

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE ENGENHARIA ELÉTRICA
ENGENHARIA DE CONTROLE E AUTOMAÇÃO

ARON ALEXANDRE MARTINS LIMA

**CLASSIFICAÇÃO DE DISFONIAS UTILIZANDO REDES NEURAIS ARTIFICIAIS E
TRANSFORMADAS WAVELET PACKET**

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO
2018

ARON ALEXANDRE MARTINS LIMA

**CLASSIFICAÇÃO DE DISFONIAS UTILIZANDO REDES NEURAIS ARTIFICIAIS E
TRANSFORMADAS WAVELETS PACKET**

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.

Orientador: Profa. Dra. Maria Eugenia Dajer

Co-orientador: Prof. Dr. Danilo H. Spatti

CORNÉLIO PROCÓPIO
2018



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



FOLHA DE APROVAÇÃO

Aron Alexandre Martins Lima

Classificação de disfonias utilizando redes neurais artificiais e transformadas Wavelet Packet

Trabalho de conclusão de curso apresentado às 13:50hs do dia 23/11/2018 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). Cristiano Marcos Agulhari - (Membro)

Prof(a). Dr(a). Paulo Rogério Scalassara - (Membro)

Dedico este trabalho à todos que me auxiliam em minha jornada.

AGRADECIMENTOS

Primeiramente gostaria de agradecer aos meus orientadores Maria Eugenia Dajer e Danilo Hernane Spatti, por me auxiliarem na escolha e estruturação deste trabalho. Agradeço ainda pela paciência, disponibilidade e empenho de ambos, mesmo diante a todos os empecilhos.

Agradeço também a meus familiares, que sempre me dão todo o suporte necessário, aos amigos, Alisson, Andriéli, Murilo e Rafael que me ajudaram a tornar este trabalho mais clara e objetivo e aos amigos Fábio e Victor, por disponibilizarem o uso da ferramenta de otimização de parâmetros internos da Perceptron Multicamadas. Bem como ao Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico e a Fabiana Zambon, pela disponibilidade das bases de dados utilizadas para este trabalho.

RESUMO

LIMA, Aron Alexandre Martins. **CLASSIFICAÇÃO DE DISFONIAS UTILIZANDO REDES NEURAIIS ARTIFICIAIS E TRANSFORMADAS WAVELETS PACKET**. 2018. 68 f. Trabalho de Conclusão de Curso (Graduação) – Engenharia de Controle e Automação. Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2018.

Grande parte das patologias laringeas causam alterações no padrão vibratório das pregas vocais. Essas alterações produzem mudanças na qualidade vocal e podem tornar-se um problema significativo, principalmente para quem faz uso ocupacional e profissional da voz. Os métodos de avaliação de voz são em geral invasivos, possuem alto custo de implementação e causam desconforto ao paciente. Por outro lado os métodos não invasivos não possuem boa resolução e robustez, já que nestes métodos o sinal de voz é analisado como um sinal periódico, sem considerar que quanto mais patológica a voz, mais aperiódico será o sinal. As Redes Neurais Artificiais do tipo *Perceptron* Multicamadas, junto às Transformadas *Wavelets Packet*, vêm sendo amplamente aplicadas no reconhecimento de padrões de sinais biológicos, apresentando boa acurácia, com altos índices de confiabilidade. Por tais motivos, este trabalho aplica às Redes Neurais Artificiais e Transformadas *Wavelets Packet* para análise de sinais vocais que ajudem na classificação de disfonias funcionais e orgânicas, bem como vozes saudáveis. Dessa forma obteve-se acurácia na classificação de 100% para saudáveis, 99,55% para disfonias orgânicas e 93,47% para disfonias funcionais, com 98% de confiabilidade.

Palavras-chave: Redes Neurais Artificiais. Transformada Wavelet Packet. Voz. Disfonias. Classificação de disfonias.

ABSTRACT

LIMA, Aron Alexandre Martins. **CLASSIFICATION OF DYSPHONIAS USING ARTIFICIAL NEURAL NETWORKS AND WAVELETS PACKET TRANSFORMS.** 2018. 68 f. Trabalho de Conclusão de Curso (Graduação) – Engenharia de Controle e Automação. Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2018.

Most laryngeal pathologies cause deviations in the vibratory pattern of the vocal folds. These deviations produce changes in vocal quality and can become a significant problem, especially for occupational and professional voice users. Voice assessment methods are generally invasive, have a high implementation cost and cause discomfort to the patient. On the other hand, noninvasive methods do not have good resolution and robustness, since they assume voice as a periodic signal, without considering that the more pathological the voice, the more aperiodic is the signal. Artificial Neural Networks as Multilayer Perceptron have been widely applied in the recognition of nonlinear biological signal patterns, presenting good accuracy, with high reliability. For these reasons, this work proposes the application of Artificial Neural Networks and Wavelet Packet Transform for voice signal analysis and for classification of functional and organic dysphonias, as well as healthy voices. In this way, 100% classification accuracy was obtained for healthy, 99.55% for organic dysphonia and 93.47% for functional dysphonia, with 98% reliability.

Keywords: Artificial Neural Network. Dysphonia. Functional. Organic. Wavelet Packet Transform.

LISTA DE FIGURAS

Figura 1 – Músculos e cartilagens da laringe (vista superior).....	18
Figura 2 – Paralisia Unilateral de Prega Vocal.....	21
Figura 3 – Videolaringoscopia de alta velocidade	22
Figura 4 – Sinal senoidal amostrado.....	24
Figura 5 – Fenômeno de aliasing, ocorreu devido a taxa de amostragem muito baixa.....	25
Figura 6 – Exemplo de <i>downsampling</i> em sinal senoidal.....	26
Figura 7 – Decomposição do sinal.....	27
Figura 8 – Reconstrução do sinal.....	27
Figura 9 – Coeficientes da decomposição <i>Wavelet Packet</i>	27
Figura 10 – Árvore de decomposição <i>Wavelet Packet</i>	28
Figura 11 – Esquerda: Decomposição para TWP; Direita: Decomposição para TWD.....	29
Figura 12 – Neurônio Artificial.....	31
Figura 13 – <i>Perceptron</i> Multicamadas.....	33
Figura 14 – <i>Perceptron</i> fase <i>forward</i> e <i>backward</i>	34
Figura 15 – Tangente hiperbólica.....	34
Figura 16 – Estrutura ART-1.....	36
Figura 17 – Fluxograma de Funcionamento da ART-1.....	37
Figura 18 – Sinal original em azul, o envelope RMS calculado em vermelho e o limiar de corte em amarelo.....	42
Figura 19 – Sinal sem silêncio.....	42
Figura 20 – Fluxograma do pré-processamento.....	43
Figura 21 – Fluxograma da segmentação.....	44
Figura 22 – Fluxograma da seleção das topologias TWP iniciais	46
Figura 23 – Exemplo de Cruzamento e TWP.....	47
Figura 24 – Cálculo da média de acertos, para seleção da topologia TWP.....	48
Figura 25 – Fluxograma do torneio.....	48
Figura 26 – Fluxograma do cruzamento TWP.....	48
Figura 27 – Fluxograma da redução das entradas e amostras.....	50
Figura 28 – Fluxograma completo da extração de características.....	50

LISTA DE TABELAS

Tabela 1 – Variações de topologias TWP.....	46
Tabela 2 – Configuração utilizada para a PMC.....	46
Tabela 3 – Saídas das RNAs especialistas.....	46
Tabela 4 – Testes para redução do vetor de amostras de entradas.....	49
Tabela 5 – Gama de ação do SIH.....	51
Tabela 6 – Arquivos de voz de teste e arquivos similares disfonia orgânica.....	52
Tabela 7 – Arquivos de voz de teste e arquivos similares disfonia funcional.....	52
Tabela 8 – Arquivos de voz de teste e arquivos similares Saudáveis.....	53
Tabela 9 – 8 melhores topologias TWP iniciais(Saudável).....	53
Tabela 10 – 8 melhores topologias TWP iniciais (Disfonia Funcional).....	53
Tabela 11 – 8 melhores topologias TWP iniciais (Disfonia Orgânica).....	54
Tabela 12 – 10 melhores topologias TWP após cruzamento (Disfonia Orgânica).....	54
Tabela 13 – 10 melhores topologias TWP após cruzamento (Saudável).....	55
Tabela 14 – 10 melhores topologias TWP após cruzamento (Disfonia Funcional).....	55
Tabela 15 – Redução da melhor topologia TWP (Disfonia Funcional).....	56
Tabela 16 – Redução da melhor topologia TWP (Saudável).....	56
Tabela 17 – Redução da melhor topologia WPT (Disfonia Orgânica).....	56
Tabela 18 – Teste de diferentes treinamento para a melhor topologia TWP (Disfonia Orgânica).....	57
Tabela 19 – Teste de diferentes treinamento para a melhor topologia TWP (Saudável).....	57
Tabela 20 – Teste de diferentes treinamento para a melhor topologia TWP (Disfonia Funcional).....	58
Tabela 21 – Melhor topologia e configuração de treinamento TWP.....	58
Tabela 22 – Parâmetros internos das RNAs, obtidas pela otimização.....	59
Tabela 23 – Melhores Famílias TWP obtidas.....	61
Tabela 24 – Acurácia por classe disfonia orgânica.....	61
Tabela 25 – Acurácia, ganho percentual por etapa.....	62

LISTA DE QUADROS

Quadro 1 - Arquivos de voz.....	39
Quadro 2 - Matriz confusão (Saudável).....	58
Quadro 3 - Matriz confusão (Disfonia Orgânica).....	59
Quadro 4 - Matriz confusão (Disfonia Funcional).....	59

LISTA DE SIGLAS

ART	Teoria da Ressonância Adaptativa (Adaptive Resonance Theory)
DC	Corrente Contínua (Direct Current)
FIR	Resposta ao Impulso Finita (Finite Impulse Response)
FPA	Filtro Passa-Alta
FPB	Filtro Passa-Baixa
GPEM CNPq	Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico
HC-FM USP	Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo
LMS	Mínimos Quadrados Médios (Least Mean Square)
PDS	Processamento Digital de Sinais
PMC	Perceptron Multicamadas
PUPV	Paralisia Unilateral de Prega Vocal
RMS	Raízes Quadradas Médias (Root Mean Square)
RNA	Rede Neural Artificial
SIH	Sistema Inteligente Híbrido
TWD	Transformada Wavelet Discreta
TWP	Transformada Wavelet Packet

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	Objetivo.....	14
1.1.1	Objetivos Específicos.....	15
1.2	Trabalhos Realizados.....	15
1.2.1	Publicados.....	15
1.2.2	Aceitos.....	16
1.2	Organização do Texto.....	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Voz.....	17
2.1.1	Fisiologia da Laringe.....	18
2.1.2	Disfonias.....	19
2.1.2.1	Disfonia Funcional.....	20
2.1.2.2	Disfonia Orgânica.....	20
2.1.2.2.1	Paralisia Unilateral de Prega Vocal (PUPV).....	20
2.1.3	Métodos de Avaliação da Voz.....	21
2.1.3.1	Métodos Invasivos e Avaliação da Voz.....	21
2.1.3.2	Métodos não-Invasivos de Avaliação da Voz.....	22
2.1.3.2.1	Análise Perceptiva-Auditiva.....	22
2.1.3.2.2	Avaliação Acústica.....	22
2.1.3.2.3	Autoavaliação da Qualidade Vocal.....	23
2.2	Processamento Digital de Sinais (PDS).....	23
2.2.1	Amostragem.....	23
2.2.1.1	<i>Aliasing</i> e Teorema de Nyquist.....	24
2.2.2	Reamostragem.....	25
2.2.2.1	<i>Downsampling</i>	25
2.2.3	Transformada <i>Wavelet</i>	26
2.2.3.1	Transformada <i>Wavelet</i> Discreta.....	26
2.2.3.2	Transformada <i>Wavelet Packet</i>	28
2.2.4	DC Offset.....	29
2.2.5	Funções do MATLAB.....	29
2.2.5.1	Envelope RMS.....	29
2.2.5.2	<i>Detrend</i>	29
2.3	Sistemas Inteligentes.....	30
2.3.1	Redes Neurais Artificiais.....	30
2.3.1.1	Neurônio Artificial.....	31
2.3.1.2	<i>Perceptron</i> Multicamadas.....	32

2.3.2	<i>Adaptive Resonance Theory(ART)</i>	35
3	MATERIAIS E MÉTODOS	38
3.1	Base de Dados	38
3.2	Ferramentas	39
3.3	Procedimentos	40
3.3.1	Pré-Processamento do Sinal.....	40
3.3.2	Segmentação.....	44
3.3.3	Extração de Características.....	45
3.3.4	Classificação.....	51
3.3.5	Pós-Processamento.....	51
4	RESULTADOS E DISCUSSÕES	52
4.1	Resultados	52
4.2	Análise dos Resultados	59
5	CONCLUSÃO	63
	REFERÊNCIAS	65

1 INTRODUÇÃO

A capacidade fonatória é a única função básica da laringe totalmente voluntária, e de aparição evolutiva, mais tardia (PINHO *et al*, 2006, p. 11). Esta capacidade é de fundamental importância para a perpetuidade da inteligência humana, possibilitando a comunicação efetiva entre indivíduos. De acordo com Sternberg *et al* (1981), é a função mais estudada devido à sua importância na sociedade.

Grande parte das patologias na laringe causam alterações na vibração das pregas vocais, ocasionando alterações na voz, podendo se tornar um problema principalmente às pessoas que a utilizam em suas profissões (PARRAGA, 2002, p.1).

Os procedimentos de avaliação vocal são classificados em invasivos e não invasivos. Como apresentado em Tsuji *et al* (2014) e González (2008), os métodos invasivos consistem na introdução de ferramentas de captura de imagens até a laringe, de modo a observar e analisar os movimentos das pregas vocais durante a fonação. Contudo, este procedimento é desconfortável ao paciente, podendo causar dores, náuseas e ansiedade, além de ser necessária a utilização de ferramentas que em alguns contextos podem ser considerados de alto custo, além de ser fundamental a experiência do examinador para analisar as imagens obtidas. Por outro lado, existem métodos não invasivos que buscam, a partir de análises dos sinais de voz, auxiliar na avaliação vocal (SCALASSARA, 2009), (FERMINO *et al*, 2016a), (FERMINO *et al*, 2016b), (BARIZÃO, 2017). Como citado por Parraga (2002, p.1), estes métodos proporcionam exames menos desconfortáveis ao paciente e possibilitam o desenvolvimento de sistemas automáticos que auxiliam o diagnóstico, com baixo custo, fornecendo uma avaliação quantitativa, padronizando os critérios de avaliação da voz.

Para a extração de características de sinais biológicos, a Transformada *Wavelet Packet* (TWP) vem sendo amplamente aplicada. Os sinais biológicos possuem características não estacionárias, o que dificulta a aplicação da Transformada de Fourier, mais usual na análise de sinais. As TWP possuem aplicações na voz, como apresentado em Scalassara (2009), Fermino *et al* (2016a),

Pavoni, *et al* (2016), Sodr  (2016), bem como em eletromiografia, apresentado em Al-timemy *et al* (2015), Barros *et al* (2017), Lima *et al* (2018) e Martins *et al* (2018), assim tamb m como em eletrocardiograma, apresentado em Agulhari (2009) e Silva *et al* (2015), entre outras  reas. Como encontrado na literatura esta ferramenta apresenta uma ampla variedade de possibilidades de adequa es para aplica o na extra o de caracter sticas, al m de possibilitar an lise tempo-frequ ncia.

