

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
ENGENHARIA DE CONTROLE E AUTOMAÇÃO

NICOLAS BASSO LENS

**CLASSIFICADOR NEURAL DE VOZES RUGOSAS**

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO  
2019

NICOLAS BASSO LENS

## **CLASSIFICADOR NEURAL DE VOZES RUGOSAS**

Trabalho de Conclusão de Curso de graduação do curso de Engenharia de Controle e Automação da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para a obtenção do título de Bacharel.

Orientadora: Prof. Dra. María Eugenia Dajer

Coorientador: Prof. Dr. Danilo Hernane Spatti

CORNÉLIO PROCÓPIO  
2019



**Universidade Tecnológica Federal do Paraná**  
**Campus Cornélio Procópio**  
**Departamento Acadêmico de Elétrica**  
**Curso de Engenharia de Controle e Automação**



**FOLHA DE APROVAÇÃO**

**Nicolas Basso Lens**

**Classificador neural de vozes rugosas**

Trabalho de conclusão de curso apresentado às 10:20hs do dia 25/04/2019 como requisito parcial para a obtenção do título de Engenheiro de Controle e Automação no programa de Graduação em Engenharia de Controle e Automação da Universidade Tecnológica Federal do Paraná. O candidato foi arguido pela Banca Avaliadora composta pelos professores abaixo assinados. Após deliberação, a Banca Avaliadora considerou o trabalho aprovado.

---

Prof(a). Dr(a). María Eugenia Dajer - Presidente (Orientador)

---

Prof(a). Dr(a). Danilo Hernane Spatti - (Coorientador)

---

Prof(a). Dr(a). Wagner Endo - (Membro)

---

Prof(a). Dr(a). Cristiano Marcos Agulhari - (Membro)

Dedico este trabalho à minha família, que sempre esteve ao meu lado me incentivando e acreditando em mim, aos meus amigos que tornaram o caminho menos árduo e à minha namorada que mesmo nos momentos de ausência é paciente e torce pelo meu sucesso.

## **AGRADECIMENTOS**

Agradeço a Deus por me prover de saúde e força sem as quais eu não seria capaz de superar todas as dificuldades encontradas no caminho em busca do meu objetivo.

Agradeço ao meu pai Aduino, minha mãe Cláudia e meu irmão Thomas por serem a base de tudo, pelo apoio incondicional e compreensão demonstrada em todos os momentos, principalmente em momentos importantes em que estive distante.

Agradeço à minha orientadora, Profa. Dra. Maria Eugenia Dajer que com enorme paciência me guiou e me ensinou lições valiosas que levarei comigo pelo novo trajeto que estarei seguindo.

Agradeço à minha companheira Nadine, que pacientemente me espera do outro lado do oceano, me passando confiança e me acalmando em momentos de dificuldade.

Agradeço também a todos os meus amigos, Abe, Alcatra, Bilico, Dani, Douglinhas, Japa, Juliano, Renato, Paludeto, Rafa, Renan, Sakuno, Túlio e Zóio por participarem de momentos inesquecíveis e não deixar que o estresse me abalasse.

Por fim, agradeço aos meus demais colegas de graduação, familiares e professores, que contribuíram direta ou indiretamente com a minha formação.

A todos vocês, muito obrigado, essa conquista é nossa!

**Quando olharmos para o futuro, veremos  
que os líderes serão os que favorecem o  
desenvolvimento de todos os outros.  
Bill Gates**

## RESUMO

LENS, Nicolas Basso. **Classificador neural de vozes rugosas**. 2019. 48 f. Trabalho de Conclusão de Curso (Graduação) – Engenharia de Controle e Automação. Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2019.

A voz humana é um dos principais meios de comunicação, e também a principal ferramenta de trabalho de diversos profissionais. Tendo em vista as dificuldades que existem em diagnósticos de avaliação vocal, e a necessidade de cuidados com a voz, o presente estudo propõe um método alternativo e não-invasivo para auxiliar os profissionais da área de fonoaudiologia, utilizando redes neurais artificiais para a classificação de vozes humanas com características de rugosidade. A extração das características do sinal em análise tempo-frequência será realizada por intermédio da transformada *wavelet packet*, e a rede neural utilizada para a classificação do distúrbio será a *Perceptron* de múltiplas camadas. O classificador desenvolvido atingiu uma alta porcentagem de acertos, mostrando-se robusto através dos testes realizados, provando ser uma alternativa promissora para uma aplicação futura como método não-invasivo que auxilia os profissionais da voz.

**Palavras-chave:** Redes Neurais Artificiais. Rugosidade. Transformada Wavelet Packet. Classificador Neural. Avaliação Vocal.

## ABSTRACT

LENS, Nicolas Basso. **Neural classifier of roughness voices**. 2019. 48 f. Trabalho de Conclusão de Curso (Graduação) – Engenharia de Controle e Automação. Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2019.

The human voice is one of the main means of communication, it is also the main tool of work of several professionals. Considering the difficulties that exist in voice assessment diagnoses and the need for voice care, the present study proposes an alternative and non-invasive method to assist professionals in the speech therapy area using artificial neural networks to classify human voices with roughness characteristics. The extraction of the characteristics of the signal in time-frequency analysis will be performed through the wavelet packet transform, and the neural network used for the disturbance classification will be multiple layers Perceptron. The developed classifier achieved a high percentage of correct answers, proving to be a promising alternative for future applications as a non-invasive method that assists speech therapy professionals.

**Keywords:** Neural Networks. Roughness. Wavelet Packet. Neural Classifier. Voice Assessment.



## LISTA DE FIGURAS

Figura 1 – Esquemático do trato vocal.....	12
Figura 2 – Exemplo de videoestrobolaringscopia.....	14
Figura 3 – Ambiguidade na amostragem.....	17
Figura 4 – Exemplo de <i>upsampling</i> .....	18
Figura 5 – Exemplo de <i>downsampling</i> .....	19
Figura 6 – Transformada <i>wavelet</i> .....	20
Figura 7 – Decomposição da transformada <i>wavelet packet</i> . Aproximação (A), Detalhe (D).....	20
Figura 8 – Exemplo de rede PMC.....	25
Figura 9 – Pré-processamento dos arquivos.....	27
Figura 10 – Processo de treinamento e aprendizagem.....	28
Figura 11 – Janelamento simples.....	29
Figura 12 – Janelamento cruzado.....	29
Figura 13 – Processo de preparação de dados.....	30
Figura 14 – Processo de treinamento.....	33
Figura 15 – Resultados dos melhores treinamentos.....	34
Figura 16 – Definição da melhor topologia.....	34
Figura 17 – Robustez da melhor topologia.....	35
Figura 18 – Desempenho da RNA.....	37
Figura 19 – Erros cometidos.....	37

## LISTA DE QUADROS

Quadro 1 – Potenciais aplicações das RNAs.....	23
Quadro 2 – Composição geral do banco de dados.....	26
Quadro 3 – Características gerais da RNA desenvolvida.....	32
Quadro 4 – Variações de topologia.....	32
Quadro 5 – Critérios pós-processamento.....	36
Quadro 6 – Matriz de confusão.....	38

## **LISTA DE SIGLAS, ACRÔNIMOS E ABREVIATURAS**

TWP	Transformada Wavelet Packet
RNA	Rede Neural Artificial
PDS	Processamento Digital de Sinais
STFT	Short Time Fourier Transform
PMC	Perceptron de Múltiplas Camadas
SINPRO- SP	Sindicato de professores de São Paulo

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>8</b>
<b>2</b>	<b>OBJETIVOS.....</b>	<b>10</b>
2.1	OBJETIVO GERAL.....	10
2.2	OBJETIVOS ESPECÍFICOS.....	10
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>11</b>
3.1	ESTUDOS DA VOZ.....	11
3.1.1	Produção da voz.....	11
3.1.2	Distúrbios da voz.....	12
3.1.2.1	Rugosidade.....	13
3.1.3	Métodos de avaliação vocal.....	14
3.1.3.1	Métodos Invasivos.....	14
3.1.3.2	Métodos Não Invasivos.....	16
3.2	PROCESSAMENTO DIGITAL DE SINAIS.....	16
3.2.1	Amostragem.....	17
3.2.2	Transformada Wavelet Packet.....	21
<b>3.3</b>	<b>SISTEMAS INTELIGENTES.....</b>	<b>22</b>
3.3.1	Redes Neurais Artificiais (RNA).....	22
3.3.2	Aplicações das RNA.....	24
3.3.3	Treinamento e aprendizagem.....	25
3.3.4	Redes Perceptron de Múltiplas Camadas (PMC).....	26
<b>4</b>	<b>MATERIAIS E MÉTODOS.....</b>	<b>28</b>
4.1	MATERIAIS.....	28
4.1.1	Banco de dados.....	28
4.1.2	Ferramentas.....	29
4.2	MÉTODOS.....	29
4.2.1	Construção de amostras e janelamento.....	30
4.2.2	Aplicação da Transformada Wavelet Packet e extração de características.....	32
4.2.3	Processos de treinamento e aprendizagem.....	33
4.2.4	Determinação da melhor topologia.....	35
4.2.4.1	Robustez da melhor topologia.....	37
<b>5</b>	<b>RESULTADOS E DISCUSSÕES.....</b>	<b>38</b>
5.1	PÓS-PROCESSAMENTO E VALIDAÇÃO DOS RESULTADOS.....	38

5.2	MATRIZ DE CONFUSÃO.....	40
6	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>41</b>
	<b>REFERÊNCIAS.....</b>	<b>42</b>

## 1 INTRODUÇÃO

Todo ser humano possui uma voz única que, ela é uma ferramenta de comunicação que carrega traços de sua faixa etária, sexo, tipo físico, personalidade e estado emocional. Para alguns, no entanto, ela representa muito mais do que isso. Professores, atores, repórteres, cantores e outros profissionais têm na voz uma indispensável ferramenta de trabalho, e precisam estar atentos aos cuidados que devem adotar para não a prejudicar (OLIVEIRA, 2013). A voz molda a transmissão da mensagem articulada, adicionando à palavra o conteúdo emocional, a entoação, a expressividade, identificando, assim, o indivíduo (PENTEADO, TEIXEIRA & PEREIRA, 2006).

Sendo a voz uma ferramenta primária das mais imediatas que o ser humano dispõe para interagir com a sociedade (BEHLAU, 2004). Tem-se que o uso inadequado de tal ferramenta acarretará em problemas não só biológicos, mas também psicológicos e, no caso de profissionais da voz, financeiros. A Associação Brasileira de Otorrinolaringologia e Cirurgia Cérvico-Fácil (2004) afirma que, no Brasil, o gasto anual destinado a lidar com os afastamentos de professores em função de problemas na voz atinge valores exorbitantes.

