) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, 2004.

ROSA, Marcelo O. **Análise Acústica da Voz para Pré-diagnóstico de Patologias da Laringe**. 1998. 261f. Dissertação (Mestrado em Engenharia Elétrica) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, 1998.

SANTANA, A. **Treinamento de Redes Neurais Artificiais Utilizando Algoritmos Genéticos em Plataforma Distribuída**. Monografia (Graduação em Sistemas de Informação). Universidade Federal de Lavras. 2012.

SCALASSARA, Paulo R. **Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias da Laringe**. 2009. 265f. Tese (Doutorado em Engenharia Elétrica) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, 2009.

SILVA, E.; OLIVEIRA, A.. **DICAS PARA A CONFIGURAÇÃO DE REDES NEURAIS**. Disponível em: < http://equipe.nce.ufrj.br/thome/grad/nn/mat_didatico/dicas_configuracao_rna.pdf>. Acesso em: 16 out. 2016.

SODRÉ, B. **Reconhecimento de Padrões Aplicados a Identificação de Patologias da Laringe**. 2016. 187f. Dissertação. (Mestrado em Engenharia Elétrica). Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

SOUZA, D. **Figura 3** Disponível em http://www.dgsotorrinolaringologia.med.br/APOST_LARINGE.html. Acesso em: 7 set 2016.

SOUZA, E. **Caracterização de Sistemas Dinâmicos através de Gráficos de Recorrência**. Dissertação (Mestrado em Física) Universidade Federal do Paraná, Curitiba, 2008.

STANGE, R.; NETO, J. **Reconhecimento de Padrões em Classificadores: Comparação de Técnicas e Aplicações**. 4º Workshop de Tecnologia Adaptativa – WTA'2010

VIEIRA, J; COSTA, S.; COSTA, W. **Análise de Quantificação de Recorrência e Análise Discriminante Aplicadas à Classificação de Sinais de Vozes Saudáveis e Sinais de Vozes Patológicas**. Artigo VII CONNEPI2012. ISBN 978-85-62830-10-5

VIEIRA, V. **Avaliação de Distúrbios da Voz por meio de Análise de Quantificação de Recorrência**. Dissertação (Mestrado em Engenharia Elétrica) Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, 2014.