De maneira equivalente as Redes Neurais Artificiais (RNAs) cada vez mais v m ganhando espa o no reconhecimento de padr es de sinais biol gicos, bem como em diversas outras  reas do conhecimento. A populariza o destas ferramentas nos  ltimos anos resultou no aprimoramento de algoritmos que tornaram mais simples o desenvolvimento destes sistemas. Outro fator importante para o aprimoramento das RNAs   o aumento na capacidade de processamento dos computadores, que acarretou em quedas dr sticas no tempo de treinamento de uma RNA e na facilidade de compartilhamento de informa es. Desta forma, na atualidade   poss vel realizar o treinamento de redes, com grandes quantidades de dados, bem como atingir altos n veis de acur cia e confiabilidade, como apresentam Lima (2012), Fermino *et al* (2016a), Pavoni *et al* (2016), Barros *et al* (2017), Bariz o (2017), Martins *et al* (2018) e Lima *et al* (2018).

Por tais motivos, este trabalho se prop e a estudar a aplica o destas duas ferramentas (TWP e RNAs) no reconhecimento de padr es de disfonias, tomando como base a classifica o sugerida por Behlau (2004), onde as disfonias est o distribu das em 3 classes (disfonias funcionais, disfonias organofuncionais e disfonias org nicas), ligadas  s suas causas. Neste trabalho ser o consideradas as disfonias funcionais e org nicas.

1.1 Objetivo

- Desenvolver um sistema autom tico baseado na rede neural artificial *Perceptron* multicamadas, capaz de reconhecer padr es de vozes com disfonias funcionais, disfonias org nicas, bem como vozes saud veis.

1.1.1 Objetivos específicos

- Realizar o pré-processamento dos sinais de vozes disfônicas e saudáveis;
- Realizar a separação de dados, utilizando RNA do tipo *Adaptive Resonance Theory* para treinamento diversificado;
- Desenvolver e implementar algoritmo de busca pela melhor topologia de TWP para cada uma das disfonias, para extração de características dos sinais;
- Realizar combinação de características extraídas das famílias TWP;
- Criar e treinar RNAs especialistas para cada disfonia;
- Testar as redes especialistas obtidas.

1.2 Trabalhos Realizados

1.2.1 Publicados

- Treinamento Otimizado de Redes Neurais Artificiais Para a Classificação de Movimentos Musculares. Encontro Nacional de Inteligência Artificial e Computacional, Uberlândia, MG, 2017.
- Reconhecimento de padrões de movimentos da mão a partir de sinais mioelétricos do antebraço utilizando redes neurais artificiais e algoritmos genéticos. V Congresso Brasileiro de Eletromiografia e Cinesiologia e X Simpósio de Engenharia Biomédica. Uberlândia, MG, 2018.
- Classification of Hand Movements from EMG Signals using Optimized MLP. International Joint Conference in Neural Network- IJCNN. Rio de Janeiro, RJ, 2018.
- Reconhecimento de padrões de movimentos da mão a partir de sinais mioelétricos do antebraço utilizando redes neurais artificiais e algoritmos genéticos. Em: Bases da Saúde e Engenharia Biomédica v. 2. cap. 15, p. 92-99. Editora Atena. Belo Horizonte, MG, 2018.

- Identificação de disfonias Utilizando Redes Neurais Artificiais. Seminário de Iniciação Científica e Tecnológica da UTFPR. Apucarana, PR, 2018.
- Classificação de Movimentos da Mão Utilizando Redes Neurais Artificiais. Seminário de Iniciação Científica e Tecnológica da UTFPR. Apucarana, PR, 2018.
- Processamento Digital de Sinais Biológicos para Aplicação em Redes Neurais Artificiais. Seminário de Iniciação Científica e Tecnológica da UTFPR. Apucarana, PR, 2018.
- Sistema de Optimização de Parâmetros Internos de Redes Neurais Artificiais. Seminário de Iniciação Científica e Tecnológica da UTFPR. Apucarana, PR, 2018.

1.2.2 Aceitos

- Optimized Artificial Neural Network for EMG Classification Using Genetic Algorithm. Journal of Control, Automation and Electrical Systems, 2018.

1.3 Organização do Texto

Este texto foi organizado da seguinte forma: a seção 2 terá por objetivo apresentar os atributos teóricos que fundamentaram este trabalho, tratando sobre a voz, processamento digital de sinais e sistemas inteligentes classificadores de padrões. A seção 3 apresenta as ferramentas e procedimentos realizados para efetividade do estudo. A seção 4 apresenta os resultados obtidos, e as análises dos mesmos para o presente trabalho. Por fim a seção 5 trata sobre as considerações finais bem como as perspectivas futuras.

2 FUNDAMENTAÇÃO TEÓRICA

É de exímia importância a realização de um estudo aprofundado dos conceitos utilizados para a realização do presente trabalho, buscando bases teóricas consolidadas que suportaram a utilização das ferramentas empregadas. Esta Seção está subdividida em três principais vertentes. A primeira trata da voz: conceitos, fisiologia, disfonias e métodos de avaliação. A segunda trata do processamento digital de sinais a serem empregados. E por último, os sistemas inteligentes classificadores de padrões e seus funcionamentos.

2.1 Voz

Segundo Sternberg *et al* (1981), existem 3 fatores básicos que denotam a Inteligência humana. São estas: capacidade de resolução de problemas, capacidade verbal e competência social. A capacidade verbal, ou linguagem, está relacionada à disponibilidade de vasto vocabulário que relaciona e organiza informações, por meio da fala e escrita.

A fala é a emissão de diferentes características da voz dada pela modulação de lábios, dentes, bochecha e língua. A voz por sua vez, é descrita como a produção de som, dada pela vibração das pregas vocais devido a pressão do ar provinda dos pulmões, que percorre a laringe. Segundo Halliday & Resnick (2016, p.344) o som é uma onda mecânica longitudinal, ou seja, uma onda que se propaga no meio material, o qual possui elasticidade, com oscilação em direção a esta propagação. Em outras palavras, Sataloff (2005 p.1) diz que o som é uma forma de movimento onde há a propagação, através de um meio elástico, de ondas de pressão emitidas por um corpo vibrante.

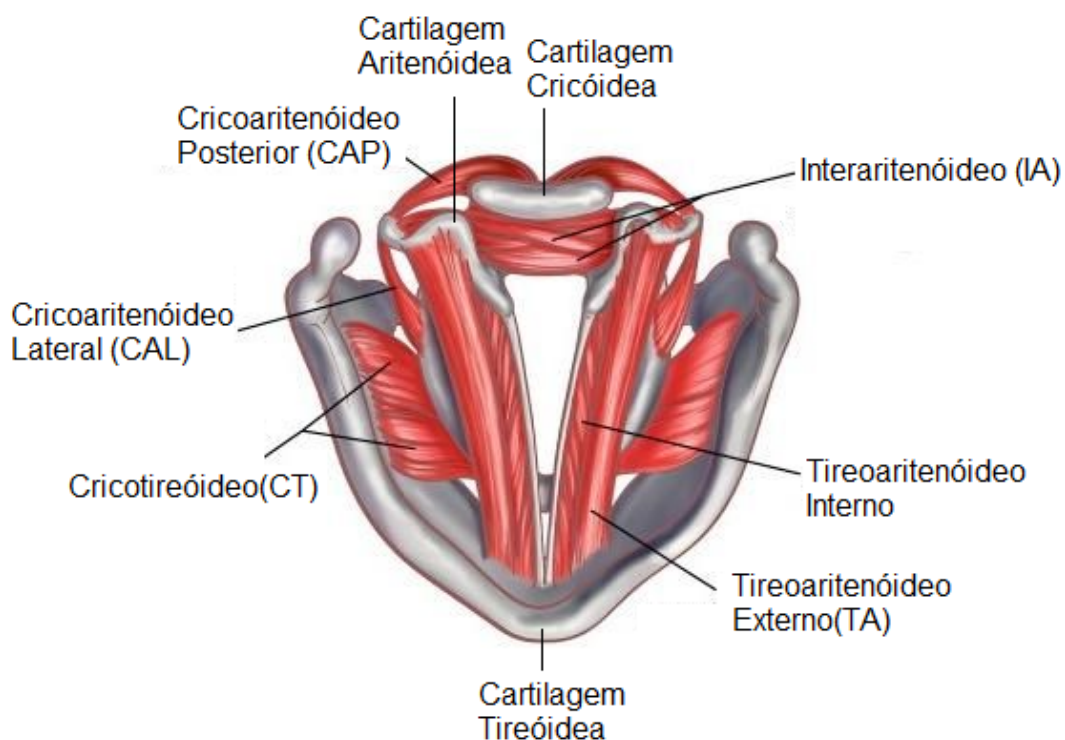
Para melhor compreensão do fenômeno da voz, será apresentada a fisiologia da laringe e o funcionamento da fonação.

2.1.1 Fisiologia da Laringe

De acordo com Pinho *et al* (2006, p.1), a laringe possui três funções. São estas: a proteção (reflexa e involuntária), respiração (podem ser iniciadas voluntariamente, mesmo possuindo controle involuntário) e fonação (voluntária). Esta última é a função menos vital e de surgimento mais tardio. Porém, é a mais estudada nos últimos anos e será alvo de estudos deste trabalho.

A laringe é formada por um conjunto de músculos e cartilagens como pode ser observado na Figura 1, cujos processos de tensão e adução resultam nos movimentos das pregas vocais para a realização de suas funções.

Figura 1 - Músculos e cartilagens da laringe (vista superior).



Fonte: Adaptado de Vocalpedia, <<https://www.vocalpedia.net/single-post/2016/10/06/Los-nódulos-y-el-músculo-vocalis>> . acesso em 11 de abr. de 2018.

O músculo Cricotireóideo (CT) é o principal tensor das pregas vocais. Ao ser contraído, a cartilagem Tireóidea e a cartilagem Cricóidea são rotacionadas. Esse movimento causa o afastamento da Cricóidea, conseqüentemente as pregas vocais são estiradas, adquirindo uma posição vertical à anterior. A contração do

músculo Tireoaritenóideo (TA) promove a adução da prega vocal e tende a encurtar as pregas vocais pela torção exercida sobre a cartilagem aritenóidea. O músculo Cricoaritenóideo Lateral (CAL) é um adutor das pregas vocais. Sua contração traciona a cartilagem aritenóidea, abaixando a cartilagem Cricóidea. O músculo Interaritenóideo (IA), faz parte da adução da região interaritenóidea, fechando a glote cartilaginosa. O músculo Cricoaritenóideo (CAP) é o abductor das pregas vocais (PINHO *et al* ,2006).

O TA, CAL e IA possuem fibras com respostas rápidas e suscetíveis a fadiga. Já o CT e CAP possuem fibras mais lentas, e menos suscetíveis a fadiga. Desta forma, as patologias causadas por fadiga ou mal uso da voz atingem principalmente os músculos mais rápidos (PINHO *et al* ,2006).

2.1.2 Disfonias

Behlau (2004) sugere que as disfonias podem ser classificadas em 3 grupos:

- Disfonias funcionais por uso incorreto da voz, inaptações da voz e alterações psicogênicas;
- Disfonias orgânicas, decorrentes de fatores orgânicos com alterações em órgãos ligados ou não a comunicação;
- Disfonias organofuncionais, causados por fatores orgânicos e funcionais.

Tem-se para o corrente trabalho, dados disponibilizados por Fabiana Zambon da UNIFESP e pelo Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEMCNPq), contendo disфонia funcional, disфонia orgânica (paralisia unilateral de prega vocal) e saudáveis explicados em detalhe na sequência.

2.1.2.1 Disfonia Funcional

As disfonias funcionais são alterações na voz, causadas pelo mal uso da voz, inaptações da voz e alterações psicogênicas, ou seja, este é um distúrbio de comportamento vocal.

2.1.2.2 Disfonia Orgânica

São problemas que não dependem do uso vocal, são causadas por fatores que influenciam diretamente na voz, como doenças neurológicas, inflamações, infecções na laringe, entre outros. Para este trabalho tem-se a patologia paralisia unilateral de prega vocal, como a representante das disfonias orgânicas.

2.1.2.2.1 Paralisia Unilateral de Prega Vocal (PUPV)

Como o próprio nome sugere, esta patologia é caracterizada pela paralisia de uma das pregas vocais. Nesta patologia, a prega vocal afetada não se movimenta em direção à linha média, como mostra a Figura 2, provocando um fechamento glótico incompleto ou ausente, comprometendo a qualidade da voz. Um paciente com PUPV pode apresentar voz fraca, soprosa e aperiódica, devido às diferentes tensões em cada uma das pregas vocais. A proteção das vias aéreas inferiores também fica comprometida, causando engasgos e tosse, pela aspiração de alimentos. (PINHO *et al* , 2006, p.82).

Figura 2 - Paralisia Unilateral de Prega Vocal



Fonte: <<http://www.wohlt.com/vocal-cord-paralysis/>>. Acesso em 11 de abr. de 2018.

2.1.3 Métodos de Avaliação da Voz

Quando se trata de avaliação da voz, existem dois tipos de métodos: os invasivos e os não invasivos, também chamados de instrumentais e não instrumentais como cita Souza *et al* (2011), os quais serão brevemente descritos neste tópico.

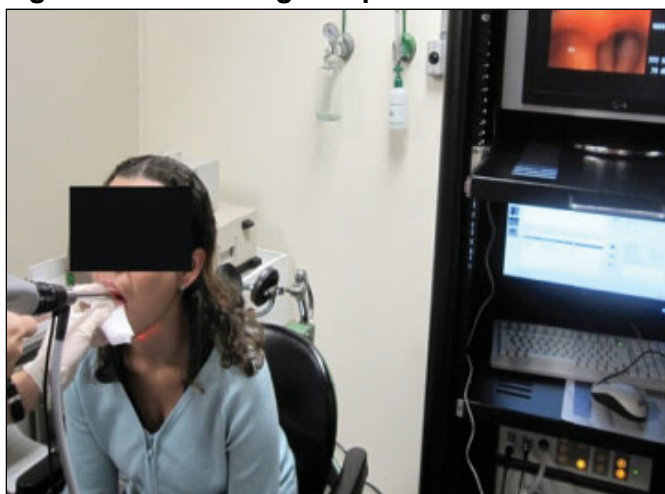
2.1.3.1 Métodos Invasivos de Avaliação da Voz

Os Métodos invasivos de avaliação da voz são caracterizados pela introdução de ferramentas capazes de observar as pregas vocais, como a videolaringoestroboscopia e videolaringoscopia, esta última apresentada na Figura 3. Como citado em Tsuji *et al* (2014), as pregas vocais vibram a uma frequência de 80 a 1000 Hz, enquanto o olho humano é capaz de perceber apenas 5 Hz, tornando inviável a observação da vibração da prega vocal durante a fonação. A videolaringoestroboscopia é um método que envolve a emissão de pulsos de luz a uma frequência de 0,5 Hz ou 1,5 Hz menor que a frequência fundamental de vibração da prega vocal. Dessa forma, com a frequência dos flashes levemente menores que a frequência das cordas vocais, há um atraso de fase entre cada ciclo vibratório iluminado, dando a sensação de *slow motion*, mostrando de forma clara as fases do ciclo glótico. A videolaringoscopia, por sua vez, captura imagens da laringe,

com objetivo de realizar diagnósticos estruturais visuais. A única diferença entre os métodos é a presença dos flashes de luz na videolaringoscopia.

Todavía, de acordo com Gonzáles (2008), este processo pode causar náusea, desconforto e até mesmo dores, já que a ferramenta é introduzida pela garganta.

Figura 3 -Videolaringoscopia de alta velocidade.



Fonte: Adaptado de Tsuji et al (2014).