Diversas doenças ligadas ao sistema de produção da voz podem ocorrer desde o nascimento até a idade mais avançada. Entre elas encontram-se inflamações, inchaços, tumores benignos e malignos, nódulos, cistos e pólipos entre outras alterações. Atualmente a avaliação destas alterações é realizada através de lentes chamadas de endoscópios, acopladas a câmeras e outros equipamentos, que são introduzidas pela boca ou nariz, com o paciente acordado, sem anestesia ou sedação nos adultos e na maioria das crianças (KRUG, 2015). Contudo, além de invasivas e desconfortantes, as técnicas convencionais possuem custo elevado e demandam muito tempo para realização e obtenção de resultados (GHASEMZADEH et al., 2015).

Métodos não-invasivos, embora ainda não desempenhem papel prioritário no procedimento de diagnóstico, tem o potencial de fornecer medidas objetivas, como valores numéricos ou computadorizados associados à voz. Esses valores podem oferecer suporte quantitativo para a tomada de decisão do profissional da área, gerando, portanto, maneiras para superar os obstáculos enfrentados pela fonoaudiologia.

Para o trabalho proposto será utilizada uma ferramenta já consagrada na área de processamento digital de sinais, conhecida como transformada *wavelet packet* (TWP). Como é abordado por Stollnitz (1995), funções *wavelet* e suas transformadas matemáticas para análise e decomposição hierárquica de funções, ou seja, representação da função em diferentes níveis de resolução, ou escalas. A decomposição de uma função *wavelet* é conhecida como transformada *wavelet* e possui suas variantes contínua e discreta. Sendo capazes de decompor funções tanto no domínio da frequência quanto no domínio do tempo, as funções *wavelet* são ferramentas poderosas de processamento de sinais e muito promissoras para trabalhos envolvendo sinais biológicos (OLIVEIRA, 2007).

Para o processo de aprendizado e automatização da identificação, utiliza-se redes neurais artificiais como um método para a estimação de parâmetros. Segundo Haykin (2001), uma rede neural é uma máquina que é projetada para modelar a maneira como o “cérebro” realiza uma tarefa particular ou função de interesse, a rede é normalmente implementada utilizando-se componentes eletrônicos ou é simulada por programação em um computador.

Considerando as tecnologias descritas, o presente trabalho busca a implementação de um classificador neural que seja capaz de diferenciar e classificar amostras de áudio que indiquem o distúrbio vocal denominado rugosidade, auxiliando portanto, os profissionais da voz a averiguarem as condições vocais do paciente e determinar as medidas cabíveis para cada caso.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

O objetivo do trabalho é desenvolver um classificador inteligente como alternativa para auxiliar o profissional da área de fonoaudiologia no diagnóstico de detecção de um distúrbio vocal através de análises acústicas não invasivas, utilizando a transformada *Wavelet Packet* e auxiliado por redes neurais artificiais.

### 2.2 OBJETIVOS ESPECÍFICOS

De modo mais detalhado, os objetivos podem ser descritos como:

- Preparar o banco de dados, assegurando que o mesmo contenha diferentes sinais de voz;
- Segmentar os arquivos de audio que compõem o banco de dados em amostras menores, tornando o processamento dos mesmos mais eficaz;
- Propor uma abordagem de detalhamento do sinal em análise tempo-frequência através da transformada *wavelet packet*;
- Tornar o processo de treinamento automatizado e robusto;
- Validar a capacidade de classificação da rede neural obtida;
- Analisar e discutir os resultados obtidos;



### 3 FUNDAMENTAÇÃO TEÓRICA

Considerando a multidisciplinaridade do trabalho proposto, o presente capítulo será dividido em três partes, sendo a primeira dedicada à uma breve introdução teórica dos conceitos da produção da voz humana, dos distúrbios vocais e seus métodos de avaliação, na segunda parte será abordada a teoria de processamento digital de sinais que será aplicada ao projeto e por fim serão abordados conceitos de sistemas inteligentes, com maior foco em redes neurais artificiais (RNA).

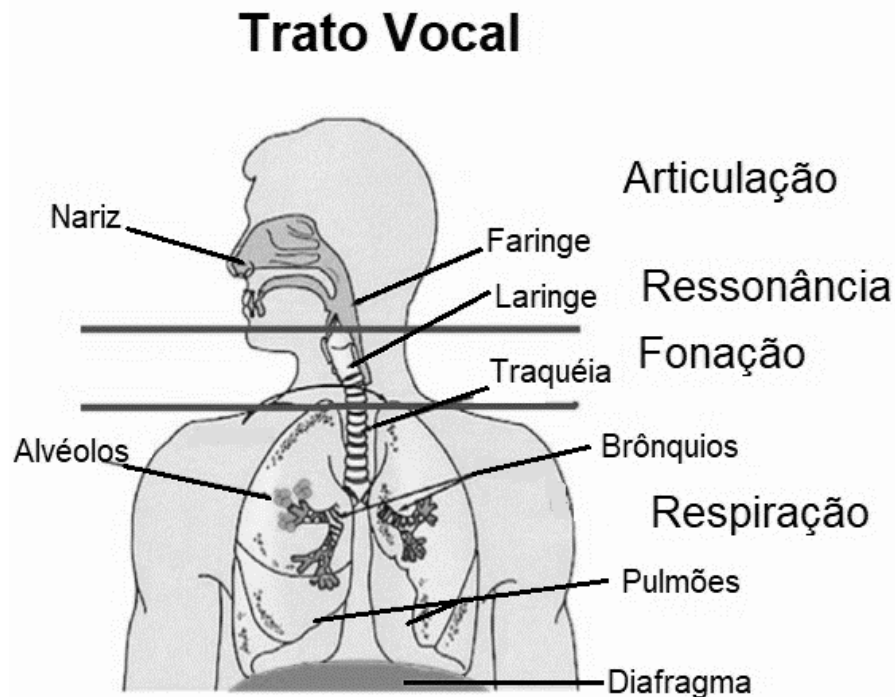
#### 3.1 ESTUDOS DA VOZ

##### 3.1.1 Produção da voz

Os seres humanos possuem diversas habilidades que lhes distinguem dos demais e os permitem atingir conquistas não obtidas por outros seres vivos. Uma destas habilidades é a fala, a fala é a expressão do pensamento manifestada por meio da voz, que é o resultado da passagem do ar pela laringe fazendo vibrar as pregas ou cordas vocais. A laringe, de acordo com Pinho, Tsuji e Bohadana (2006), é um órgão extremamente complexo, responsável por diversas atividades fisiológicas, entre elas a respiração, a fonação e a proteção, e atua em conjunto com outros órgãos, formando o aparelho fonador humano.

A laringe está localizada na região cervical do corpo humano entre a traqueia e a base da língua. Ela possui características fibromusculares, várias cartilagens dentre elas a tiroide, cricoide e a epiglote, todas revestidas com as membranas mucosas que são impulsionadas pelos músculos da laringe, onde as dobras das membranas mucosas geram as pregas vocais, que se localiza na parte interna da laringe e que promove através dos seus movimentos o som (voz) que usamos para nos comunicar (PHILLIPI, 2009). A Figura 1 ilustra o esquemático do trato vocal e seus principais elementos.

Figura 1 – Esquemático do trato vocal.



Fonte: Adaptado de Miranda (2016).

### 3.1.2 Distúrbios da voz

Como discute Gayotto (2006), eufonia é o nome dado à uma voz saudável, aquela onde há o equilíbrio entre a força do ar e da musculatura das cordas vocais. Quando falta esse equilíbrio, tem-se uma voz que é chamada de disfonia. A disfonia não é uma doença, mas um sintoma, uma manifestação de um mau funcionamento de algum dos sistemas ou estruturas que atuam na produção da voz.

Realizando uma analogia, sabe-se, por exemplo, que um atleta olímpico deve dedicar anos de treinamentos específicos para sua respectiva modalidade para que possa obter um bom desempenho, e que os atletas despreparados fisicamente estão mais propensos a lesões. O uso da voz também é uma atividade física e pode chegar à demanda energética e muscular do porte de um atleta olímpico, dependendo da atividade vocal (como por exemplo um cantor de ópera), e assim como os cuidados com um atleta exigem conhecimentos específicos, os cuidados com o profissional da voz também exige conhecimento específico (CAMPOS, 2002).

De acordo com Behlau (2004), um dos temas mais subjetivos de sua área, é definir as distinções entre uma voz normal e uma voz disfônica, pois pode-se

apresentar diferentes percepções sobre o grau de qualidade vocal do paciente. Um dos distúrbios mais recorrentes em vozes é a rugosidade. A rugosidade tem como característica, a voz com aspecto trêmulo, como se houvesse acúmulo de mucosa dificultando a passagem do ar.

### 3.1.2.1 Rugosidade

O distúrbio vocal estudado no presente trabalho é denominado rugosidade, e como Anand, Skowronski, Shrivastav e Eddins (2017) e Zitta (2005) sugerem, a rugosidade se trata do acréscimo de tremulações na voz, produzindo ruídos auditivos de baixa frequência e, nos diagramas espectrográficos, mascarando os harmônicos inferiores. De maneira mais simplista, a rugosidade pode ser definida como a irregularidade percebida através das vibrações das pregas vocais.

Para que o classificador proposto atinja os objetivos do trabalho, os autores, Omori, Kojima, Kakani, Slavik e Blaugrund (1996) indicam que através de análises no espectro de energia pode-se encontrar parâmetros para a classificação de vozes rugosas mesmo em casos com *jitter* e *shimmer* normais. O projeto deste trabalho utilizará o grau de energia para realizar a classificação dos arquivos de áudio.

### 3.1.3 Métodos de avaliação vocal

Os autores Gragone e Nagano (2004) também fazem referência à importância de buscar profissionais de saúde responsáveis pelo cuidado da voz, de acordo com os autores, o otorrinolaringologista é o médico especialista responsável pelo tratamento clínico e cirúrgico, e o fonoaudiólogo é o profissional que realizará a avaliação do perfil vocal, o que inclui uma série de testes. Entre os testes a serem realizados, existem os que se enquadram nos métodos invasivos e não invasivos.

#### 3.1.3.1 Métodos Invasivos

Os métodos invasivos de avaliação vocal são baseados em imagens internas da laringe do paciente, obtidas, por exemplo, através de exames como a videolaringoestroboscopia, tais exames permitem avaliar as condições das pregas

vocais dos pacientes eficientemente, porém causam grande desconforto ao indivíduo que está sendo examinado.

Segundo González (2008), devido à velocidade, o movimento vibratório das pregas vocais não pode ser perceptível ao olho humano, sendo necessário o auxílio de uma micro câmera com capacidade de captura em alta velocidade em conjunto com um aparato de luz etroboscópica que será inserida via oral no paciente. Grinfeld (2017) explica que além de um custo elevado, o exame é de difícil realização em pessoas não colaborativas, pois normalmente (salvo casos de pacientes na faixa etária pediátrica) é realizado com o paciente acordado e sem o uso de qualquer tipo de anestesia. A Figura 2 ilustra um exemplo de um exame de videolaringoestroboscopia.