2.1.3.2 Métodos Não-Invasivos de Avaliação da Voz

2.1.3.2.1 Análise Perceptiva-Auditiva

Esta é a técnica mais utilizada, e oferece informações qualitativas da voz e do comportamento vocal e é considerada o primeiro passo para o diagnóstico de patologias de voz. É analisada comumente por fonoaudiólogos, a partir de áudios ou vídeos gravados, da emissão de vogais sustentadas, fala encadeada ou canto. Nesta técnica avaliam-se: a qualidade vocal, a relação fonte-filtro, o ataque vocal, o tipo articulatório, a presença de distúrbios articulatórios, a ressonância, a velocidade, entre outros fatores (PIMENTA, 2016 , p. 27).

2.1.3.2.2 Avaliação Acústica

A avaliação acústica é baseada na amplitude, periodicidade, duração e composição espectral do sinal de voz, podendo ser utilizada na fase de diagnóstico

dando suporte às imagens videolaringoestroboscópicas, para triagem de pessoas disfônicas e não disfônicas. Permite análise acústica gráfica, garantindo maior eficiência na devolutiva ao paciente no que diz respeito ao entendimento do problema vocal. Sua confiabilidade depende da qualidade no registro dos áudios utilizados para análise acústica, como também na metodologia da análise (PINHO *et al*, 2006, p.82).

2.1.3.2.3 Autoavaliação da Qualidade Vocal

Trata-se de questionários desenvolvidos para auto-avaliação e verificação do impacto de alterações da voz na vida do paciente. Junto aos dados da avaliação perceptiva-auditiva auxiliam de forma não-invasiva a avaliação multidimensional da voz (PIMENTA, 2016, p.29).

2.2 Processamento Digital de Sinais (PDS)

Conforme conceituado por Oppenheim *et al* (1998, p.1), um sinal é algo que contém informação de estado e comportamento de sistemas físicos, representados matematicamente em função de uma ou mais variáveis independentes, normalmente tempo, as quais podem ser contínuas ou discretas.

Para sua manipulação artificial, precisa-se que os sinais, naturalmente contínuos, sejam discretizados, seguindo normas que garantem perda mínima de informação, de modo a alocar o mínimo em memória digital, para o armazenamento de informações, já que memória é um recurso finito. O PDS lida com a transformação de sinais discretizados em amplitude e tempo.

Neste tópico será abordado os conceitos teóricos necessários para aplicação de forma satisfatória destas transformações.

2.2.1 Amostragem

Como dito, os sinais são naturalmente analógicos, ou seja, contínuos no tempo. Contudo os dispositivos de processamento de sinais são digitais, finitos em

armazenamento, amplitude e precisão. Dessa forma, é necessário aproximar o sinal amostrado do real de uma maneira eficiente, preocupando-se com precisão e armazenamento dos dados resultantes.

De acordo com Oppenheim *et al* (1998, p.140), o processo que converte o sinal análogo para um sinal discreto é chamado de amostragem. Recebe este nome devido ao procedimento realizado, onde capturam-se determinadas amostras ($x[n]$), sendo n o índice da amostra que pertence ao conjunto dos números inteiros, a cada intervalo de tempo dado pelo período de amostragem (T_s), a partir do sinal contínuo ($x_c(t)$), delimitado pela taxa ou frequência de amostragem (F_s). Onde

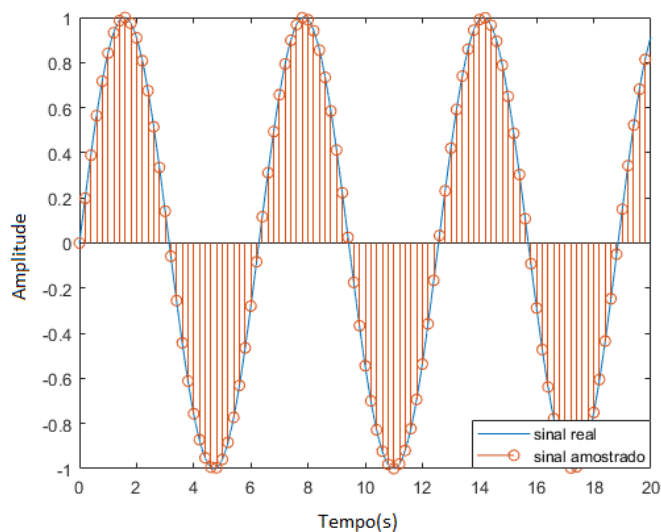
$$T_s = \frac{1}{F_s} \quad (1)$$

Dessa forma temos:

$$x[n] = x_c(nT_s), \quad -\infty < n < \infty \quad (2)$$

Esse procedimento pode ser observado na Figura 4.

Figura 4 - Sinal senoidal amostrado.



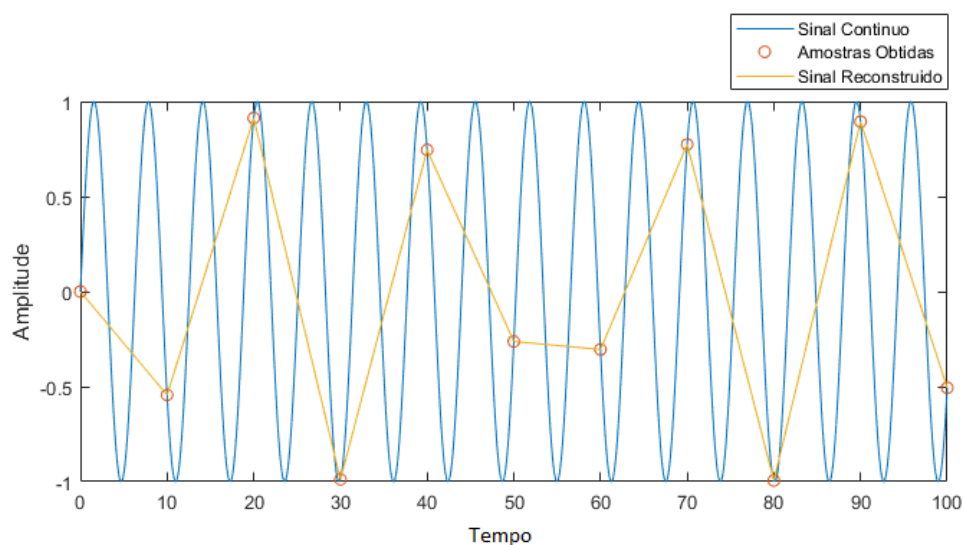
Fonte - Autoria Própria.

2.2.1.1 Aliasing e Teorema de Nyquist

Para garantir a precisão da informação sintetizada ou para que não haja perdas faz-se necessário que o teorema de Nyquist seja respeitado. O teorema sugere que a frequência de amostragem deve ser pelo menos duas vezes maior que a largura de banda do sinal, sendo essa a taxa de Nyquist. Por simplicidade,

assumimos que a largura de banda é igual à frequência máxima do sinal (HAYKIN e VEEN, 2001, p.292). Caso este teorema não seja respeitado, pode ocorrer o fenômeno chamado de *aliasing*, onde o sinal reconstruído não é exatamente igual ao sinal real, pois a taxa de amostragem é muito baixa. Esse fenômeno pode ser observado na Figura 5.

Figura 5 - Fenômeno de aliasing, produzido por taxa de amostragem muito baixa.



Fonte - Autoria Própria.

2.2.2 Reamostragem

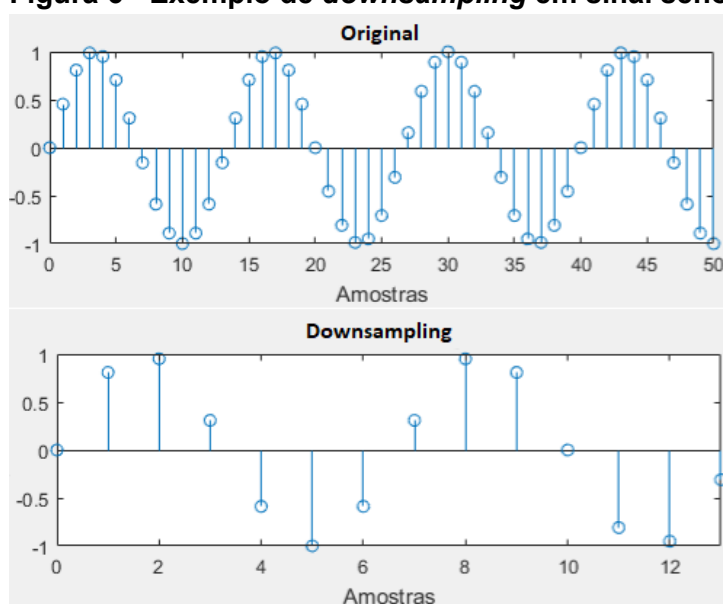
O processo de reamostragem pode ser realizado quando há a necessidade de adequação de dados, compactação de arquivos, entre outros. Consiste na alteração da taxa de amostragem do sinal. Normalmente, a nova taxa de amostragem é reduzida devido ao fato de ser uma melhor opção para o armazenamento de informação, e outros fatores que serão discutidos a seguir (WEEKS, 2007, p. 162).

2.2.2.1 Downsampling

Downsampling ou decimação é o processo de reamostragem onde a nova taxa de amostragem é menor que a taxa de amostragem inicial. Dependendo da taxa de *downsampling*, esse procedimento é realizado retirando amostras de acordo

com a nova taxa, tendo essa como principal vantagem em relação a seu processo inverso, upsampling, apenas são retiradas algumas amostras, e não são criadas novas amostras artificialmente, não há surgimento de dados “fantasmas”, no sinal. Deve-se atentar ao critério de Nyquist, ou seja, a nova taxa de amostragem, deve ser ainda pelo menos duas vezes maior que a largura de banda do sinal original, para que não se perca informação (WEEKS, 2007, p.159). Um exemplo de downsampling é apresentado na Figura 6.

Figura 6 - Exemplo de downsampling em sinal senoidal.



Fonte - Autoria Própria.

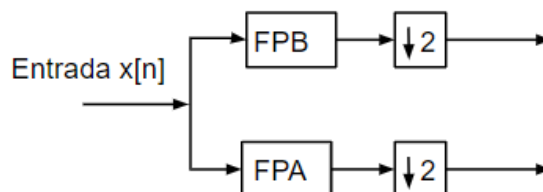
2.2.3 Transformadas *Wavelets*

2.2.3.1 Transformada *Wavelet* Discreta

De acordo com Weeks (2007, p.275), a Transformada Wavelet Discreta (TWD) é utilizada em diversas aplicações de processamento de sinais, como compressão de vídeos, compressão de comunicação na internet, reconhecimento de objetos e análise numérica. O sinal de entrada passa por níveis de decomposição, cada um destes níveis pode ser visto como um par de filtros de resposta impulsiva finita (FIR). Um dos filtros deste par é um filtro passa-baixa (FPB) e o outro um filtro passa-alta (FPA). Cada um deles possui um *downsampler* em sua saída, como

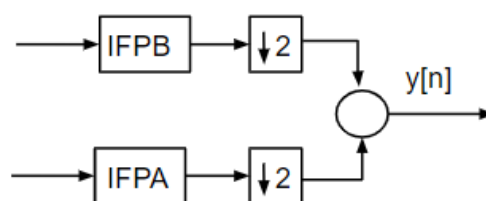
mostra a Figura 7. O processo de reconstrução do sinal é a passagem por um *upsampler* e pelo filtros inversos, como mostra a Figura 8 (WEEKS, 2007, p. 276).

Figura 7 - Decomposição do sinal.



Fonte - Autoria Própria.

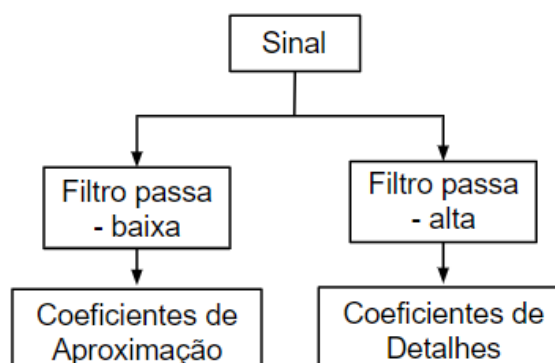
Figura 8 - Reconstrução do sinal.



Fonte - Autoria Própria.

De acordo com Parraga (2002, p.51), o filtro passa-alta produz coeficientes de detalhes da decomposição *Wavelet*, enquanto o filtro passa-baixa está associado com os coeficientes de aproximação da decomposição da *Wavelet*, como apresentados na Figura 9. Para a maioria dos sinais, incluindo os sinais da voz, a maior parte das informações está na componente de baixa frequência.

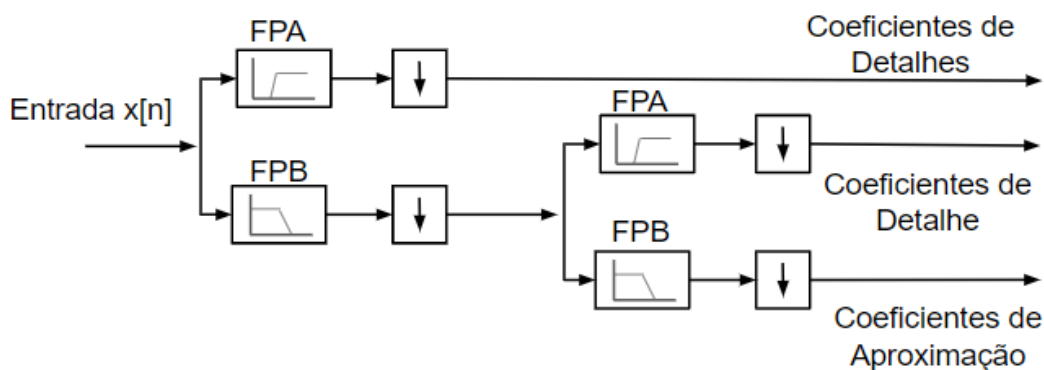
Figura 9 - Coeficientes da decomposição *Wavelet*.



Fonte - Autoria Própria.

Dessa forma, a implementação da TWD pode ser explicada como a sucessiva decomposição do sinal em sub-bandas, obtendo-se a árvore de decomposição da *wavelet*. Decomposição esta que é ilustrada na Figura 10 (PARRAGA, 2002, p.51).

Figura 10 - Árvore de decomposição *Wavelet*.



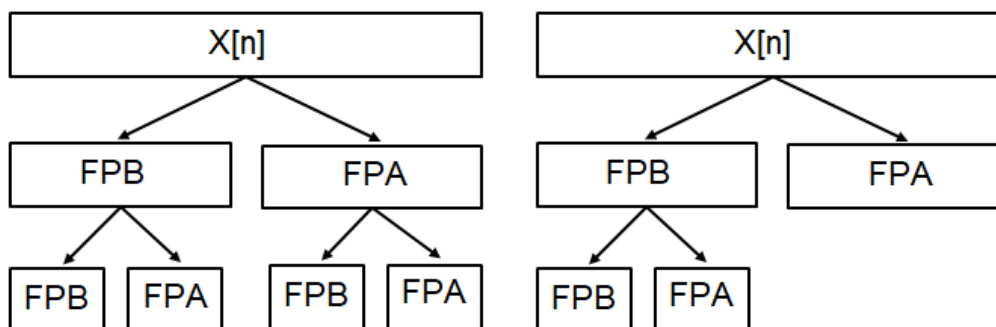
Fonte - Autoria Própria

Os filtros possuem diversos perfis de onda, já sugeridos na literatura e amplamente implementados os quais chamamos de famílias *wavelets*.

2.2.3.2 Transformada *Wavelet Packet* (TWP)

A TWP é uma generalização da TWD. A principal diferença é que na TWD o sinal é dividido em coeficientes de aproximação e detalhes, ao passar pelos dois filtros (passa-baixa e passa-alta). No segundo nível de decomposição apenas os coeficientes de aproximação, são novamente divididos em coeficientes de aproximação e detalhes, já na TWP o mesmo procedimento acontece, porém nos níveis posteriores de decomposição os coeficientes de detalhes também são divididos (PARRAGA, 2002, p. 56). Essa diferença pode ser compreendida pela Figura 11.

Figura 11 - Esquerda: Decomposição para TWP; Direita: Decomposição para TWD.



Fonte - Autoria Própria.

2.2.4 DC Offset

Existem arquivos de áudio onde a componente de frequência zero ou componente DC - *direct current* (corrente contínua), não é de fato zero. Este fenômeno é chamado de DC Offset, o qual dificulta o desempenho da detecção de silêncio. Para realização deste procedimento, faz-se necessária a remoção deste elemento. (LU e HANKINSON, 1998).

2.2.5 Funções do MATLAB

2.2.5.1 Envelope RMS

A função envelope do MATLAB, retorna limitantes superiores e inferiores para determinado sinal. O envelope RMS, por sua vez, retorna limitantes superiores e inferiores calculados a partir do RMS de uma janela deslizante de comprimento 'w'. Neste trabalho, essa janela será uma das ferramentas utilizadas para retirada de intervalos de silêncio do áudio.

2.2.5.2 Detrend

A função '*detrend*' do MATLAB é utilizada para remover linhas de tendência de um sinal. A função calcula a tendência baseada nos mínimos

quadrados e subtrai do original. Removendo de maneira grossa a tendência do sinal. Processo importante para a remoção do DC *offset*.

2.3 Sistemas Inteligentes

2.3.1 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) são sistemas baseados no cérebro humano, e são descrita por Haykin (2001, p. 28), da seguinte maneira: Uma rede neural artificial é um processador maciçamente paralelamente distribuído constituído de unidade de processamento, que armazena conhecimentos e torna-os disponível para o uso. Se assemelhando ao cérebro “humano” em dois aspectos, são estes, o conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem e as forças de conexão entre neurônios, chamadas de pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

Segundo Silva *et al* (2010, p. 21), as RNAs possuem diversas aplicações, entre elas:

- Avaliação de imagens captadas por satélite;
- **Classificação de padrões de escrita e de fala;**
- Reconhecimento de faces em visão computacional;
- Controle de trens de grande velocidade;
- Previsão de ações no mercado financeiro;
- Identificação de anomalias em imagens médicas;
- Identificação automática de perfis de crédito para clientes de instituições financeiras;
- Controle de aparelhos eletrônicos e eletrodomésticos.

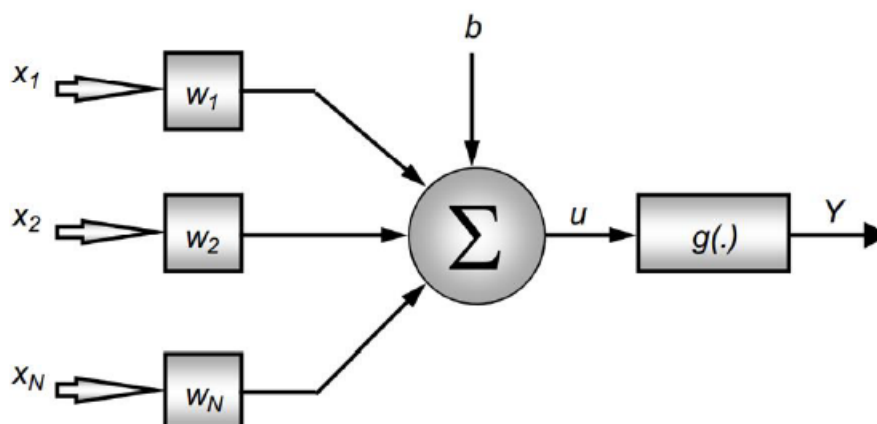
As RNAs possuem características e capacidades que as tornam robustas para a classificação de padrões, como cita Haykin (2001, p. 29) e Silva *et al* (2010, p. 25) a não-linearidade, a adaptação por experiência, capacidade de aprendizado, habilidade de generalização, organização de dados, tolerância a falhas, armazenamento distribuído e facilidade de prototipagem.

As RNAs podem ser lineares ou não lineares. São chamadas adaptativas, pois seus parâmetros internos (pesos sinápticos), são ajustados a partir da apresentação de exemplos relacionados a um determinado padrão, ou seja adquirem conhecimento (se adaptam) a partir de experiências. Com a aplicação do treinamento, a rede é capaz de extrair a relação entre informações que compõem a aplicação. Após o processo de treinamento uma RNA é capaz de generalizar padrões, tornando possível a estimação de soluções. Como a rede é baseada em características e informações de um processo, a rede é capaz de se organizar internamente de modo a possibilitar o agrupamento de padrões. Devido ao grande número de interconexões, a RNA torna-se tolerante a falhas quando parte de sua estrutura interna é sensivelmente corrompida. As informações obtidas a respeito do comportamento de determinado processo são armazenadas em diferentes pesos sinápticos de diferentes neurônios, o que possibilita manter a robustez do sistema mesmo se ocorrer alguma eventualidade de inoperação de algum neurônio. A maioria das arquiteturas de RNAs, são de simples implementação, já que seus resultados são obtidos a partir de algumas operações matemáticas elementares.

2.3.1.1 Neurônio Artificial

Como apresentado por Silva *et al* (2010, p. 33) o modelo de neurônio artificial mais utilizado é o proposto por McCulloch e Pitts (1943), este é o mais simples que engloba as principais características. O modelo pode ser visto na Figura 12.

Figura 12 - Neurônio Artificial.



Fonte - SILVA *et al.*, 2010, p. 34.

Na Figura 12, $\{x_1, x_2, \dots, x_N\}$ são sinais de entrada, $\{w_1, w_2, \dots, w_N\}$ são pesos sinápticos, $\{\Sigma\}$ o Combinador Linear, $\{b\}$ o limiar de ativação, $\{u\}$ o potencial de ativação, $\{g\}$ a função de ativação e $\{Y\}$ o sinal de saída.

Sinais de entrada: São sinais ou medidas do meio externo, apresentando valores assumidos pelas variáveis de uma aplicação específica. Os sinais de entrada são usualmente normalizados para auxiliarem na eficiência computacional dos algoritmos utilizados na aprendizagem. São análogos aos impulsos elétricos externos captados pelos dendritos biológicos.

Pesos sinápticos: São valores utilizados para ponderar as entrada da rede, quantificando as suas relevâncias em relação à funcionalidade de determinado neurônio.

Combinador linear: Tem por função agregar os sinais de entrada ponderados pelos pesos sinápticos, com objetivo de produzir um valor de potencial de ativação.

Limiar de ativação: É uma variável que especifica qual será o patamar apropriado para que o resultado produzido pelo combinador linear possa gerar um valor de disparo em direção à saída do neurônio.

Potencial de ativação: É o resultado produzido pela diferença entre o produzido pelo combinador linear e o limiar de ativação. Se o valor é positivo, ou seja, se $u \geq b$ então o neurônio produz um potencial excitatório; caso contrário, o potencial será inibitório.

Função de ativação: É utilizado para limitar a saída do neurônio dentro de um intervalo de valores razoáveis.

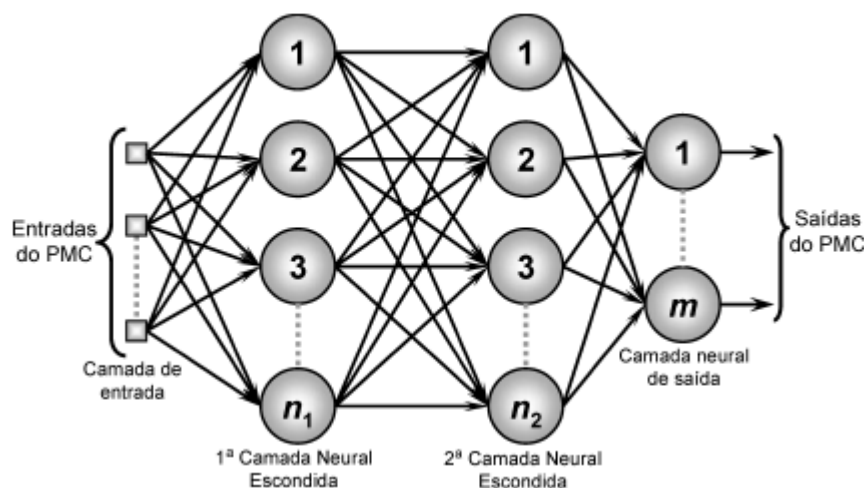
Sinal de saída: Consiste do resultante produzido pelo neurônio em relação ao conjunto de sinais de entrada, podendo ser utilizado por outros neurônios que estão interligados a este.

2.3.1.2 Perceptron Multicamadas

As RNAs do tipo *Multilayer Perceptron* ou *Perceptron Multicamadas* (PMC), têm por principal característica a presença de camadas intermediárias (escondida ou oculta), com o número de camadas intermediárias e números de

neurônios por camada variáveis importantes que dependem do problema analisado, além de uma camada neural de saída. Desta forma esta é composta por uma camada de entrada, camadas neurais intermediárias, e camada neural de saída, como pode ser observado na Figura 13. A camada de saída pode possuir mais de um neurônio, tornando possível a solução de problemas onde as classes não são linearmente separáveis. As PMCs possuem elevadas possibilidades de aplicação nas diversas áreas do conhecimento, sendo muito utilizadas para aproximação universal de funções, reconhecimento de padrões, identificação e controle de processos, previsão de séries temporais e otimização de sistemas (SILVA *et al*, 2010, p.91).

Figura 13 - Perceptron Multicamadas.



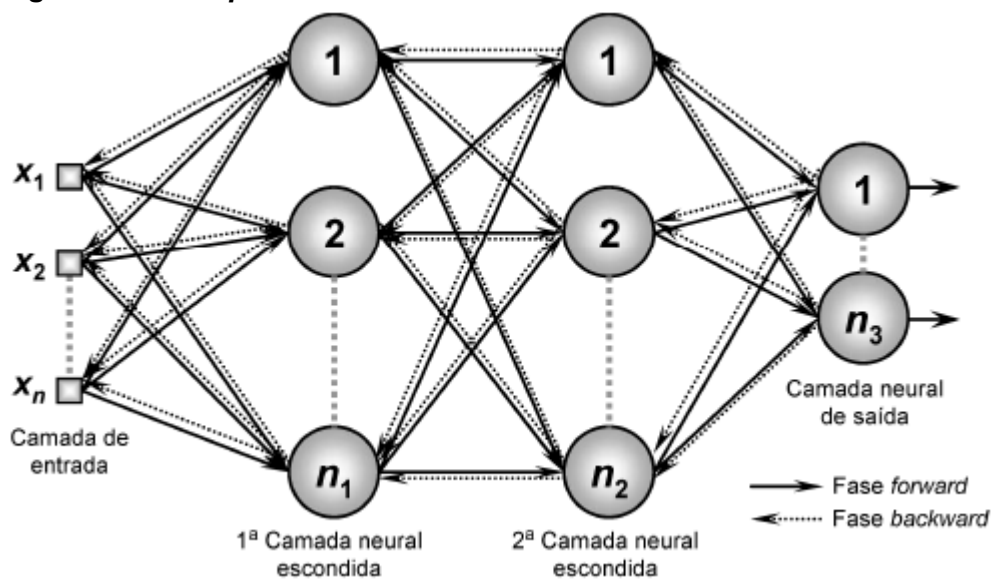
Fonte - SILVA *et al.*, 2010, p. 92.

A PMC, como cita Haykin (2001, p.183), possui arquitetura *feedforward*, ou seja, sem realimentação, com treinamento supervisionado com um algoritmo de retropropagação de erro (*error back-propagation*), algoritmo este baseado nas regras de aprendizagem para correção de erro. Podem ser vistos como uma generalização do algoritmo de filtragem adaptativa, mínimos quadrados médios (LMS).

O processo de aprendizagem, de acordo com Silva *et al* (2010, p.94), é realizado em duas fases, como mostrado na Figura 14, onde na primeira etapa denominada de fase *forward*, os sinais de entrada, de um conjunto de treinamento são inseridas na rede e são propagados camada a camada até a produção da respectiva saída. Em seguida é calculado o erro, da saída com o valor desejado, disponível já que esta rede possui treinamento supervisionado. Estes erros são

utilizados para o ajuste de pesos sinápticos dos neurônios, chamada de fase *backward*.

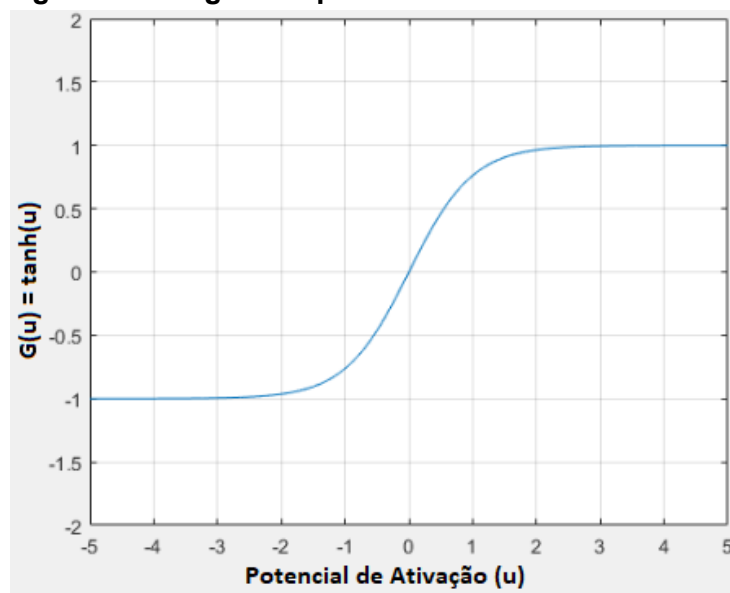
Figura 14 -Perceptron fase forward e backward.



Fonte - SILVA et al., 2010, p. 94.

Segundo Haykin (2001, p. 206) uma PMC, pode em geral aprender mais rápido se a função de ativação for antissimétrica do que quando for não-simétrica. Um exemplo popular de função anti-simétrica é a tangente hiperbólica, como mostra a Figura 15.

Figura 15- Tangente hiperbólica.



Fonte - Autoria própria.

2.3.2 Adaptive Resonance Theory (ART)

Segundo Sivanandam *et al* (2006, p. 277) a Teoria da Ressonância Adaptativa - *Adaptive Resonance Theory* (ART), refere-se a uma classe de arquiteturas neurais auto-organizáveis (não supervisionadas) que agrupam padrões e produzem modelos de vetor de peso apropriados.

As redes neurais artificiais convencionais não conseguiram resolver o dilema plasticidade/estabilidade (aprender sem destruir o conhecimento adquirido no passado) onde o sistema deve ser adaptativo para incorporar mudanças do ambiente, ao passo que ainda é estável na preservação do conhecimento previamente adquirido. Sendo esta a principal capacidade da ART, após sucessivas apresentações de amostras, a rede apresenta plasticidade/estabilidade.

Esta arquitetura foi proposta por Grossberg (1976), baseada em fenômenos biológicos como a visão, fala, o desenvolvimento cortical e as interações cognitivo-emocionais, tendo em vista 3 princípios biológicos:

Normalização de sinais - capacidade destes sistemas em se adaptar a alterações contínuas do ambiente;

Intensificação de contraste - potencial na identificação de detalhes mínimos a partir de sucessiva observação do ambiente;

Memória de curto prazo - possibilidade de armazenamento momentâneo de informações sensoriais provindas de mecanismos de intensificação de contraste (SILVA *et al*, 2010, p .265).