**Figura 2 – Exemplo de videoestrobolaringoscopia.**



**Fonte: Adaptado de González (2008).**

### 3.1.3.2 Métodos Não-Invasivos

Os métodos não invasivos por sua vez, são exames que não necessitam da inserção de alguma ferramenta no corpo do paciente, e de acordo com Freitas (2010) são definidos da seguinte forma:

- **Avaliações áudio-perceptuais:** o profissional deve julgar auditivamente uma amostra vocal do indivíduo. Em geral, tais amostras são compostas por vogais sustentadas e/ou falas encadeadas, e ficam registradas em arquivos de áudio ou vídeo para que possa ser realizada a análise completa.
- **Avaliações acústicas:** possibilitam a obtenção dos dados automaticamente através de um processamento computacional, essas avaliações permitem os profissionais a quantificar, mensurar e descrever as características da voz, e podem agir em conjunto com as análises áudio-perceptuais, demonstrando uma visão mais detalhada sobre o processo de geração do sinal de voz e suas modificações e disfonias.

### 3.2 Processamento digital de sinais

Durante as últimas décadas, a área de processamento digital de sinais (PDS) tem crescido e se mostrado importante, tanto na teoria quanto nas aplicações práticas tecnológicas. Uma das principais razões para esse sucesso na indústria, é o desenvolvimento e uso de softwares e hardwares de baixo custo (INGLE; PROAKIS, 2010).

Nesta seção, serão abordados os conceitos de amostragem (*sampling*) e a transformada *wavelet packet*, esta última será responsável na transformação dos sinais do banco de dados para satisfazer os requisitos do projeto.

#### 3.2.2 Amostragem

O mundo em que vivemos é analógico, porém os sinais que nos interessam são contínuos. Considerando que nossos computadores são digitais, ou seja, eles armazenam somente dados discretos e finitos, se faz necessário aproximar os sinais analógicos e digitais antes de partir para um processo de armazenamento ou manipulação de dados.

Segundo Zayed (1993), a amostragem tem como principal função, na área de comunicação e processamento de sinais, a transferência entre analógico e digital de um sinal. Por sua vez, Manolakis e Ingle (2011) comentam que o modo mais comum para realizar a amostragem é dito periódico, ou seja, a partir de um sinal analógico serão obtidas amostras, extraindo os valores em instantes igualmente divididos no tempo.

Por fim, Weeks (2012) diz que o período de amostragem  $T_s$  é o intervalo entre duas amostras, que, por sua vez, é exatamente o inverso da frequência de amostragem  $F_s$ , essa relação é demonstrada pela equação (1).

$$T_s = \frac{1}{F_s} \quad (1)$$

Reorganizando a equação (1), temos:

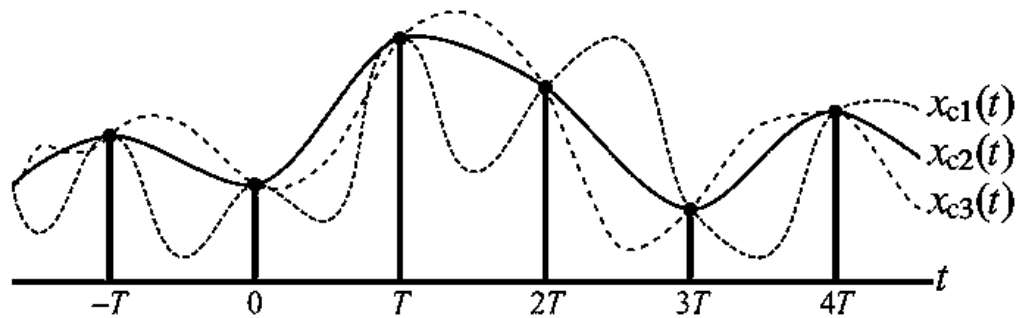
$$F_s = \frac{1}{T_s} \quad (2)$$

Pode-se então definir que,

- $F_s$  é a frequência de amostragem (ciclos por segundo ou Hertz) ou, a taxa de amostragem (amostras por segundo);
- $T_s$  é o período de amostragem (segundos), que delimita o intervalo entre duas amostras.

Porém, como explanam Manolakis e Ingle (2011), um número infinito de sinais podem gerar a mesma combinação de amostras, e a única maneira de se obter o valor ideal é definir algumas restrições sobre o comportamento do sinal em tempo contínuo, e a resposta para essas questões está na relação entre os espectros contínuo e discreto do sinal desejado. A Figura 3 demonstra um caso não ideal, onde três sinais diferentes com o mesmo período de amostragem resultam em amostras idênticas.

Figura 3 – Ambiguidade na amostragem.



Fonte: Adaptado de Manolakis e Ingle (2011).

Ingle e Proakis (2010), descrevem o intervalo de frequências de um sinal como largura de banda (*bandwidth*), e indicam como ideal, o uso do critério de Nyquist. O critério de Nyquist relaciona uma frequência de amostragem ideal com a largura de banda, para que o sinal possa ser reconstruído corretamente no futuro, a equação (3) representa o critério de Nyquist:

$$F_s \geq 2B \quad (3)$$

Onde,

- $F_s$ , é a frequência de amostragem;
- $B$  é a largura de banda.

Baseadas no critério de Nyquist, Weeks (2012) descreve duas situações problemáticas, nomeadas sobreamostragem (*oversampling*) e subamostragem (*undersampling*).

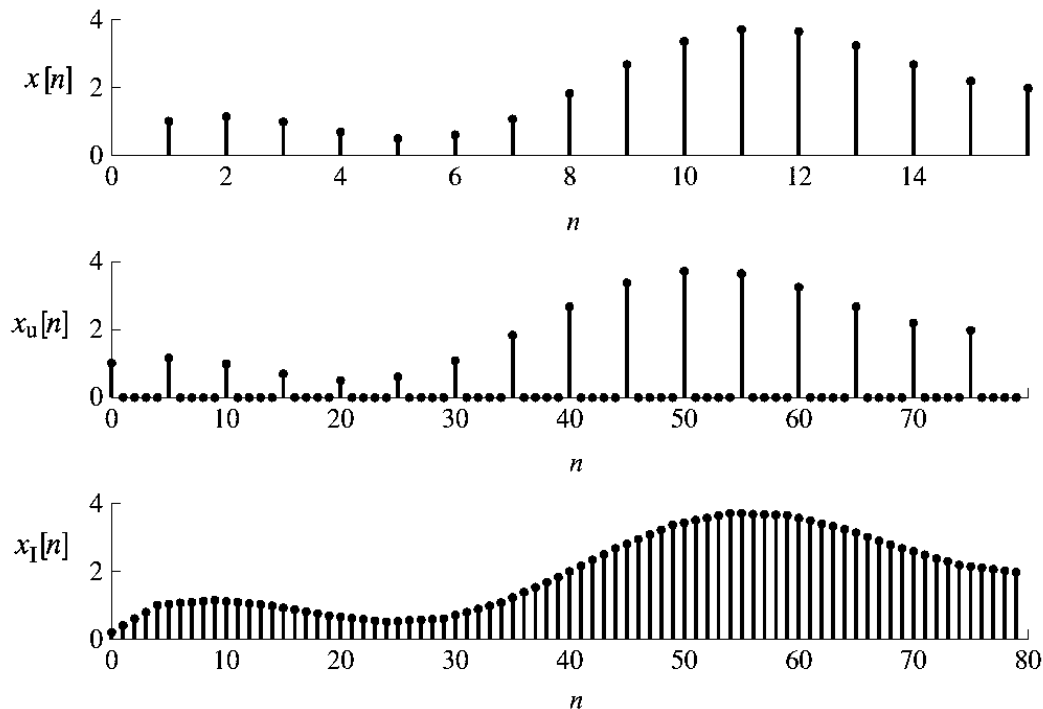
- **Sobreamostragem (*oversampling*):** essa situação representa a coleta de amostras que possuem frequência maior do que o necessário (maior do que duas vezes a largura de banda), tendo como resultado dados difusos que aumentam o tempo necessário de processamento do sinal e a necessidade de memória para armazenamento.
- **Subamostragem (*undersampling*):** é o contrário da sobreamostragem, representa a coleta de amostras com frequência menor do que o necessário (ou menos do que duas vezes a largura de banda), resultando em dados errôneos e inabilitando a reconstrução do sinal.

Por fim, o último conceito que será abordado dentro do tópico de amostragem é a conversão da taxa de amostragem, este conceito possibilita que o banco de dados possua uma única taxa de amostragem. Conforme Oppenheim e Schafer (1989) definem, existem dois métodos utilizados para realizar a alteração da taxa de amostragem, e são classificados da seguinte forma:

- **Upsampling:** é uma técnica que permite aumentar a taxa de amostragem em  $N$  vezes, inserindo um número  $N-1$  de zeros entre as amostras originais do sinal;
- **Downsampling:** é uma técnica de redução da taxa de amostragem por um fator  $D$ , reduzindo a apenas uma amostra um segmento de  $D$  pontos.

As Figuras 4 e 5 ilustram exemplos de *upsampling* e *downsampling* respectivamente:

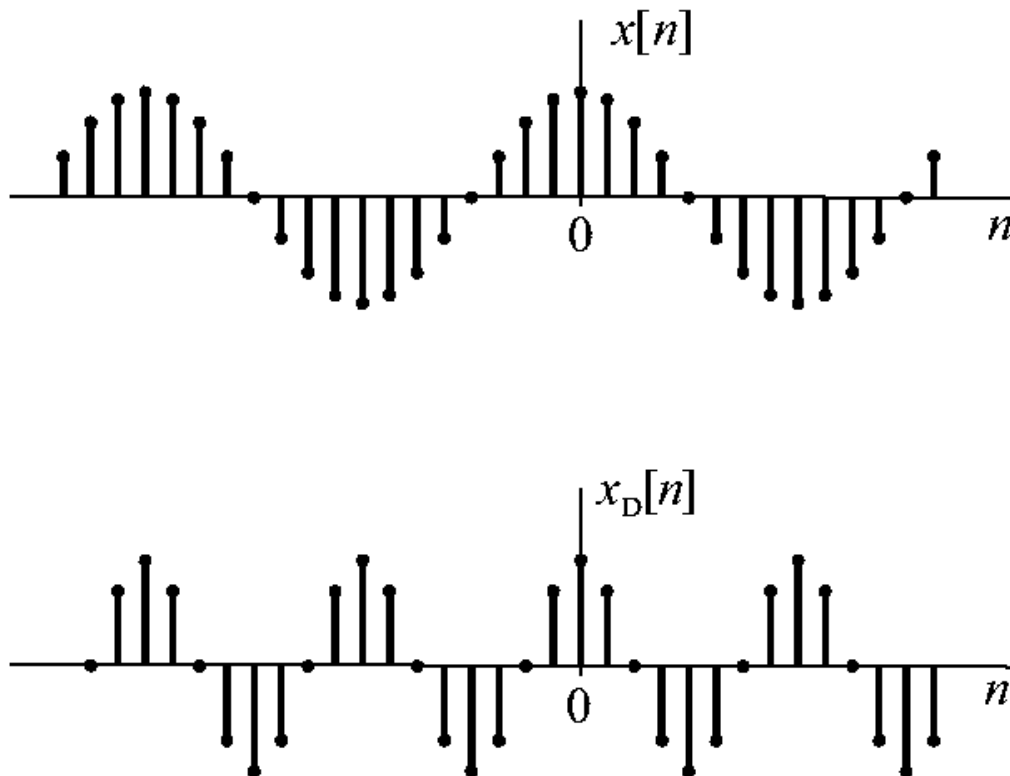
Figura 4 – Exemplo de *upsampling*.