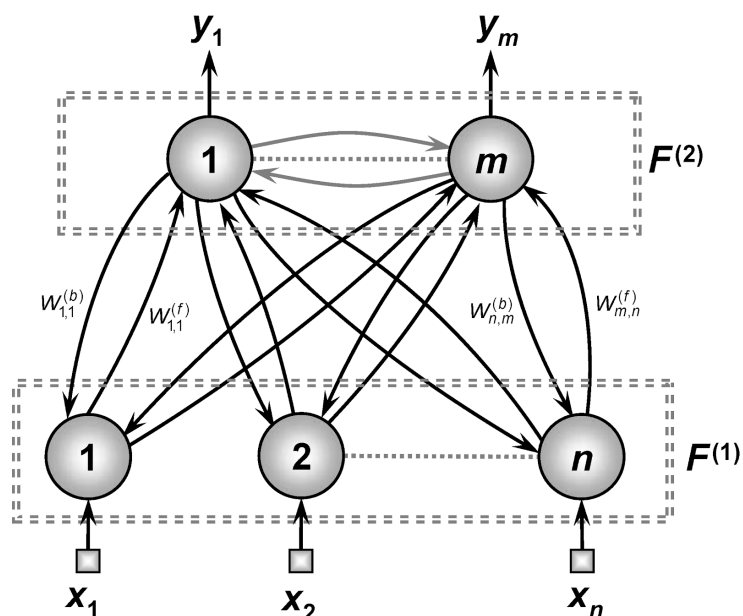
Existe algumas topologias de arquitetura ART, como a ART-1, ART-2, ART-3, ART-Map, etc, porém dada a aplicabilidade e simplicidade de implementação este trabalho terá como foco a ART-1.

A ART-1 possui como entrada padrões binários e treinamento não-supervisionado, possuindo como principal aplicação a classificação de padrões vetorial.

Silva *et al* (2010, p.267) descreve o funcionamento da ART-1 da seguinte forma: Dada uma amostra de entrada vetorial binária, o sistema consiste na classificação destas em uma das classes representativas do problema, usando medidas de similaridade. Como a aprendizagem é não-supervisionada, não há saída desejada.

Como apresentado na Figura 16, a ART-1 possui duas camadas neurais, onde todos os neurônios da camada 1 estão conectados a todos os neurônios da camada 2.

Figura 16 - Estrutura ART-1.



Fonte - SILVA et al., 2010, p. 266.

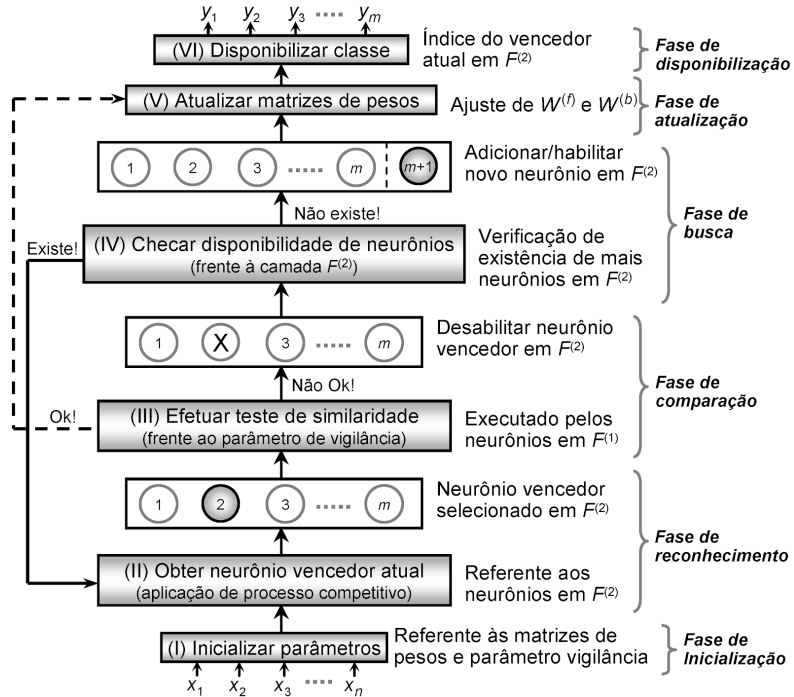
A primeira camada (camada de entrada ou camada de comparação), possui quantidade de neurônios de tamanho equivalente ao tamanho do vetor de entrada. Esta tem por finalidade repassar, por meio de uma matriz de pesos reais $W^{(f)}$, características das componentes do vetor de entrada de modo a estimular a ativação de um dos neurônios da saída, a camada 2. Sendo assim, o fluxo de informação é do tipo *feedforward*, em direção à camada 2.

Já a camada 2, camada de saída ou de reconhecimento, tem a missão do reconhecimento e incorporação do conhecimento trazido de um vetor de entrada a classe representada por um neurônio (SILVA et al., 2010, p. 267).

As informações de realimentação provindos da camada 2 para a 1 (*feedback*), são ponderadas por uma matriz de pesos binários. A missão da camada 1 é comparar as entradas x com essa matriz de pesos, para verificação de similaridade deste com as classes. Se o vetor de pesos do neurônio vencedor forem parecidos com o vetor de entrada, levando-se em conta um parâmetro de vigilância, haverá o processo de ressonância, associando o vetor de entrada ao vetor do neurônio vencedor, por meio da atualização das matrizes de pesos.

A Figura 17 mostra o fluxograma de seu funcionamento.

Figura 17- Fluxograma de Funcionamento da ART-1.



Fonte - SILVA et al., 2010, p. 270.

3 MATERIAIS E MÉTODOS

Esta seção diz respeito a base de dados, bem como aos ferramentais e metodologia adotados para realização do presente trabalho, onde foram utilizados procedimentos já bem descritos na literatura, e outros propostos neste estudo, como a aplicação de técnicas de *clustering* para a garantia de diversidade no treinamento de uma PMC, e cruzamento de famílias TWP, os quais serão detalhados nesta seção.

3.1 Base de Dados

Os dados utilizados neste estudo são arquivos de áudios contendo a vogal /e/ sustentada do português brasileiro, de indivíduos que concordaram em participar de pesquisas relacionadas ao estudo da voz e foram aprovados pelos comitês de ética dos locais onde foram coletados. Os arquivos de áudios estão divididos em 3 grupos: vozes saudáveis, vozes com disfonia funcional e vozes com disfonia orgânica.

Os áudios de disfonia funcional foram coletados na Universidade Federal de São Paulo (UNIFESP) e cedidos por Fabiana Zambon responsável pela coleta, de 30 professoras com idade entre 23 e 49 anos, com média de 26,2 anos e desvio padrão de 7,5, todas do sexo feminino. Os áudios foram capturados utilizando uma placa de som externo *Andrea PureAudio* USB e microfone Auricular *karsect* HT-2 com taxa de amostragem de 44100 Hz, como descrito por Zambon (2011, p. 13). O microfone para captura dos sinais foi posicionado a 2 cm de distância da boca, em um ângulo de 45°.

Os áudios de disfonia orgânica e saudáveis foram coletados no Ambulatório de Voz do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) e cedido pelo Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEMCNPq), para realização deste trabalho. Contendo informações de voluntários do sexo masculino e feminino, contando com 45 áudios de disfonia orgânica, mais precisamente paralisia unilateral de prega vocal e 51 saudáveis. Foram utilizados

para a captura dos áudios o *software Audacity* junto a um computador *Pentium II*, Placa de som e microfone *headset* unidirecional AKG-C520 L, com taxa de amostragem de 44100 Hz. Contudo a aquisição dos arquivos foi feita durante um longo período de tempo e, devido às capturas mais antigas terem sido realizadas por placa de aquisição de som desatualizada, parte dos arquivos possuem taxa de amostragem de 22050 Hz. O microfone foi situado a 4 cm de distância da boca, em um ângulo de 45°, como descrito por Pimenta (2016).

Os arquivos gerais que compõem este estudo estão apresentados na quadro 1 .

Quadro 1 - Arquivos de voz.

CARACTERÍSTICA	TAXA DE AMOSTRAGEM (HZ)	QUANTIDADE
Funcional	44100	30
Orgânica	44100/22050	45
Saudáveis	44100/22050	51
TOTAL	-	126

Fonte: Autoria Própria.

3.2 Ferramentas

Todos os procedimentos deste trabalho foram realizados com auxílio da ferramenta de desenvolvimento MATLAB®, plataforma esta que facilita o desenvolvimento e automatização de rotinas.

O MATLAB possui *toolboxes* consolidadas de processamento digital de sinais e redes neurais artificiais, as quais auxiliam na extração de características, estudo do sinal, classificação e análise de resultados em um mesmo ambiente.

3.3 Procedimentos

Segundo Duda *et al* (2001, p. 4) um bom projeto de classificador de padrões deve ser realizado nas seguintes etapas: pré-processamento, segmentação e agrupamento, extração de característica, classificação e pós-processamento.

O pré-processamento é a etapa responsável por adequar e normalizar o sinal. É neste momento que possíveis desníveis, problemas com aquisição, ruídos indesejados, entre outros fatores, precisam ser corrigidos.

A segmentação e agrupamento são de exímia importância à classificação de padrões, além de serem a ponte perfeita entre os dados pré-processados e a extração de características. Estas etapas são responsáveis por segmentar ou fracionar os dados em conjuntos menores com objetivo de extrair características de grupos com menos informação, resultando em maior diversidade de características, além de ser imprescindível na geração de novas amostras para o treinamento da RNA.

Extração de características é o processo onde há a transformação dos dados em informações mensuráveis ou características de determinado dado. A grande dificuldade nesta etapa é realizar a extração de forma que dados semelhantes estejam em apenas uma classe, um processo que pode se tornar penoso ao projetista do extrator.

Classificação é a generalização de características, dentro de grupos pré-definidos (treinamento), para agrupamento por meio de regras de dados apresentados ao sistema treinado.

Pós-processamento é a etapa onde se analisa o resultado obtido pelo classificador. É neste momento que podemos definir a qualidade do treinamento realizado, a confiabilidade do resultado, bem como a acurácia da ferramenta.

3.3.1 Pré-processamento do Sinal

Para garantir a qualidade dos resultados foi preciso atentar-se a alguns detalhes dos sinais de voz que podem interferir na classificação final, como por exemplo: a uniformidade das taxas de amostragem entre os áudios, momentos de

silêncio, *DC offset* e artefatos nos sinais, que podem provir da ferramenta ou interferências externas. Porém, a remoção ou adequação destes fatores deve ser realizada com cautela, já que uma informação relevante para uma disфонia pode ser um ruído na voz, o que dificulta a aplicação de filtros. As formas como estes fatores foram resolvidos serão tratadas neste tópico.

Devido ao fato de haver diferentes taxas de amostragem (44100 Hz e 22050 Hz) dentre os arquivos de áudio, para garantir a uniformidade das informações, foi realizada a reamostragem, com o objetivo de uniformizar os áudios. Para este estudo foi escolhido realizar um *downsampling* de 2 (redução da taxa de amostragem) de 44100 Hz para 22050 Hz. Já que o processo inverso -- o *upsampling* -- gera pontos artificiais a partir da interpolação, podendo danificar informações do sinal, como citado.

Após a adequação das taxas de amostragem dos áudios, outro problema resolvido foi um possível *DC offset*. Para solucionar isso, foi utilizada a função *detrend* do MATLAB para eliminar possíveis desvios do sinal. Esta função busca uma linha de tendência do sinal capturado a partir do método dos mínimos quadrados e subtrai esta função do sinal, o que remove o *offset* de corrente contínua do áudio.

Em seguida, pôde-se iniciar o processo de remoção de silêncio do áudio. Para isso, foi implementado um filtro onde é aplicada uma função do MATLAB envelope 'RMS' no sinal com janela de 1000 amostras. Foram calculados limiares com base nos mínimos e médias do sinal envelopado. Desse modo, uma amostra foi selecionada para compor o cálculo do limiar se fosse menor que a média do envelope dividido por 2 e maior que o mínimo do envelope vezes 2 com mostra a equação 3, limiares selecionados empiricamente a partir de testes.

$$L(m) = Env(n), \text{ se } \min(Env) \times 2 < Env(n) < \text{média}(Env)/2 \quad (3)$$

Onde $L(m)$ são as m amostras utilizadas para o cálculo do limiar de corte, $Env(n)$ é o valor do envelope para cada amostra n do sinal original.

O limiar final foi igual a média das amostras que compõem o limiar anterior, com apresentado na Equação 4.

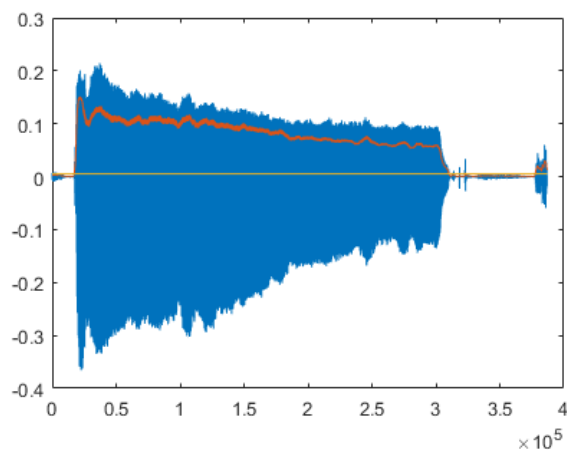
$$L_f = \text{média}(L) \quad (4)$$

Sendo L_f o limiar médio, calculado a partir das amostras que compõem limiar.

Dessa forma, o sinal resultante foram todos os pontos com amplitude maior que o limiar calculado, eliminando-se assim os momentos de silêncio, ponto a ponto. Além disso, as janelas iniciais e finais do sinal foram removidas, para que somente um intervalo em que há constância na voz seja utilizado.

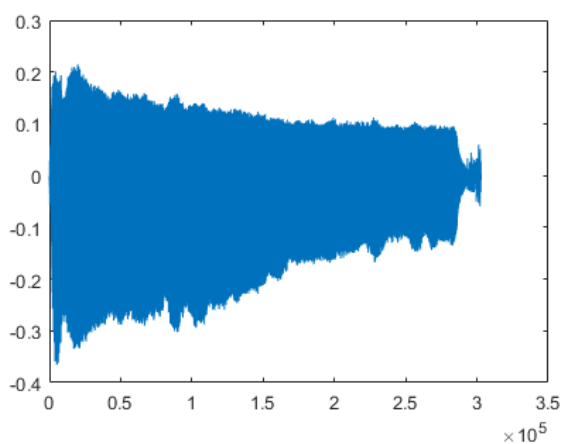
As Figuras 18 e 19 mostram o resultado deste processo, na Figura 18 pode-se observar o sinal original em azul, o envelope RMS calculado em vermelho e o limiar de corte em amarelo. O sinal resultante pode ser observado na Figura 19.

Figura 18- Sinal original em azul, o envelope RMS calculado em vermelho e o limiar de corte em amarelo.



Fonte - Autoria Própria.

Figura 19 - Sinal sem silêncio.



Fonte - Autoria Própria.

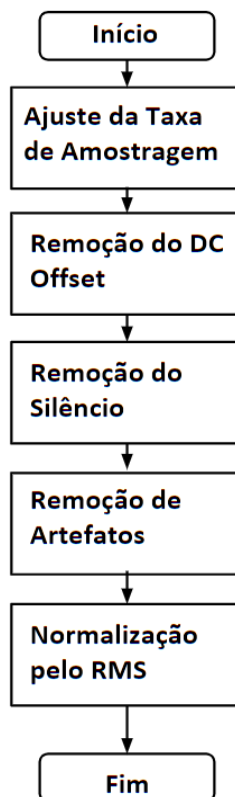
Após a remoção de um possível *DC offset* e dos intervalos de silêncio foi preciso remover trechos de áudio onde há interferências externas que podem confundir a classificação das disfonias. Portanto, foi necessário remover trechos que possuem: vozes de outras pessoas, artefatos provindos dos equipamentos de captura de áudio, engasgos, tosses, entre outros. Para realização deste procedimento foi necessário ouvir cada um dos áudios e remover manualmente cada um destes trechos. É recomendado não aplicar filtros, pois vozes disfônicas podem apresentar ruídos aditivos semelhantes aos que seriam filtrados. Por fim foi feita a normalização pelo RMS de todos os sinais concatenados, conforme a Equação 5.

$$v_f(n) = v_0(n) / \sqrt{\frac{1}{n_T} \cdot \sum_{i=1}^{n_T} |v_0(i)|} \quad (5)$$

Onde $v_f(n)$ é a amostra final calculada $v_0(n)$ é a amostra inicial, n_T é o número total de amostras.

O pré-processamento completo pode ser visto a partir do fluxograma da Figura 20.

Figura 20 - Fluxograma do pré-processamento.

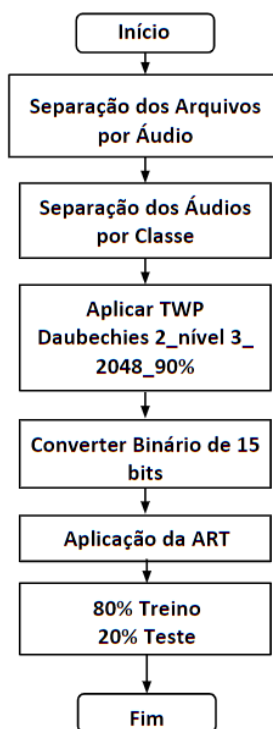


Fonte - Autoria Própria.

3.3.2 Segmentação

A etapa seguinte foi a separação dos arquivos nos conjuntos de treinamento e teste que foram utilizados na PMC para cada uma das classes (saudáveis, disfonia orgânica e disfonia funcional). A ferramenta utilizada para a separação dos arquivos em treinamento e teste foi a RNA do tipo ART. Para redução de dados foi extraída a energia da TWP com configuração genérica: família Daubechies 2 com nível 3 de decomposição, janelamento de 2048 amostras e 50% de sobreposição das janelas, facilitando a utilização da ART. As energias de cada áudio foram transformados para binários de 15 bits com range de 0 a 100. Para garantir maior diversidade no treinamento da PMC utilizada para classificação das patologias, os conjuntos de informações obtidas para cada grupo (saudáveis, funcionais e orgânicos) foram analisados separadamente pelo sistema ART com parâmetro de vigilância igual a 0,6 para verificação da semelhança entre os áudios de cada grupo. Assim, foram separados 80% dos arquivos para treinamento e 20% para teste. Este processo pode ser visto no fluxograma da Figura 21.

Figura 21 - Fluxograma da segmentação.



Fonte - Autoria Própria.

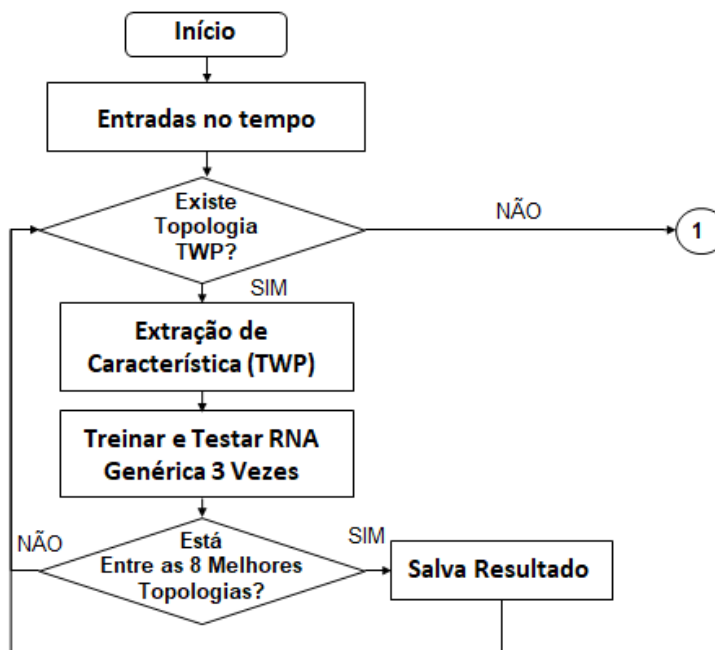
3.3.3 Extração de Características

Para a etapa de extração de característica foi escolhida a TWP, por extrair características do sinal tanto no tempo quanto na frequência. Além de ter demonstrado bons resultados ao ser aplicada junto às RNAs, como pode ser visto em Fermino, et al (2016), Pavoni, et al (2016), Martins, et al (2018), Barros, et al (2017), Lima, et al (2018). A TWP foi aplicada nos sinais no tempo, a aplicação prévia da mesma na etapa anterior foi realizada apenas com intuito de redução de amostras.

Para a definição de quais configurações TWP, a qual chamaremos de topologia TWP, que foram utilizadas neste trabalho, foi implementado um sistema automatizado, com funcionamento apresentado no diagrama da Figura 22, cujas entradas foram os áudios no tempo pré-processados, apresentados anteriormente.

Com objetivo de verificar qual topologia possui melhor resultado na classificação junto aos arquivos de áudio, um conjunto de topologias TWP que possui as variantes descritas conforme a Tabela 1 (Todas as possíveis configurações foram geradas), foi treinado e testado junto a RNAs do tipo PMC, com configuração apresentada na Tabela 2. As entradas consideradas são: a energia, o logaritmo da energia e a entropia de Shannon (informações obtidas pela *toolbox* do MATLAB). O procedimento descrito foi realizado três vezes, com RNAs especialistas em apenas uma característica (saudáveis, disfonia orgânica e disfonia funcional). Ou seja, uma RNA é capaz de identificar se uma amostra é saudável ou não, outra é capaz de identificar se uma amostra possui disfonia orgânica ou não e a última é capaz de identificar se uma amostra possui disfonia funcional ou não. Assim, a saída da PMC será do tipo *one-of-c-class*, onde a saída será [1 0] caso a amostra pertencer à classe cujo a rede é especialista e [0 1] caso contrário, a Tabela 3, mostra a distribuição das saídas para cada uma das RNAs especialistas. Dessa forma saudáveis, disfonia orgânica e disfonia funcional são características das redes, e pertence ou não pertence são classes de cada uma das redes.

Figura 22 - Fluxograma da seleção das topologias TWP iniciais.



Fonte - Autoria Própria.

Tabela 1 - Variações de topologias TWP.

Tamanho da Janela	1024	2048	4096	8192
Sobreposição (overlap(%))	50	75	85	90
Nível de Decomposição	3	4	5	-
Família TWP	'db2' a 'db10'	'sym2' a 'sym10'	'coif1' a 'coif5'	-

Fonte - Autoria Própria.

Tabela 2 - Configuração utilizada para a PMC.

Treinamento PMC	
Camadas Intermediárias	3
Neurônios por Camada	[10 9 7]
Épocas	200
Taxa de Aprendizagem	0.3
Algoritmo de Aprendizagem	Levenberg-Marquardt
Função de Ativação (interm.)	Tangente Hiperbólica
Função de Ativação (saída)	Rampa Linear

Fonte - Autoria Própria.

Tabela 3 - Saídas das RNAs especialistas.

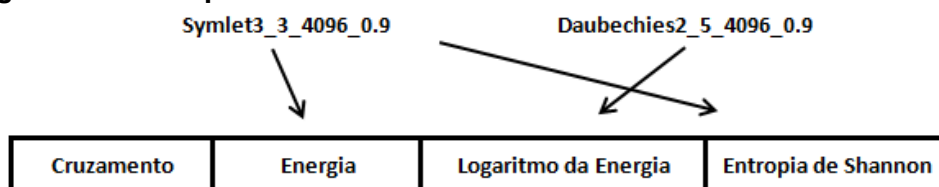
Saídas respectivas a cada uma das 3 RNAs Especialistas				
RNA Especialista				
Saudáveis Funcionais Organicos				
	Saudáveis	[1 0]	[0 1]	[0 1]
Saídas	Funcionais	[0 1]	[1 0]	[0 1]
	Organicos	[0 1]	[0 1]	[1 0]

Fonte - Autoria Própria.

Nas iterações do sistema, as topologias foram utilizadas para treinar e testar as RNAs três vezes cada. O resultado de cada classificação foi salvo. Nas iterações seguintes, caso ainda existissem topologias TWP a serem testadas, o treinamento e teste junto a RNA padrão foram realizados novamente, ou seja, este procedimento foi realizado até que não houvessem mais topologias a serem testadas. Foram selecionados os 8 melhores resultados, Como mostra o fluxograma da Figura 22.

Caso não houvessem mais topologias a serem testadas, os dados de saída (energia, logaritmo da energia e entropia de Shannon) foram cruzados, treinados e testados 3 vezes junto à RNA, em busca de novos melhores resultados caso possível. A notação designada para este procedimento foi “família TWP_nível de decomposição_tamanho da janela_sobreposição da janela”. Somente foi possível realizar o cruzamento caso o janelamento de dados e a sobreposição de dados forem iguais entre as topologias, para garantir a equivalência entre as amostras cruzadas. Como mostra a Figura 23.

Figura 23 - Exemplo de Cruzamento e TWP.



Fonte - Autoria Própria.

Dez melhores topologias, dentre as oito iniciais mais seus cruzamentos possíveis foram selecionadas. Cada uma delas foi treinada e testada 10 vezes com objetivo de garantir que a inicialização aleatória dos pesos sinápticos da PMC não interfira no resultado final. Eliminou-se 30% dos piores resultados e foi calculada a média dos resultados restantes, com mostra a Figura 24, este procedimento foi chamado de torneio, com fluxograma apresentado na Figura 25, a qual foi utilizada para comparação e seleção da melhor topologia. O processo de cruzamento é apresentado no fluxograma da Figura 26.

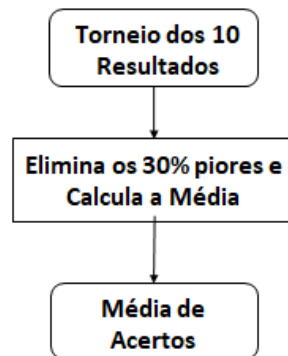
Figura 24 - Cálculo da média de acertos, para seleção da topologia TWP.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	Média
Taxa total de acertos (%)	91,59	92,46	85,27	91,36	87,50	92,05	85,42	95,00	94,36	95,27	93,15

30% - eliminado

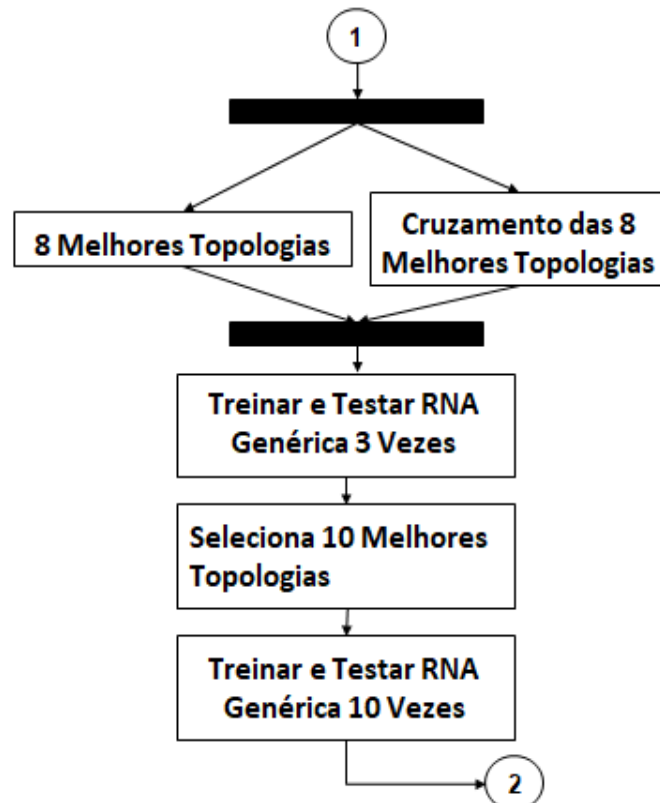
Fonte - Autoria Própria.

Figura 25 - Fluxograma do torneio.



Fonte - Autoria Própria.

Figura 26 - Fluxograma do cruzamento TWP.



Fonte - Autoria Própria.

Logo após este procedimento foram testadas, junto à RNA de configuração citada, diferentes configurações retirando informações da melhor topologia de TWP com objetivo de reduzir os dados mantendo alta acurácia na classificação. Dessa forma foram realizados testes conforme a Tabela 4. Seguindo a metodologia adotada anteriormente foram realizados 10 testes, 30% são eliminados, e foi calculada a média dos restantes.

Tabela 4 - Testes para redução do vetor de amostras de entradas.

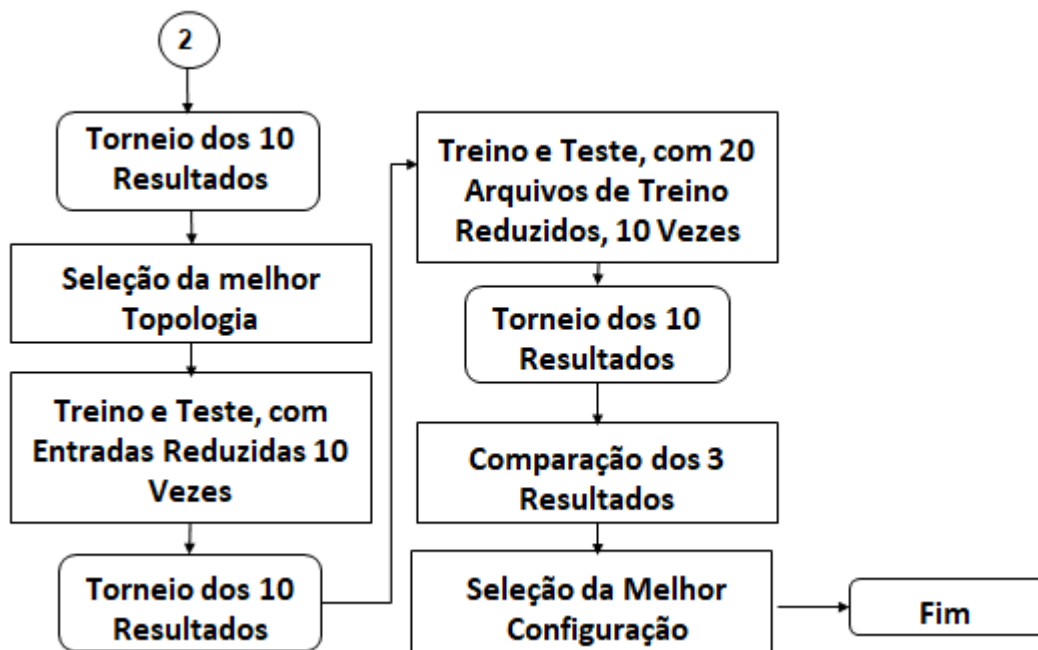
Inicial	[Energia Logaritmo da Energia Shannon]
Teste 1	[Energia]
Teste 2	[Logaritmo da Energia]
Teste 3	[Shannon]
Teste 4	[Energia Logaritmo da Energia]
Teste 5	[Energia Shannon]
Teste 6	[Logaritmo da Energia Shannon]

Fonte - Autoria Própria.