Fonte: Adaptado de Manolakis e Ingle (2011).



Figura 5 – Exemplo de *downsampling*.



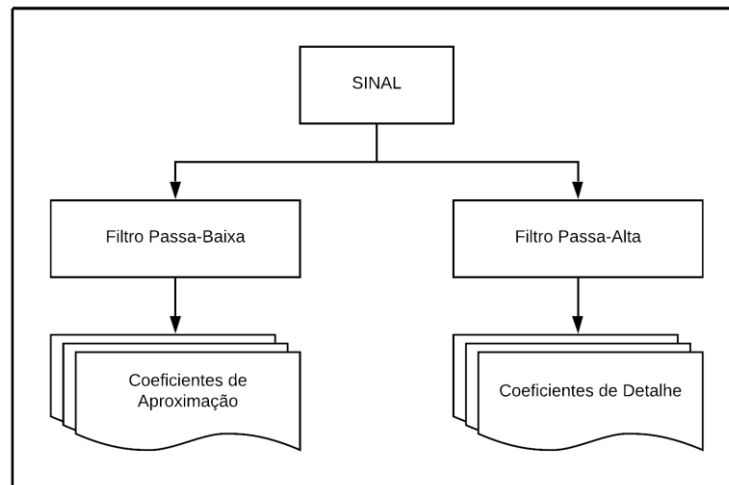
Fonte: Adaptado de Manolakis e Ingle (2011).

### 3.2.3 Transformada Wavelet Packet

A transformada *wavelet*, quando comparada com uma ferramenta muito utilizada, a transformada de Fourier de tempo curto ou, em inglês, *short time Fourier transform* (STFT), se mostra mais viável, pois ela permite um melhor precisão de tempo e frequência em casos não estacionários, tendo maiores resoluções de tempo em segmentos de alta frequência e melhores resoluções de frequência em segmentos de baixa frequência.

Conforme Prarraga (2002) aborda, a função *wavelet* está relacionada ao filtro passa-altas, produzindo os coeficientes de detalhe, e a função de escalonamento está relacionada ao filtro passa-baixas, produzindo os coeficientes de aproximação. Essa relação está ilustrada na Figura 6.

**Figura 6 – Transformada *wavelet*.**

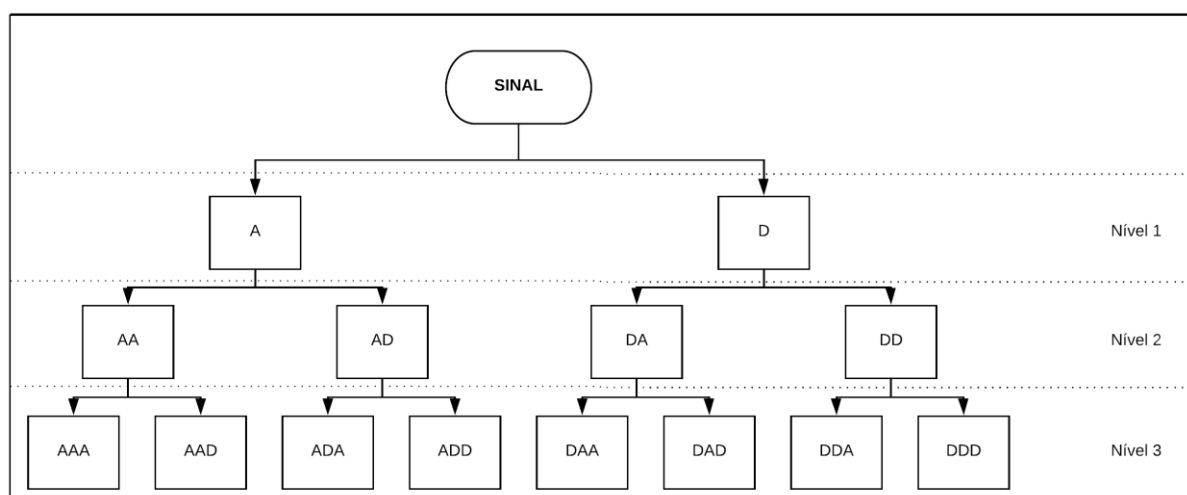


**Fonte: Autoria própria.**

Em busca de maior versatilidade, Graps (1995), sugere o uso da transformada *wavelet packet* (TWP), que é composta por uma combinação de diversas transformadas *wavelet* e como característica tem os seus coeficientes calculados por algoritmos recursivos, ou seja, os resultados da computação de cada coeficiente serão as raízes para os próximos cálculos.

A decomposição do sinal utilizando a TWP possui como vantagem o fato de as informações presentes no coeficiente de detalhe não serem ignoradas ao início de um novo nível, deste modo, ao realizar uma decomposição, o nível seguinte conterá mais dois coeficientes gerados a partir da decomposição do coeficiente de aproximação e do coeficiente de detalhe. A Figura 7 ilustra tal decomposição.

**Figura 7 – Decomposição da transformada *wavelet packet*. Aproximação (A), Detalhe (D).**



**Fonte: Autoria própria.**

### 3.3 SISTEMAS INTELIGENTES

Os sistemas inteligentes são utilizados com o intuito de criar uma inteligência artificial criada para resolver problemas inexatos, através de treinamentos, por possuírem algo que se denomina capacidade de aprendizagem.

Nesta seção, serão abordados conceitos de uma das ramificações do ramo de sistemas inteligentes, as redes neurais artificiais (RNA), bem como suas aplicações e características, mantendo o foco voltado para uma rede em específico, chamada de perceptron de múltiplas camadas (PMC), que será utilizada para a implementação do classificador.

#### 3.3.1 Redes Neurais Artificiais (RNA)

Segundo Haykin (1999), uma rede neural pode ser definida como um processador constituído de unidades de processamento simples, com a capacidade de adquirir conhecimento experimental através de um processo de aprendizagem, armazená-lo em matrizes de pesos sinápticos e disponibilizá-lo para uso.

Como algumas das principais características das RNA, Da Silva, Spatti e Flauzino (2010) citam:

- **Habilidade de generalização:** é a capacidade da rede de generalizar o conhecimento adquirido, de forma que soluções desconhecidas possam ser estimadas após o processo de treinamento;
- **Adaptação por experiência:** permite que a rede adquira conhecimento a partir da experimentação, ou seja através de uma sucessiva introdução de exemplos;
- **Tolerância a falhas:** por possuir um alto nível de interconexões entre neurônios artificiais, a RNA é um sistema que tolera falhas;
- **Capacidade de aprendizado:** através dos processos de treinamento, a RNA é capaz de relacionar múltiplas variáveis que constituem a aplicação;

- **Organização de dados:** a RNA é capaz de agrupar padrões que apresentam determinadas características em comum;
- **Facilidade de prototipagem:** redes neurais artificiais, de acordo com a área de aplicação, podem ser prototipadas em hardware ou software de maneira simples;
- **Armazenamento distribuído:** devido à essa característica, uma RNA se torna robusta, pois caso neurônios se tornem inoperantes, a mesma ainda possuirá conhecimento a respeito do comportamento de determinado processo.

### 3.3.2 Aplicações das RNA

Considerando suas características, as redes neurais artificiais são ferramentas de grande potencial para a resolução de problemas dos mais diversos ramos das ciências. Como citam Hagan (2014) e Da Silva, Spatti e Flauzino (2010), alguns exemplos e ramos de aplicação das redes neurais artificiais estão descritos no Quadro 1:

**Quadro 1 – Potenciais aplicações das RNAs.**

<b>Ramo</b>	<b>Aplicação</b>
Medicina	Classificação e predição de câncer, diagnóstico de doenças cardíacas, projeto de próteses
Química	Obtenção de novos compostos poliméricos, tratamento de água, classificador de produtos
Fala	Reconhecimento de fala, identificação de distúrbios vocais, classificação de patologias, compressão de som
Automotivo	Deteção de falhas, sistemas de controle de injeção de combustível, sistemas de frenagem automática
Militar	Identificação de ameaças, sonares, radares, controle de direção de projéteis, processamento de sinais
Robótica	Controle de trajetória, veículos autônomos, manipulação de controladores, sistemas de visão
Bancária	Previsão de caixa, estimação de risco de crédito, leitura de documentos, reconhecimento de assinaturas
Produção	Sistemas de inspeção, projeto e análise de produtos, controle de qualidade, estimação de processos
Aeroespacial	Simulação de falhas, desempenho de piloto automático, sistemas de controle de aeronaves

Fonte: Adaptado de Hagan (2014) e Da Silva, Spatti e Flazuino (2010).

### 3.3.3 Treinamento e aprendizagem

A habilidade mais importante das redes neurais artificiais, é a capacidade de aprendizagem a partir dos dados fornecidos, e com isso aumentar sua performance. O aprendizado é realizado através de um processo de ajustes aplicado aos pesos sinápticos denominado de treinamento, e acontece quando a rede neural chega a uma solução generalizada para uma classe de problemas.

Haykin (1999), discute de maneira geral, dois tipos de treinamentos para redes neurais artificiais:

- **Treinamento Supervisionado:** é aquele onde as respostas já são de conhecimento, e as respostas da rede são comparadas com

estas a fim de aproximar o comportamento da rede em uma simulação mais próxima do ideal.

- **Treinamento Não-Supervisionado:** diferente do primeiro, este tipo de treinamento ocorre quando as respostas ideais não são conhecidas, e a responsabilidade de encontrá-las é da própria rede, que deve se organizar e ajustar seus padrões com o objetivo de reconhecer particularidades e agrupá-las (princípio da clusterização).

Quanto ao aprendizado, Kriesel (2007), define, de modo genaralizado dois tipos existentes para uma rede neural:

- **On-line:** a cada entrada exposta à rede, são realizados ajustes de valores para cada erro de amostra. Este método é utilizado para aplicações em ambientes muito dinâmicos, que requerem rapidez e nos quais existe a passagem de muitas entradas para a rede.
- **Off-line:** ao contrário do primeiro exemplo, neste caso a rede recebe uma alta quantidade de padrões simultaneamente, e em seguida calcula os erros e ajusta os pesos sinápticos. Este método é conhecido por *batch training procedures* (procedimento de treinamento em lote), porque todos os valores são corrigidos ao mesmo tempo.