Durante as etapas anteriores foram utilizadas 80% das amostras totais para treinamento, dessa forma é fácil entendermos que a quantidade de amostras referentes a não pertencer a característica apresentada a uma rede especialista é maior que os dados referentes ao pertencer a mesma, o que pode causar *overfitting* na PMC. Por exemplo para treinamento da RNA especialista em vozes saudáveis existem 6954 amostras de vozes, sendo 3127 saudáveis e 3827 disfônicas, ou seja, mais amostras que não pertence à característica do que amostras que pertencem à característica. Por tal motivo foram separados 20 configurações de treinamento com seleção de números equivalentes de amostras em cada uma das classes, com objetivo de aumentar a acurácia do sistema. A seleção foi realizada a partir de uma permutação aleatória e seleção da quantidade que equivale à quantidade de amostras que pertencem à característica.

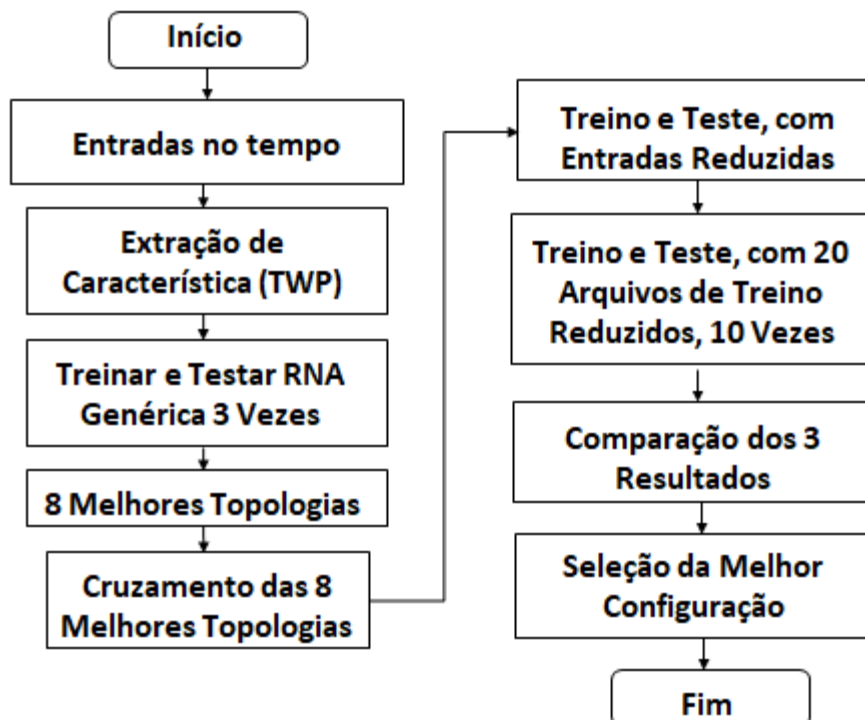
Por fim, foi selecionada a topologia e conjunto de treinamento que obteve melhor resultado. O processo de seleção com entradas reduzidas e amostras reduzidas podem ser observados no fluxograma da Figura 27 e a planificação do processo completo de extração de característica pode ser visto na Figura 28.

Figura 27 - Fluxograma da redução de entradas e amostras.



Fonte - Autoria Própria.

Figura 28 - Fluxograma completo da extração de características.



Fonte - Autoria Própria.

3.3.4 Classificação

Finalmente, para otimização dos parâmetros internos das PMCs, as entradas foram normalizadas com intervalo de 0 a 1, e enviadas a um Sistema Inteligente Híbrido (SIH), proposto por Yoshizumi et al (2018). O SIH é composto pela RNA do tipo PMC, responsável pela classificação e pelo *Particle Swarm Optimization* (PSO), ou Otimização por Enxame de Partículas. Responsável por otimizar número de camadas variando de 1 a 3, número de neurônios por camada variando de 0 a 63 e taxa de aprendizagem variando de 0.025 a 0.5, da PMC, o sistema foi configurado com 15 partículas e 20 iterações. Como se pode observar na Tabela 5.

Tabela 5 - Gama de ação do SIH.

	Quantidade de Neurônios			Taxa de Aprendizagem
	1ª Camada	2ª Camada	3ª Camada	
Mínimo	0	0	0	0.0025
Máximo	63	63	63	0.5

Fonte - Autoria Própria.

O sistema de classificação final foi composto por três RNAs especialistas, onde cada uma delas terá por objetivo apontar se determinado dado pertence ou não a um grupo específico de disfonia (funcional e orgânica) e saudável, como apresentado por Lima *et al* (2018), por isso são chamadas RNAs especialistas.

3.3.5 Pós-Processamento

A partir das respostas obtidas nos testes pelas RNAs, foi realizado um ajuste de confiabilidade da rede (limiar de saturação) para 0,98, ou seja, classificações <0,02 recebem 0 e >0,98 recebem 1, caso contrário, considera-se incerteza. Em seguida, podem ser analisadas as porcentagem de acertos da RNA, a partir da matriz de confusão.

4 RESULTADOS

4.1 Resultados

Nesta seção, serão apresentados os resultados, baseando-se na metodologia previamente apresentada.

A partir da metodologia apresentada e com os arquivos de áudios pré-processados, foi possível agrupar os arquivos em treinamento e teste para a classificação, com auxílio da RNA do tipo ART, para isso os arquivos foram enumerados, de modo a facilitar a visualização do agrupamento. Foram selecionados para compor o teste 20% dos arquivos. Foram selecionados aqueles com maior similaridade com o restante. Estes podem ser observados nas Tabelas 6, 7 e 8, junto aos arquivos aos quais estes são similares. Os 80% restante dos arquivos foi utilizado para o treinamento das RNAs como citado anteriormente.

Tabela 6 - Arquivos de voz de teste e arquivos similares disфонia orgânica.

Arquivos Disфонia Orgânica	Similaridade
2	[1 3 4 12 14 16 19 20 23 24 27 28 29 30 33 34 35 40 45]
8	[1 6 13 15 17 26 28 37 39 42 43]
10	[3 4 9 12 20 22 23 24 26 31 33 34 36 37 40]
15	[1 6 8 13 17 26 28 37 39 42 43]
18	[4 6 9 14 24 30 33 36 37 40 42]
28	[1 2 3 4 6 8 12 13 14 15 16 17 19 20 23 24 26 27 29 30 33 34 35 37 39 40 42 43 45]
35	[1 2 3 4 12 14 16 19 20 23 24 27 28 29 30 33 34 40 45]
36	[3 4 6 9 10 12 14 18 20 22 23 24 30 31 33 34 37 40 42]
37	[1 3 4 6 8 9 10 12 13 14 15 17 18 20 22 23 24 26 28 30 31 33 34 36 39 40 42 43]
44	[7 14 33 34]

Fonte - Autoria Própria.

Tabela 7 - Arquivos de voz de teste e arquivos similares disфонia funcional.

Arquivos Disфонia Funcional	Similares
1	[2 3 5 12 15 18 19 27 30]
3	[1 2 5 11 12 13 14 15 18 19 25 29 30]
4	[2 8 11 13 14 20 23 25 29]
5	[1 2 3 12 15 18 19 27 30]
22	[7 9 16 17 25 26 28]
23	[2 4 8 11 13 14 20 25 29]
27	[1 5 7 12 15 18 19 30]
30	[1 2 3 5 12 15 18 19 27]

Fonte - Autoria Própria.

Tabela 8 - Arquivos de voz de teste e arquivos similares Saudáveis.

Arquivos Saudáveis	Similares
1	[3 5 7 8 11 22 23 24 32 33 36 44 47 49 50]
2	[7 8 11 13 14 19 21 22 23 24 25 31 32 38 41 42 43 44 49]
7	[1 2 8 11 13 14 19 21 22 23 24 25 31 32 38 41 44 49]
9	[5 10 11 12 15 18 21 30 34 36 38 39 41 42 45 47 48 49 50]
14	[2 7 8 11 13 21 22 23 24 25 31 32 38 41 44 49]
33	[1 3 5 8 11 23 24 36 47 49 50]
39	[5 9 10 11 12 15 18 21 30 34 36 38 41 42 45 47 48 49 50]
41	[2 5 7 8 9 10 11 12 13 14 15 18 21 22 23 24 25 30 31 32 34 36 38 39 42 44 45 47 48 49 50]
45	[5 9 10 11 12 15 21 24 29 30 35 39 41 47 49 50]
48	[8 9 10 11 12 15 18 21 30 31 34 36 37 38 39 41 42 47 50]

Fonte - Autoria Própria.

Com os arquivos de treinamento e teste separados iniciou-se o processo de seleção da topologia TWP. A partir dos 9936 testes totais realizados junto à PMC, variando as configurações conforme a Tabela 1 por 3 vezes, para cada uma das 3 RNAs especialistas, foi possível identificar a 8 melhores topologias para cada uma das RNAs especialistas. A topologia WPT é descrita da seguinte forma: “família TWP_nível de decomposição_tamanho da janela_sobreposição da janela”. As Tabelas 9, 10 e 11 apresentam as 8 melhores topologias encontradas para a classificação na RNA especialista em saudáveis, disfonias funcionais e disfonias orgânicas, respectivamente.

Tabela 9 - 8 melhores topologias TWP iniciais(Saudável).

Symlet 3_5_8192_0.9
 Symlet 3_3_8192_0.85
 Symlet 2_3_1024_0.9
 Symlet 2_3_1024_0.5
 Daubechies 2_5_8192_0.85
 Daubechies 2_4_1024_0.85
 Daubechies 2_3_8192_0.85
 Daubechies 2_3_1024_0.85

Fonte - Autoria Própria.

Tabela 10 - 8 melhores topologias TWP iniciais (Disfonia Funcional).

Symlet 8_4_4096_0.85
 Symlet 7_4_4096_0.85
 Symlet 7_3_1024_0.5
 Symlet 6_4_8192_0.9
 Symlet 6_4_4096_0.9
 Symlet 5_4_1024_0.75
 Daubechies 9_3_4096_0.85
 Daubechies7_4_1024_0.85

Fonte - Autoria Própria.

Tabela 11 - 8 melhores topologias TWP iniciais (Disfonia Orgânica).

Symlet 8_3_4096_0.75
Symlet 2_5_4096_0.5
Daubecheis 2_3_4096_0.5
Daubechies 9_3_8192_0.5
Daubechies 8_4_8192_0.9
Daubechies 8_3_8192_0.9
Daubechies 2_4_4096_0.75
Coifflets 1_3_2048_0.75

Fonte - Autoria Própria.

Com as 8 melhores topologias selecionadas, foi possível iniciar os cruzamentos das topologias. As Tabelas 12, 13 e 14 apresentam as 10 melhores topologias, junto à média de acertos totais obtida, calculada conforme descrito na Metodologia deste trabalho para cada uma das RNAs especialistas após os cruzamentos.

Tabela 12 - 10 melhores topologias TWP após cruzamento (Disfonia Orgânica).

Média de Acertos (%) - 10 Melhores Topologias Disfonia Orgânica								
Energia		Logaritmo da Energia		Entropia de Shannon		Janela	Overlap	
Família	Nível	Família	Nível	Família	Nível			
Db2	3	Db2	3	Sym2	2	4096	0,5	95,79
Db2	3	Db2	3	Db2	3	4096	0,5	94,92
Db8	3	Db8	3	Db8	3	8192	0,9	93,15
Db8	4	Db8	3	Db8	3	8192	0,9	92,42
Db8	4	Db8	4	Db8	3	8192	0,9	92,31
Sym2	5	Sym2	5	Db2	3	4096	0,5	92,22
Db2	3	Sym2	5	Db2	3	4096	0,5	91,83
Sym8	3	Sym8	3	Sym8	3	4096	0,75	91,77
Db8	4	Db8	3	Db8	4	8192	0,9	91,44
Sym8	3	Sym8	3	Db2	4	4096	0,75	90,18

Fonte - Autoria Própria.

Tabela 13 - 10 melhores topologias TWP após cruzamento (Saudável).

Média de Acertos (%) - 10 Melhores Topologias Saudável								
Energia		Logaritmo da Energia		Entropia de Shannon		Janela	Overlap	
Família	Nível	Família	Nível	Família	Nível			
Db2	5	Sym3	3	Db2	5	8192	0,85	96,27
Sym2	3	Sym2	3	Sym2	3	1024	0,5	94,67
Sym3	5	Sym3	5	Sym3	5	8192	0,9	92,80
Sym2	3	Sym2	3	Sym2	3	1024	0,9	92,69
Db2	5	Sym3	3	Sym3	3	8192	0,85	92,63
Db2	5	Db2	5	Sym3	3	8192	0,85	91,81
Db2	5	Db2	5	Db2	5	8192	0,85	91,51
Db2	4	Db2	4	Db2	4	1024	0,85	91,20
Sym3	3	Sym3	3	Sym3	3	8192	0,9	90,16
Db2	3	Db2	4	Db2	3	1024	0,85	89,33

Fonte - Autoria Própria.

Tabela 14 - 10 melhores topologias TWP após cruzamento (Disfonia Funcional).

Média de Acertos (%) - 10 Melhores Topologias Disfonia Funcional								
Energia		Logaritmo da Energia		Entropia de Shannon		Janela	Overlap	
Família	Nível	Família	Nível	Família	Nível			
Sym7	4	Sym8	4	Sym7	4	4096	0,85	89,87
Sym5	4	Sym5	4	Sym5	4	1024	0,5	89,78
Sym8	4	Sym8	4	Db9	3	4096	0,85	88,43
Sym6	4	Sym6	4	Sym6	4	4096	0,9	88,16
Db9	3	Sym8	4	Sym8	4	4096	0,85	87,82
Db9	3	Db9	3	Db9	3	4096	0,75	87,57
Sym8	4	Sym7	4	Db9	3	4096	0,85	87,39
Db9	3	Sym7	4	Db9	3	4096	0,85	87,22
Sym8	4	Sym7	4	Sym8	4	4096	0,85	86,08
Sym7	4	Sym7	4	Sym7	4	4096	0,85	85,89

Fonte - Autoria Própria.

Em seguida, foram realizados testes reduzindo os dados, conforme descrito anteriormente, em que a melhor topologia encontrada no processo anterior tem os dados de entrada reduzidos, onde a energia, logaritmo da energia e entropia de shannon são removidos de todas as maneiras possíveis, conforme apresentado nas Tabelas 15,16 e 17, as quais mostram ainda as taxas médias de acertos obtidas, para cada uma das RNAs especialistas.

Tabela 15 - Redução da melhor topologia TWP (Disfonia Funcional).

Média de Acertos(%) - Disfonia Funcional	
[Energia Shannon]	88,93
[Energia Logaritmo da Energia]	86,89
[Logaritmo da Energia Shannon]	86,85
[Shannon]	85,49
[Energia]	84,54
[Logaritmo da Energia]	83,01

Fonte - Autoria Própria.

Tabela 16 - Redução da melhor topologia TWP (Saudável).

Média de Acertos(%) - Saudável	
[Energia]	97,24
[Logaritmo da Energia]	95,88
[Energia Shannon]	94,22
[Shannon]	93,59
[Energia Logaritmo da Energia]	93,23
[Logaritmo da Energia Shannon]	92,21

Fonte - Autoria Própria.

Tabela 17 - Redução da melhor topologia WPT (Disfonia Orgânica).