#### 3.3.4 Redes Perceptron de Múltiplas Camadas (PMC)

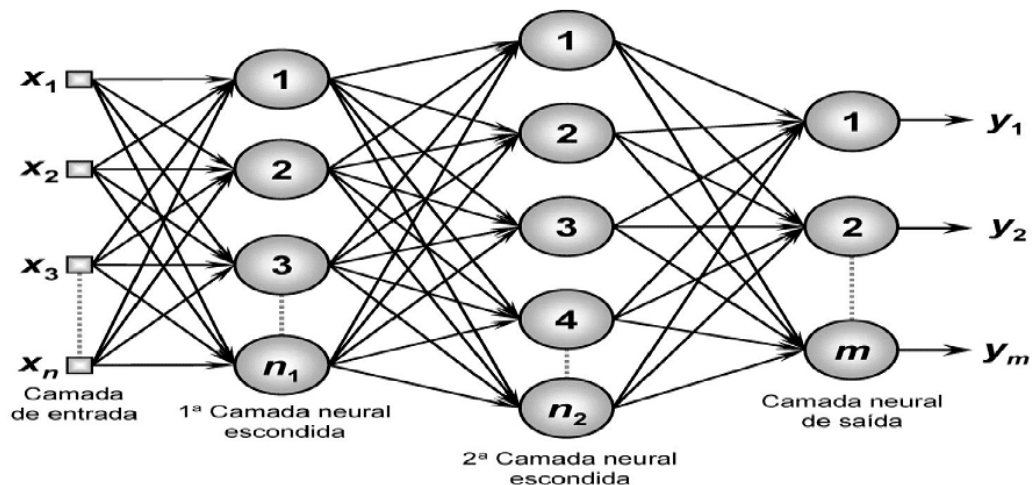
Como é abordado por Bishop (2007), a rede neural artificial que possui melhor desempenho na resolução de classificação de padrões é a rede PMC. Sendo assim, a rede PMC é considerada a melhor opção para o presente projeto. Uma rede Perceptron de múltiplas camadas, tem como característica a existência de no mínimo uma camada intermediária (ou camada escondida) de neurônios artificiais, que estará situada entre as camadas de entrada e saída. Da Silva, Spatti e Flauzino (2010), descrevem as redes PMC como uma das arquiteturas mais versáteis em questão de

aplicabilidade, pois estas redes possuem grande potencial em aplicações de inúmeros tipos de problemas relacionados às mais diversas áreas do conhecimento.

Segundo Haykin (2010), a popularidade das redes PMC se consolidou a partir da publicação do livro *Parallel Distributed Processing* (Rumelhart e McClelland, 1986), no qual o algoritmo de *backpropagation* (algoritmo de retropropagação) foi consolidado. O processo de aprendizagem realizado a partir de tal algoritmo é denominado de aprendizagem por retro propagação.

A rede PMC pertence à arquitetura chamada *feedforward* de camadas múltiplas, cujo treinamento é realizado de forma supervisionada. (DA SILVA, SPATTI, FLAUZINO, 2010). Esta arquitetura indica que o processamento dos dados segue um fluxo de “alimentação para frente”, ou seja, da entrada para a saída, sem a existência de laços ou realimentações. A Figura 8 ilustra o exemplo de uma rede PMC.

Figura 8 – Exemplo de rede PMC.



Fonte: Da Silva, Spatti e Flauzino (2010).

## 4 MATERIAIS E MÉTODOS

### 4.1 MATERIAIS

Nesta seção serão expostos os materiais e ferramentas necessários para atingir o objetivo proposto no trabalho.

#### 4.1.1 Banco de dados

Os sinais vocais a serem analisados estão presentes em um banco de dados fornecido pela Dra. Fabiana Zambon do ambulatório de saúde vocal do SINPRO-SP (sindicato de professores de São Paulo), para o registro dos sinais, os pacientes permaneceram sentados em ambiente controlado, cerca de 70% das gravações foram realizadas por meio de um microfone de cabeça da marca *Plantronics* (Audio 20) ajustado a cerca de 1 centímetro do canto da boca do paciente, conectado à uma placa de som *Sound Blaster Audigy 7.1*. Os áudios provenientes destas gravações foram registrados em arquivos de formato “.wav” por meio do programa *Sound Forge* com taxas de 44.100 amostras por segundo. Contudo, cerca de 30% dos sinais registrados provém de gravações mais antigas nas quais o microfone de cabeça estava acoplado à um modelo obsoleto de placa de som, resultando em captura de somente 22050 amostras por segundo.

As amostras estudadas contem a pronúncia da vogal “e” sustentada e os pacientes, todos do sexo feminino, possuíam entre 21 e 66 anos. A composição geral dos arquivos presentes no banco de dados está representada no Quadro 2:

**Quadro 2 – Composição geral do banco de dados.**

Predominância	Quantidade
Rugosa	27
Não rugosa	63
Total	90

Fonte: Autoria própria.



#### 4.1.2 Ferramentas

As etapas práticas do trabalho serão realizadas dentro do ambiente computacional do *software* MATLAB®, pois o mesmo permite maior facilidade na implementação dos códigos e agilidade para processar os dados paralelamente necessitando de alto esforço computacional.

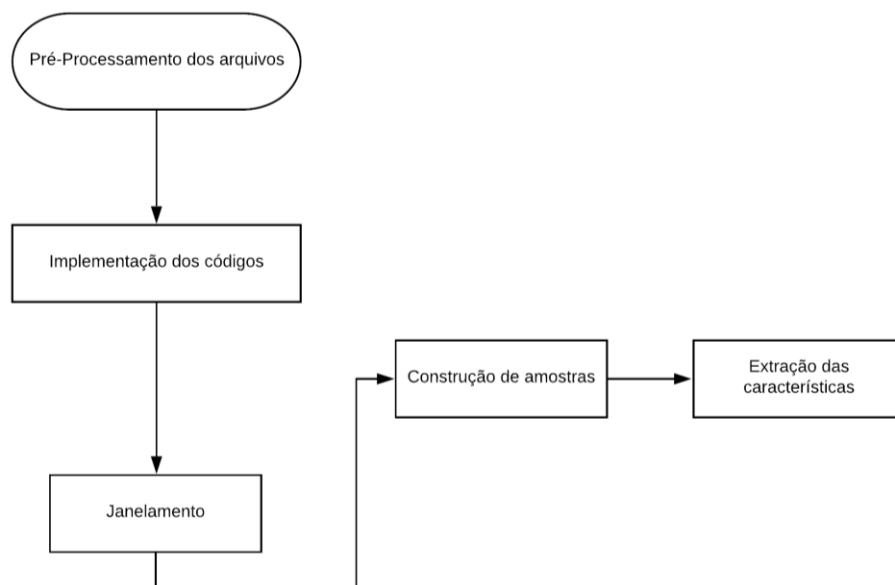
Serão utilizadas também algumas funções já implementadas em pacotes instalados em conjunto com o *software*, como por exemplo as *toolboxes* de processamento de sinais para a aplicação de transformadas e obtenção de parâmetros, e de redes neurais para realizar o processo de treinamento e aprendizagem do sistema inteligente.

#### 4.2 MÉTODOS

Com o banco de dados adquirido, serão abordados, então, os métodos a serem seguidos de acordo com a proposta do trabalho.

A primeira etapa do projeto é definida Pré-processamento dos arquivos, e consiste na implementação dos códigos através do *software* MATLAB® para a realização do janelamento e construção das amostras e, posteriormente, extrair as características dos sinais. A Figura 9 ilustra esta etapa.

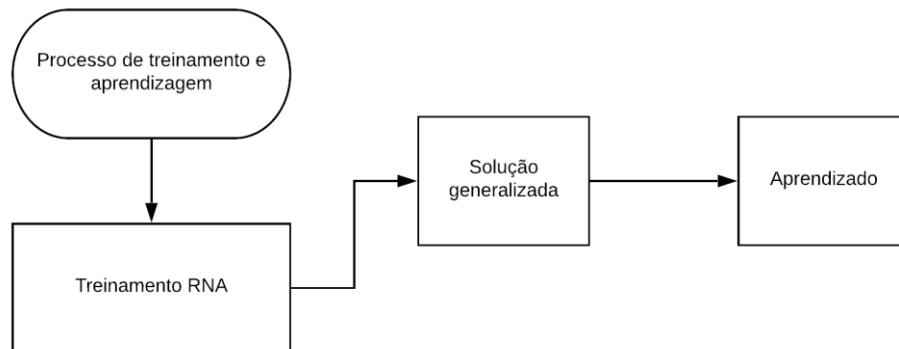
**Figura 9 – Pré-processamento dos arquivos.**



**Fonte: Autoria própria.**

A segunda etapa são os processos de treinamento e aprendizagem, nos quais serão utilizadas entradas obtidas a partir da criação de amostras para testar diversas topologias em busca da melhor topologia para o presente trabalho. A Figura 10 ilustra este procedimento.

**Figura 10 – Processo de treinamento e aprendizagem.**



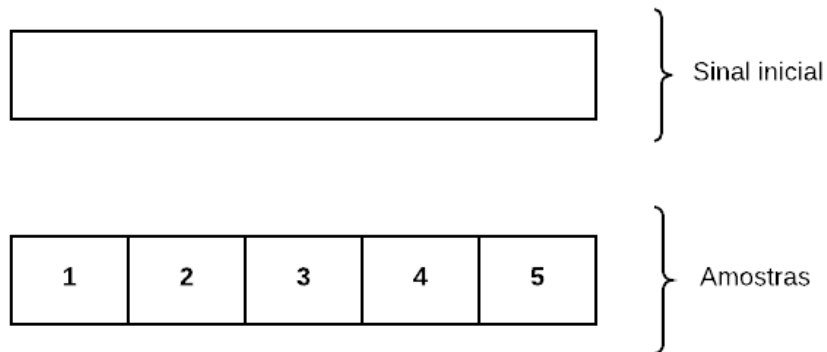
**Fonte: Autoria própria.**

#### 4.2.1 Pré-processamento: amostras e janelamento

Os métodos de janelamento dos sinais são indispensáveis para o resultado desejado do presente estudo. Estes são responsáveis por reduzir o tempo necessário de aplicação da Transformada *Wavelet Packet* e da extração das características, bem como por produzir um maior número de amostras que serão utilizadas durante o treinamento, proporcionando, assim, um aumento no desempenho e uma mais robusta validação da Rede Neural Artificial.

Para tornar clara a explicação dos processos de janelamento, utiliza-se um sinal imaginário qualquer de 500 pontos. O janelamento simples consiste na segmentação de um sinal completo em janelas de mesmo comprimento (não havendo sobreposição entre as janelas), de maneira mais simplificada, o sinal original será dividido em janelas de tamanho desejado. É possível então, segmentar o sinal original de 500 pontos em 5 partes de 100 pontos, como demonstra a Figura 11.

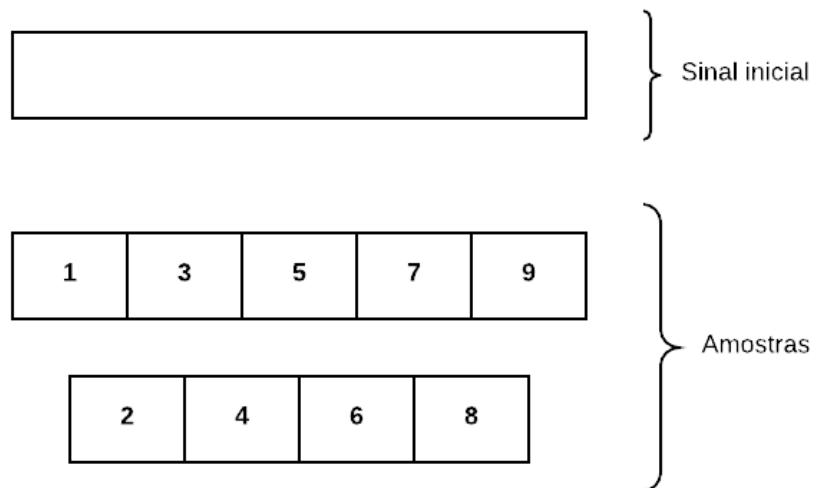
**Figura 11 – Janelamento simples.**



**Fonte: Autoria própria.**

Entretanto, no caso do janelamento cruzado, para este sinal, após a primeira janela de 100 pontos ser retirada do sinal original, a próxima segmentação é realizada com sobreposição de 50% entre as janelas, proporcionando, assim, um número ainda maior de segmentos. O exemplo de janelamento cruzado está ilustrado na Figura 12.

**Figura 12 – Janelamento cruzado.**



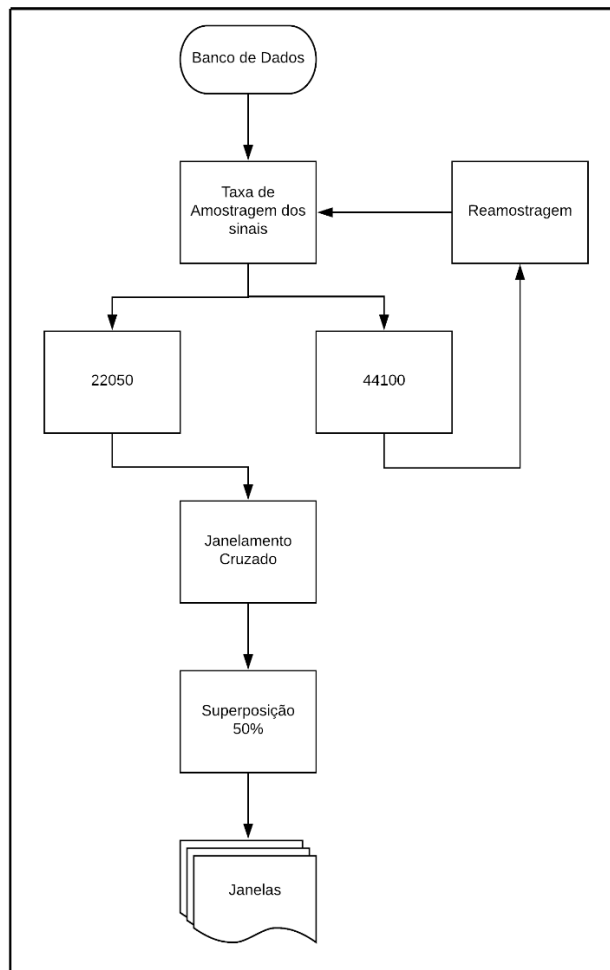
**Fonte: Autoria própria.**

Com o intuito de aumentar o número de amostras e garantir melhor resultado do treinamento da RNA, optou-se no presente estudo, pelo uso do janelamento cruzado com superposição de 50% nas janelas, cada janela contendo um total de 4096 pontos. Ressalta-se que, antes da realização do janelamento cruzado, é necessário reamostrar os arquivos de áudio a uma taxa de 22050 amostras por

segundo para garantir a uniformidade dentre os sinais, a reamostragem se dará através da função “resample” do *software* MATLAB®.

Destaca-se ainda, que por questões de organização esse processo será feito duas vezes, separando os áudios com vozes rugosas dos áudios com vozes não rugosas. A Figura 13 ilustra tal processo de preparação de dados.

**Figura 13 – Processo de preparação de dados.**



**Fonte: Autoria própria.**

#### 4.2.2 Aplicação da Transformada Wavelet Packet e extração de características

A energia do sinal, quando associada às faixas de frequência dos níveis de resolução é capaz de apontar um desvio vocal. Tendo em mente tal capacidade, e ressaltando que a rugosidade consiste em irregularidade na vibração das pregas vocais, resultando em ruídos presentes em baixa frequência, essa característica foi utilizada para o desenvolvimento do presente trabalho.

O nível de energia pode ser, segundo Hsu (1995), calculado por meio da equação (4):

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad (4)$$

Onde,

- $x[n]$  representa o conjunto de pontos no segmento analisado;
- $E$  é o valor da energia obtido.

Na sequência é criado um processo automatizado para extrair as características das janelas criadas a partir do processo de segmentação e compor as amostras que posteriormente serão utilizadas como entrada para a rede neural artificial.

Dentre a variedade de funções *wavelet* existentes, as *wavelets* Daubechies (DAUBECHIES, 1992) mostram-se como as mais eficientes para caracterizar os detalhes através das *wavelet* mãe que se situam nas sub-bandas de baixa resolução, os coeficientes, enquanto as *wavelet* pai caracterizam a tendência do sinal (WANG, 2000). Com isso em mente, em cada segmento, foi aplicada a transformada *wavelet packet* utilizando a família Daubechies 2 no nível 3 de resolução, gerando 8 folhas para cada janela. Para cada folha o nível de energia é calculado, resultando, enfim, em 8 amostras que serão utilizadas como entradas para a rede neural artificial.

#### 4.2.3 Processos de treinamento e aprendizagem

Posteriormente à criação de amostras, tem-se como objetivo utilizá-las nos processos de treinamento e aprendizagem através de uma rotina automatizada, a rede do tipo Perceptron de múltiplas camadas será treinada com um algoritmo supervisionado.

Tendo em mente ainda que o banco de dados não é homogêneo quanto à duração dos arquivos de áudio e à quantidade de arquivos contendo vozes rugosas e não-rugosas, com o intuito de equilibrar o número de amostras, foi feita a remoção de amostras excedentes da categoria de vozes não-rugosas, resultando em 5809

amostras de cada categoria. As características da RNA aplicada estão descritas no Quadro 3.

**Quadro 3 – Características gerais da RNA desenvolvida.**

<b>Arquitetura</b>	Perceptron de camadas múltiplas
<b>Treinamento</b>	Supervisionado – <i>Off-line</i>
<b>Algoritmo de treinamento</b>	Levenberg-Marquardt ( <i>backpropagation</i> )
<b>Taxa de aprendizagem</b>	0.005
<b>Critério de parada (Erro)</b>	$10^{-4}$
<b>Épocas</b>	2600
<b>Proporção Treinamento/Validação</b>	70% / 30%
<b>Número total de amostras</b>	11618

Fonte: Autoria própria.

Com o objetivo de apresentar maior robustez e não descartar uma possível configuração promissora, para todas as topologias, foram realizados 15 treinamentos na rotina automatizada. As topologias receberam variações no número de neurônios das primeira e segunda camadas escondidas, os detalhes sobre as variações de topologias são apresentados no Quadro 4.

**Quadro 4 – Variações de topologia.**

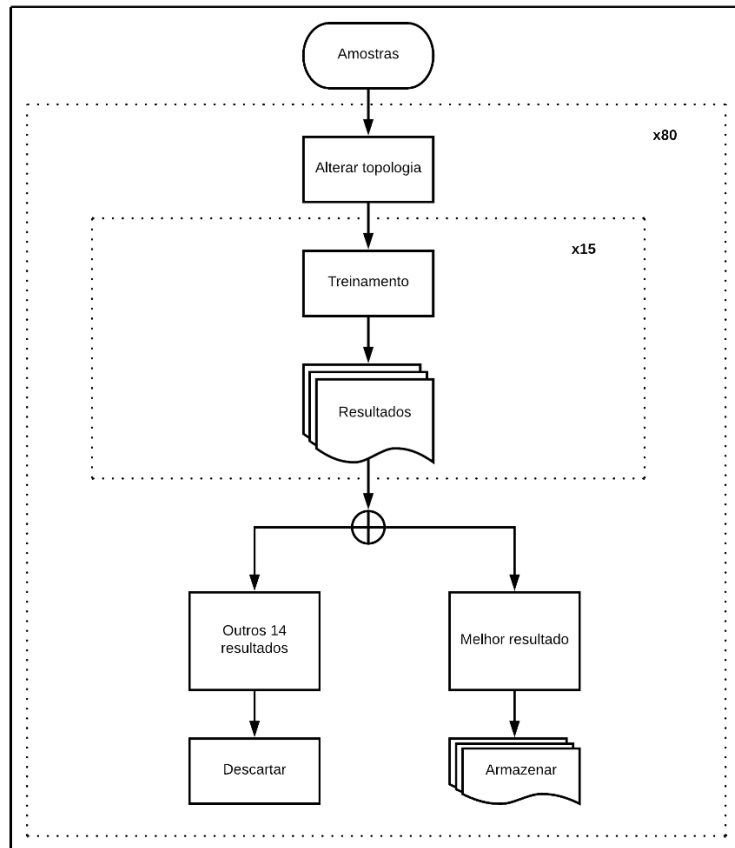
<b>Neurônios na primeira camada escondida</b>	7, 9, 11, 13, 15, 17, 19, 21
<b>Neurônios na segunda camada escondida</b>	3, 5, 7, 9, 11, 13, 15, 17, 19, 21
<b>Total de variações de topologia</b>	80
<b>Treinamentos por topologia</b>	15
<b>Total de treinamentos</b>	1200

Fonte: Autoria própria.

### 4.2.3 Determinação da melhor topologia

Para cada variação de topologia obtém-se um resultado diferente, a topologia que apresentar a maior porcentagem de acertos será considerada a melhor topologia. A Figura 14 ilustra o processo de treinamento da RNA em busca da melhor topologia.

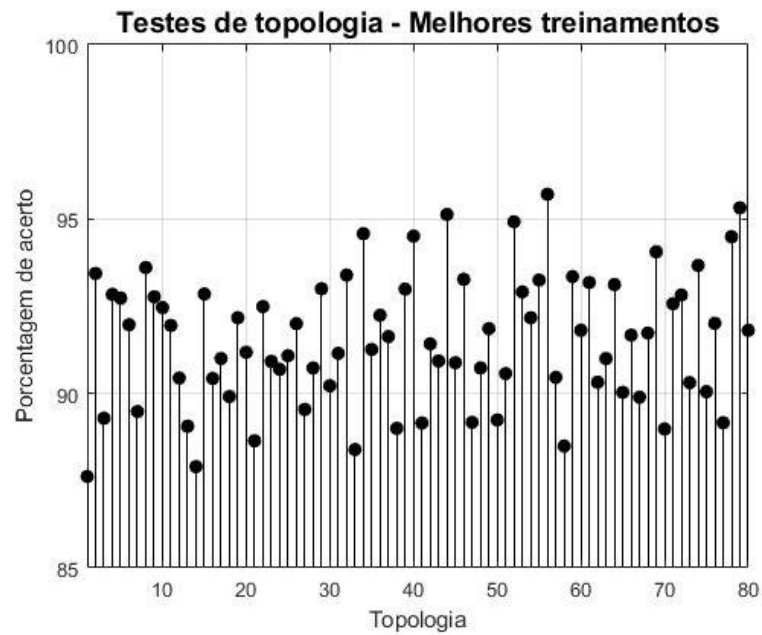
**Figura 14 – Processo de treinamento.**



**Fonte: Autoria própria.**

Após realizados os 15 treinamentos para cada uma das 80 variações de topologia, registrou-se os resultados dos treinamentos que apresentaram maior taxa de acerto durante a etapa de validação. A Figura 15 representa os resultados obtidos no melhor treinamento de cada variação.

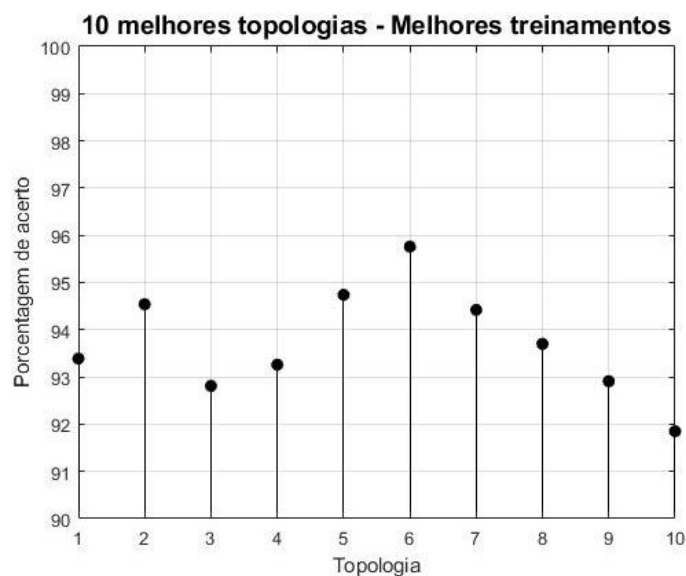
**Figura 15 – Resultados dos melhores treinamentos.**



**Fonte: Autoria própria.**

Como último passo para definir a melhor topologia, foram selecionadas as 10 topologias com maior porcentagem de acerto e novamente realizados 15 treinamentos para cada uma, os resultados dos treinamentos com maior porcentagem de acerto foram novamente comparados para a definição final da melhor topologia. Como resultado obteve-se a topologia com 17 neurônios na primeira camada e 19 neurônios na segunda camada com uma porcentagem de acerto de 95.76%. A Figura 16 representa este resultado.

**Figura 16 – Definição da melhor topologia.**



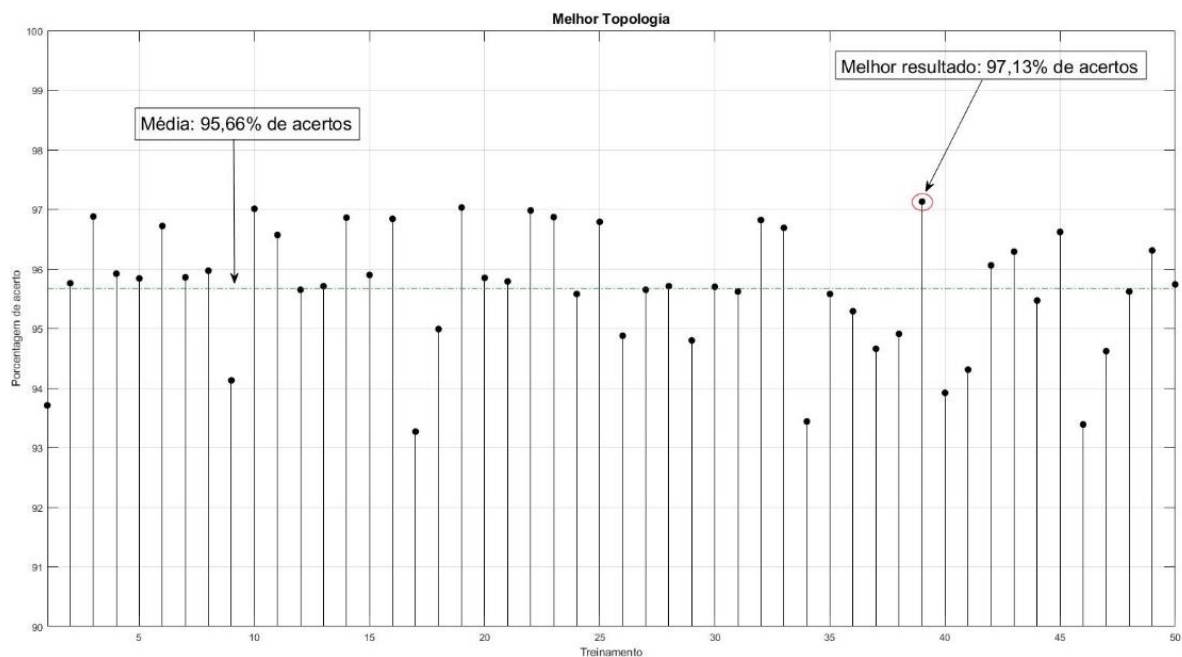
**Fonte: Autoria própria.**



#### 4.2.3.1 Robustez da melhor topologia

Com a melhor topologia definida, inicia-se a próxima rotina de treinamento com o propósito de verificar a robustez da mesma. Foram realizados 50 treinamentos com a quantidade de neurônios do melhor resultado intacta ( $N1=17$  e  $N2=19$ ), variando-se apenas os pesos sinápticos e limiares de ativação (aleatórios para cada novo treinamento). O resultado está representado por meio da Figura 17.

**Figura 17 – Robustez da melhor topologia.**



**Fonte: Autoria própria.**

Como mostra a Figura 18, os resultados se mantêm coerentes, obtendo uma média de 95,66% de taxa de acertos. Podendo, portanto, alegar que a capacidade de generalização da rede neural artificial desenvolvida é fundamentada.

Vale também salientar que, como ilustra a Figura 18, o melhor resultado obtido apresentou taxa de acerto de 97,13%. Sendo, desta forma, o melhor treinamento da melhor topologia, este resultado será utilizado para os demais procedimentos.

## 5 RESULTADOS E DISCUSSÕES

Neste capítulo serão abordados os critérios de pós-processamento e discutidos os resultados da validação. Em seguida, será apresentada a matriz de confusão, demonstrando de forma mais objetiva as estatísticas relacionadas à classificação realizada.

### 5.1 PÓS-PROCESSAMENTO E VALIDAÇÃO DOS RESULTADOS

O processo de pós-processamento representa a transformação da resposta da validação em valores discretos para definir as classes desejadas, simplificando a discussão e permitindo adotar uma margem de confiabilidade para os resultados.

Para o presente projeto, a margem de confiabilidade adotada foi de 97% na classificação, atribuindo os critérios com valores de 0 a 1, os valores encontrados entre 0.03 e 0.97 serão adotados como imprecisão. A explicação dos critérios pode ser observada no Quadro 5.

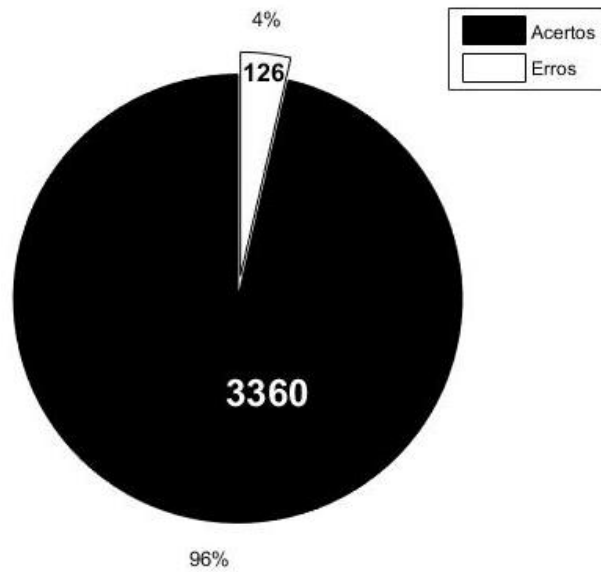
**Quadro 5 – Critérios pós-processamento.**

<b>Valor de saída</b>	<b>Valor adotado</b>	<b>Classificação</b>
$0 \leq Y \leq 0,03$	-1	Voz não-rugosa
$0,03 < Y < 0,97$	0	Imprecisão
$0,97 \leq Y \leq 1$	1	Voz rugosa

**Fonte: Autoria própria.**

Atribuindo todos os casos de imprecisão como erros cometidos, é possível representar de forma visual o desempenho da rede neural artificial criada para o conjunto de amostras destinadas à validação. A Figura 18 ilustra este desempenho.

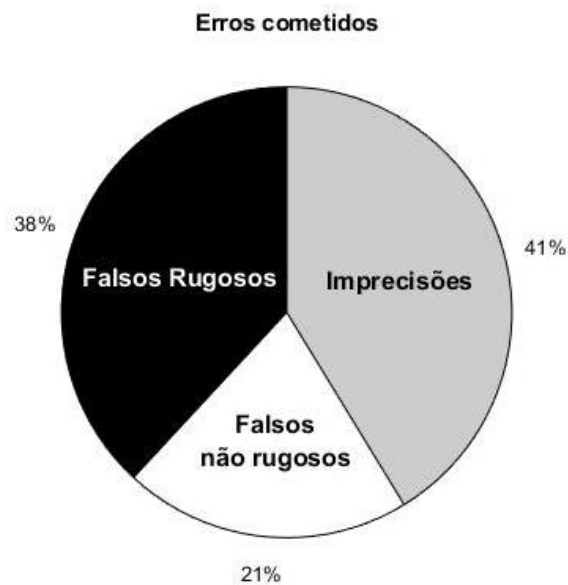
**Figura 18 – Desempenho da RNA.**



**Fonte: Autoria própria.**

Por fim, através da Figura 19 pode-se observar quais os erros cometidos pela rede, em suas respectivas proporções, levando em consideração o número total de erros igual a 126, temos 48 amostras de vozes não-rugosas classificadas como rugosas, 26 amostras de vozes rugosas classificadas como não-rugosas e 52 amostras caracterizando imprecisão.

**Figura 19 – Erros cometidos.**



**Fonte: Autoria própria.**

## 5.2 MATRIZ DE CONFUSÃO

A matriz de confusão (ou matriz de erro), expõe as informações da validação da melhor topologia de forma estruturada, proporcionando uma visualização objetiva dos aspectos relevantes. A matriz de confusão está representada no Quadro 6.

**Quadro 6 – Matriz de confusão.**

	Classificação		
<b>Número de amostras</b> 3486	<b>Rugosa</b> 1718	<b>Não-Rugosa</b> 1716	<b>Imprecisão</b> 52
<b>Rugosas</b> 1740	<b>Verdadeira rugosa</b> 1670	<b>Falsa não-rugosa</b> 26	<b>Taxa de falsa Rugosa</b> 2,7939%
<b>Não-Rugosas</b> 1746	<b>Falsa rugosa</b> 48	<b>Verdadeira não-rugosa</b> 1690	<b>Taxa de falsa não-rugosa</b> 1,5152%
	<b>Acurácia rugosa</b> 97,2061%	<b>Acurácia não-rugosa</b> 98,4848%	<b>Acurácia geral</b> 96,3855%

Fonte: Autoria própria.

Onde,

- **Acurácia rugosa:** representa a taxa de acertos na classificação de vozes rugosas;
- **Acurácia não-rugosa:** representa a taxa de acertos na classificação de vozes não-rugosas;
- **Acurácia geral:** representa a taxa de acertos na classificação de todas as amostras;
- **Taxa de falsa rugosa:** representa a taxa de erros na classificação de vozes rugosas;
- **Taxa de falsa não-rugosa:** representa a taxa de erros na classificação de vozes não-rugosas.

## 6 CONSIDERAÇÕES FINAIS

A integração entre as áreas tecnológicas e biológicas se mostrou, ao longo das décadas, extremamente importante, proporcionando ao ser humano maior qualidade de vida. A busca por tecnologias capazes de auxiliar profissionais da área de fonoaudiologia é apenas uma das infinitas possibilidades de benefício que as RNA podem oferecer.

O presente trabalho, apresentou um método para a classificação de arquivos de áudio com parâmetros subjetivos do distúrbio vocal denominado rugosidade. Por intermédio de ferramentas computacionais, obteve-se um classificador neural com média de confiabilidade acima dos 95% apesar do banco de dados ser formado por arquivos gravados entre 2009 e 2011.

Deve-se, porém, ressaltar que os estudos realizados neste trabalho, não propõem, de forma alguma, desconsiderar a importância e aplicação de diagnósticos realizados por profissionais da saúde da voz. É apresentado apenas um recurso capaz de auxiliar profissionais das áreas médicas, se tornando uma opção às análises acústicas já empregadas nos diagnósticos.

Como motivação para futuros estudos baseados neste trabalho, alguns pontos de interesse podem ser enfatizados, tais como a investigação de diferentes parâmetros, a utilização de arquivos de áudio mais recentes, bem como, a sustentação de diferentes vogais nos mesmos, a classificação de outros distúrbios vocais e as diferentes possibilidades de famílias *wavelet* a serem utilizadas em casos distintos.

## REFERÊNCIAS

- ABORL-CCF – Associação Brasileira de Otorrinolaringologia e Cirurgia Cérvico-Facil. **Consenso Nacional Sobre Voz Profissional: Voz e Trabalho, Uma Questão de Saúde e Direito do Trabalhador.** Rio de Janeiro, 2004.
- ANAND, S.; SKOWRONSKI, M. D.; SHRIVASTAV, R.; EDDINS, D. A. **Perceptual and Quantitative Assessment of Dysphonia Across Vowel Categories.** *Journal of Voice*, 2017.
- BEHLAU, M.; AZEVEDO, R.; PONTES, P. **Conceito da Voz Normal e Classificação das Disfonias.** *Voz – O Livro do Especialista.* v.1. Rio de Janeiro: Revinter, p.53-84, 2004.
- BISHOP, C. M. **Pattern Recognition and Machine Learning.** New York, US: Springer, 2007.
- CAMPOS, C. A. H.; COSTA, H. O. O. **Tratado de Otorrinolaringologia.** São Paulo: Rocca, 2002.
- CARVALHO, R. T. S.; CAVALCANTE, C. C.; CORTEZ, P. C. **Wavelet transform and artificial neural networks applied to voice disorders identification.** In *Nature and Biologically Inspired Computing (NaBIC)*, 2011.
- DAUBECHIES, I. **Ten Lectures on Wavelets.** Philadelphia, Pennsylvania: Society for Industrial Mathematics, 1992.
- DA SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas.** São Paulo, SP: Artliber, 2010.
- FREITAS, S. V. **Correlação Entre Avaliação Acústica e Perceptual na Caracterização de Vozes Patológicas.** 132 f. Relatório do Estado-da-Arte. Faculdade de Engenharia Universidade do Porto, Porto, 2010.
- GAYOTTO, L. H. da C. **Dinâmicas de Movimento da Voz.** *Revista Distúrbios da Comunicação.* 2006; v. 17, no. 3, p. 41-49.
- GHASEMZADEH, H. et al. **Detection of Vocal Disorders Based on Phase Space Parameters and Lyapunov Spectrum.** *Biomedical Signal Processing and Control*, 22, p. 135-145, 2015.
- GONZÁLEZ, I. V. **Videolaringoestroboscopia: Una Técnica Para Visualizar las Cuerdas Vocales.** *Estudios de fonética experimental*, no. 17: p. 418-432, 2008.
- GRAGONE, M. L. S.; NAGANO, L. **A Voz Que Ensina: O Professor e a Comunicação Oral em Sala de Aula.** Rio de Janeiro: Revinter, 2004.
- GRAPS, A. **An Introduction to Wavelets.** *Computational Science & Engineering*, IEEE 2(2): p. 50-61, 1995.

HAGAN, M. T. **Neural Network Design**. Second Edition. Boston, MA: Martin Hagan, 2014.

HAYKIN, S. S. **Neural Networks: A Comprehensive Foundation**. New Jersey, US: Prentice Hall, 1999.

INGLE, V. K; PROAKIS, J. G. **Digital Signal Processing Using MATLAB**. Cengage Learning, Stamford, p. 213-289:476-499, 2012.

KRIESEL, D. **A Brief Introduction to Neural Networks**. 2007. Disponível em <[http://www.dkriesel.com/\\_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf](http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf) > Acesso em: 30 out. 2017.

KRUG, E. <http://coral.ufsm.br/otorrino/index.php/otorrinolaringologia/sintomas-e-doencas-mais-comuns/2-uncategorised/8-alteracoes-da-voz> - Universidade Federal de Santa Maria. Acesso em 18/10/2017 13:53

LIMA, A. A. M. **Classificação de Disfonias Utilizando Redes Neurais Artificiais e Transformadas Wavelet Packet**. Trabalho de Conclusão de Curso em Engenharia de Controle e Automação - Universidade Tecnológica Federal do Paraná, Cornélio Procópio, PR, 2018.

MANOLAKIS, D. G.; INGLE, V. K. **Applied Digital Signal Processing: Theory and Practice**. Cambridge, UK: Cambridge University Press, 2011.

MIRANDA, K. M. **Cuidados com a Voz**. Equipe de Multiprofissionais – Gerência de Ensino Especial Secretaria de Estado da Educação, Cultura e Esporte, 2016.

OLIVEIRA, J. **A Importancia da Saúde Vocal Para Profissionais**. Rev. Espaço Aberto, Ago/2013. Ed. 152. USP São Paulo.

OLIVEIRA, H. M. **Análise de Sinais Para Engenheiros. Uma Abordagem via Wavelets 1 ed.** [S.I.]: BRASPORT, (2007).

OMORI, K.; KOJIMA, T.; KAKANI, R.; SLAVIT, D. H.; BLAUGRUND, S. M. **Acoustic Characteristics of Rough Voice" Subharmonics**. Journal of Voice Vol. 11, No. 1, pp. 40-47, 1996.

OPPENHEIM, A. V.; SCHAFER, R. W. **Discrete-Time Signal Processing**. Upper Saddle River, New Jersey, p. 140-178:185-212:439-464, 1999.

PAVONI, H. E. **Classificação de Sinais Vocais em Parâmetros não Acústicos Utilizando Redes Neurais Artificiais**. Trabalho de Conclusão de Curso em Engenharia Elétrica - Universidade Tecnológica Federal do Paraná, Cornélio Procópio, PR, 2017.

PENTEADO; TEIXEIRA; PEREIRA. **A Voz do Professor: Relações Entre Trabalho, Saúde e Qualidade de Vida**. Rev. Bras. De Saude Ocupacional, vol 25, p. 109-129, 1995/96.

PHILIPPI, F. F. **Anatomia das Pregas Vocais e Fisiologia da Fonação**. USP, 2009.

PINHO, S. M. R.; TSUJI, D. H.; BOHADANA, S. C. **Fundamentos em Laringologia e Voz**. Rio de Janeiro, RJ: Revinter, 2006.

PRARRAGA, A. **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. 163 f. Dissertação. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

STOLLNITZ, E. J.; DEROSE, T. D.; SALESIN, D. H. **Wavelets for Computer Graphics: A Primer - part 1**. *IEEE Computer Graphics and Applications*, p. 76–84, 1995.

WANG, J. Z. (2000). **Semantics-Sensitive Integrated Matching for Picture Libraries and Biomedical Image Databases**. Dissertação de PHD, Department of Biomedical Informatics of Stanford University.

WEEKS, M. **Digital Signal Processing Using MATLAB and Wavelets**, second edition, US: Jones & Bartlett Learning, p. 69-110:149-168, 2012.

ZAYED, A.I. **Advances in Shannon's Sampling Theory**, CRC Press, Boca Raton, 1993.

ZITTA, S. M. **Análise Perceptivo-Auditiva e Acústica em Mulheres com Nódulos Vocais**. 131 f. Dissertação de Pós-Graduação. CEFET, Paraná, Curitiba, 2005.

GRINFELD, A. <http://www.portaldootorrino.com/exame.php?id=3> acesso em 27/10/2017 as 23:22. Hospital Otorrinolaringológico. Bahia, Salvador, 2017.