Média de Acertos(%) - Disfonia Organica	
[Logaritmo da Energia]	96,29
[Energia Logaritmo da Energia]	95,42
[Energia Shannon]	93,94
[Logaritmo da Energia Shannon]	91,10
[Shannon]	90,77
[Energia]	62,58

Fonte - Autoria Própria.

A taxa média de acertos do procedimento de redução dos treinamentos, aplicado com objetivo de evitar *overfitting* pela superioridade numérica de uma classe em relação a outra, pode ser vista nas Tabelas 18, 19 e 20, para cada uma das RNAs.

Tabela 18 - Teste de diferentes treinamentos para a melhor topologia TWP (Disfonia Orgânica).

Média de Acertos(%) - Disfonia Organica	
Treinamento	Média
1	93,68
2	95,52
3	92,64
4	93,68
5	93,53
6	95,03
7	94,97
8	95,21
9	90,35
10	93,02
11	94,44
12	93,46
13	93,79
14	92,07
15	94,14
16	92,74
17	94,63
18	94,09
19	94,54
20	94,27

Fonte - Aatoria Própria.

Tabela 19 - Teste de diferentes treinamentos para a melhor topologia TWP (Saudável).

Média de Acertos(%) - Saudável	
Treinamento	Média
1	93,10
2	93,41
3	92,21
4	94,34
5	95,04
6	94,80
7	94,75
8	93,17
9	94,42
10	94,87
11	95,32
12	93,27
13	93,27
14	93,84
15	93,11
16	95,63
17	93,24
18	89,69
19	92,63
20	95,23

Fonte - Aatoria Própria.

Tabela 20 - Teste de diferentes treinamentos para a melhor topologia TWP (Disfonia Funcional).

Média de Acertos(%) - Disfonia Funcional	
Treinamento	Média
1	85,06
2	83,04
3	89,76
4	88,19
5	78,09
6	86,93
7	86,91
8	82,63
9	86,84
10	84,51
11	88,75
12	87,54
13	86,10
14	89,62
15	90,35
16	86,52
17	89,62
18	85,63
19	88,52
20	88,88

Fonte - Autoria Própria.

Finalmente a melhor topologia e configuração de treinamento TWP foi selecionada para cada uma das RNAs especialistas, conforme mostra a Tabela 21.

Tabela 21 - Melhor topologia e configuração de treinamento TWP.

Melhor Topologia	
Disfonia Funcional	Treinamento 15 Symlet 7_4_Symlet 8_4_Symlet 7_4_4096_0.85
Disfonia Orgânica	Daubechies 2_3_Daubechies 2_3_Symlet 2_5_4096_0.5
Saudável	Energia Daubechies 2_5_8192_0.85

Fonte - Autoria Própria.

Desta forma a última etapa realizada foi a otimização dos parâmetros internos das RNAs utilizando o PSO. Os resultados são apresentados nas matrizes de confusão dos Quadros 2, 3 e 4.

Quadro 2 - Matriz confusão (Saudável).

	Saudáveis	Disfônicas	Incerteza
Saudáveis	739 100%	0	0
Disfônicas	0	1028 100%	0

Fonte - Autoria Própria.

Quadro 3 - Matriz confusão (Disfonia Orgânica).

	Orgânicos	Funcionais e Saudáveis	Incerteza
Orgânicos	213 98,15%	0	4 1,85%
Funcionais e Saudáveis	0	904 99,89%	1 0,11%

Fonte - Autoria Própria.

Quadro 4 - Matriz confusão (Disfonia Funcional).

	Funcionais	saudáveis e Orgânicos	Incerteza
Funcionais	1305 89,44%	7 0,48%	147 10,08%
Saudáveis e Orgânicos	0	2157 96,08%	88 3,92%

Fonte - Autoria Própria.

Os parâmetros internos obtidos para as RNAs especialistas, pelo sistema otimizado com o PSO, podem ser observados na Tabela 22.

Tabela 22 - Parâmetros internos das RNAs, obtidos pela otimização.

Parâmetros internos das RNAs especialistas				
Característica	Neurônios 1ª camada	Neurônios 2ª camada	Neurônios 3ª camada	Taxa de aprendizagem
Saudáveis	39	12	18	0,1265
Disfonias Funcionais	24	59	26	0,0798
Disfonias Orgânicos	4	12	55	0,2987

Fonte - Autoria Própria.

4.2 Análise dos Resultados

Nesta seção serão apresentadas as discussões dos resultados obtidos neste trabalho.

Deve-se inicialmente comentar sobre fatores relevantes relacionados aos arquivos utilizados neste trabalho. Os arquivos foram coletados durante um longo

período, para diferentes trabalhos, com diferentes objetivos, contendo diferentes informações, sendo gentilmente cedidos pelos pesquisadores citados. Para que a realização deste estudo fosse possível, foi necessário realizar uma criteriosa seleção de quais informações seriam consideradas, levando em conta principalmente a quantidade de arquivos com uma mesma informação já que para o treinamento efetivo de uma RNA, faz-se necessário considerável massa de dados, dessa forma a quantidade de arquivos já reduzida, se tornou ainda mais enxuta. Devido ao fato de serem arquivos dos mais distintos estudos da voz, existe uma grande variabilidade intraclasse (como por exemplo para os arquivos de disfonias orgânicas, contendo PUPV, temos arquivos de pessoas com PUPV direita ou esquerda, com diferentes níveis de paralisia, ou seja, prega vocal paralisada com diferentes posições relativas a posição central das pregas vocais), com poucos arquivos representando algumas variações, o que dificulta o treino de RNAs com capacidade de generalização. Por isso, uma base de dados com abundância de arquivos direcionados ao estudo da voz utilizando RNAs não necessitam de um pré-processamento tão minucioso, onde a adequação das taxas de amostragem, remoção de silêncio, ou até a remoção do DC offset, caso todos os arquivos utilizados forem capturados pelo mesmo equipamento (uma RNA é capaz de identificar que a componente é equivalente em todos os arquivos), não seriam de preocupação do projetista da RNA. Desta forma tem-se no presente trabalho um estudo onde a performance da rede se demonstrou robusta diante a adversidades.

A RNA do tipo ART auxiliou no aproveitamento máximo das discrepâncias intraclasses, de maneira que pelo menos um arquivo de determinada característica estivesse no conjunto de treinamento, atenuando o efeito destas no resultado final obtido.

Para a seleção da melhor topologia TWP foi necessário um longo tempo de processamento, por tal motivo foi feita segmentação dos testes em processos, de modo a realizá-los de maneira paralela, em mais de uma unidade de processamento. A partir dos resultados obtidos podemos verificar que dentre as famílias TWP testadas, apenas as famílias apresentadas na Tabela 23, obtiveram após o cruzamento resultado entre os 10 melhores, o que nos auxilia reduzir a região de busca consideravelmente para trabalhos futuros, onde é possível perceber

que as famílias Coiflets não apresentam bons resultados, enquanto as famílias Daubechies e Symlets se destacam. Exaltando as família Daubechies 2, Symlet 2 e Symlet 8, as quais estão contidas nas entradas dos sistemas especialistas. Pôde-se observar ainda que a quantidade de amostras selecionadas no janelamento, são predominantemente 4096 e 8192, janelas maiores, podendo ser explicados, por representarem períodos de tempo suficientes, onde informações relevantes as características estão contidas.

Tabela 23 - Melhores Famílias TWP obtidas.

Disfonia Orgânica	Daubechies 2	Daubechies 8	Symlet 2	Symlet 8
Saudável	Daubechies 2	Symlet 2	Symlet 3	
Disfonia Funcional	Daubechies 9	Symlet 5	Symlet 6	Symlet 7 Symlet 8

Fonte - Autoria Própria.

Outro fator importante que deve ser ressaltado é que a melhor solução para Disfonia Orgânica com entradas reduzidas possui acurácia superior aos testes anteriores, onde somente o cruzamento foi realizado, contudo ao analisar a acurácia por classe, apresentada na Tabela 24, vemos que é mais vantajoso o caso Inicial.

Tabela 24 - Acurácia por classe disfonia orgânica.

	Acurácia (%)	
	Inicial	Entradas Reduzidas
Orgânicos	94,60	91,11
Não Orgânicos	96,43	98,48

Fonte - Autoria Própria.

A proposta dos cruzamentos das topologias TWP, em relação às informações obtidas pela TWP (energia, logaritmo da energia, entropia de shannon), se mostrou efetiva em relação a apenas uma topologia contendo todas as informações, como pode ser observado na Tabela 21, para os três casos de RNAs especialistas obteve-se uma topologia com cruzamento.

Os diferentes testes de TWP se demonstraram importantes visto que a configuração final escolhida, pôde reduzir a quantidade de entradas ou amostras, o que diminui a complexidade da PMC, obtendo-se ainda resultado com melhor acurácia. Como podemos observar na Tabela 21, para a RNA especialista em amostras saudáveis, somente a energia da topologia TWP selecionada foi suficiente para a generalização da classe. Para o caso das disfonias funcionais um treinamento com quantidade de amostras de pertencer e não pertencer iguais, obteve-se uma melhor acurácia, além de reduzir o tempo de treinamento da rede.

Com os resultados apresentados nos Quadros 2, 3 e 4, é possível mostrar que o presente trabalho obteve acurácia de 100% para saudáveis, 99,55% para disfonias orgânicas e 93,47% para disfonias funcionais, com confiabilidade de 0,98, onde as amostras que não foram classificadas de maneira correta, estiveram em sua maioria na incerteza dada por esta confiabilidade, e não no erro de fato.

A Tabela 25 mostra o ganho percentual por procedimento realizado, mostrando a eficiência de cada um em relação ao seu melhor precedente.

Tabela 25 - Acurácia, e ganho percentual por etapa.

	Acurácia (%)				
	Topologia Inicial	Cruzamento	Entrada Reduzida	Amostras Reduzidas	Otimização
Disfonia Funcional	89,78	89,87 (+0,09 %)	88,93 (-0,85 %)	90,35 (+0,48 %)	93,47 (+3,12 %)
Saudável	94,67	96,27 (+1,6%)	97,24 (+0,97%)	95,63 (-1,61 %)	100 (+2,76 %)
Disfonia Orgânica	94,92	95,79 (+0,87 %)	96,29 (+0,5 %)	95,52 (-0,77 %)	99,55 (+3,26 %)

Fonte - Autoria Própria.

Com as informações apresentadas na Tabela 25 podemos verificar a importância de cada uma das etapas desde a obtenção da primeira melhor topologia TWP. Comprovando a eficiência dos procedimentos propostos, na busca de uma solução ótima.

5 CONCLUSÃO

O presente trabalho propôs-se a desenvolver um sistema automático capaz de reconhecer as características (disfonias funcionais, disfonias orgânicas, bem como saudáveis), utilizando RNAs do tipo PMC, dentre um conjunto de arquivos de áudios, cedidos por Fabiana Zambon e pelo GPEMCNPq. Para isso fez-se necessário atingir uma série de objetivos específicos, que são: a aplicação de um conjunto de técnicas de pré-processamento dos arquivos para a aplicação a rede; aplicação da RNA do tipo ART, para separação de arquivos em treinamento e teste, garantindo maior diversidade possível no treinamento; aplicação da TWP para extração de características, onde foram testadas um conjunto de configurações as chamadas topologias, variando-se a família TWP, o nível de decomposição do sinal, o tamanho da janela em que a TWP foi aplicada e a sobreposição entre as janelas; aplicação do cruzamento entre topologias, para a obtenção de entradas que culminam em melhor acurácia a RNA. Além dos procedimentos descritos, os parâmetros internos da mesma é otimizada utilizando a ferramenta proposta em Yoshizumi, et al (2018).

A partir destes procedimentos pôde-se obter acurácias para cada uma das características (saudável: 100%, disfonia funcional: 93,47% e disfonia orgânica: 99,55%) que demonstram a capacidade das ferramentas utilizadas com alta confiabilidade (98%) na classificação de características em vozes disfônicas. Dessa forma pode-se garantir o sucesso de todos os objetivos deste trabalho, obtendo-se como resultado um sistema automático capaz de reconhecer padrões de vozes saudáveis, com disfonias funcionais e disfonias orgânicas.

Com os objetivos principal e específicos alcançados um conjunto de contribuições ao estado da arte puderam ser notados, como a utilização de técnicas de agrupamento, mais especificamente a RNA do tipo ART, que garantiu boa diversidade no treinamento da PMC, mesmo com a quantidade de dados limitada; a observação das famílias TPW com melhor resultado para a gama de dados utilizadas, onde a família *Coiflet* não aparece dentre as testadas, como um dos melhores após o cruzamento e *Daubechies 2*, *Symlet 2* e *Symlet 8* se mostraram mais genéricas, ou seja, possuem melhores resultados sem dependência com

determinada característica; e por fim os cruzamentos de configurações de TWP, que aumentaram consideravelmente a acurácia da PMC (1,6% para os saudáveis, 0,87% para as disfonia orgânicas e 0,09% para as difonias funcionais), para todos os casos testados. Todos estes fatores não observados na literatura, propostos e testados neste trabalho.

Para trabalhos futuros espera-se a busca da melhor topologia TWP a partir de um algoritmo evolutivo ou de busca local, com universo de busca mais restrito, tomando como base o presente trabalho, inclusão das disfonias Organofuncionais, a partir de uma base de dados com maior volume e o desenvolvimento de algoritmo evolutivo que busque a melhor topologia TWP, junto aos melhores parâmetros internos da PMC, integrados em um único sistema.

REFERÊNCIAS

- AGULHARI, C. M. **Compressão de Eletrocardiogramas Usando Wavelets**. Dissertação de Mestrado em Engenharia Elétrica - Faculdade de Engenharia Elétrica e de Computação, Universidade de Campinas, Campinas, SP, 2009.
- AL-TIMEMY, A. H., KHUSHABA, R. N., BUGMANN, G., ESCUDERO, J. **Improving the performance against force variation of emg controlled multifunctional upper-limb prostheses for transradial amputees**. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2015 .
- BARIZÃO, A. H. **Estimação do Grau de Parâmetros Subjetivos Vocais Aplicando Redes Neurais Artificiais**. Trabalho de Conclusão de Curso em Engenharia Elétrica - Universidade Tecnológica Federal do Paraná, Cornélio Procópio, PR, 2017.
- BARROS, F. K. H., LIMA, A. A. M., GUIDOTTI, F. A.,S., YOSHIKUMI, V. H., ARAUJO, R. M., DAJER, M. E., SPATTI D. H. **Treinamento Otimizado de Redes Neurais Artificiais Para a Classificação de Movimentos Musculares**. Encontro Nacional de Inteligência Artificial e Computacional, Uberlândia, MG, 2017.
- BEHLAU, M. **A voz: O livro do especialista. Vol. I**. Rio de Janeiro, RJ: Revinter, 2004.
- DUDA, R. O., HART, P. E., STORK, D. G. **Pattern Classification**. 2ª Ed. New Jersey, US: John Wiley & Sons, 2012.
- FERMINO, M. A. et al **Classificação de Distúrbios Vocais Aplicando Redes Neurais Artificiais e Transformada Wavelet Packet**. 12th IEEE/IAS International Conference on Industry Applications – INDUSCON. Curitiba, PR, 2016a.
- FERMINO, M. A. et al **Reconhecimento de Distúrbios Vocais Utilizando Redes Neurais Artificiais**. XXV Congresso Brasileiro de Engenharia Biomédica – CBEB. Foz do Iguaçu, PR, 2016b.
- GONZÁLEZ, I. V. **Videolaringostroboscopia: una técnica para visualizar las cuerdas vocales**. Revista Estudios de fonética experimental, vol. 17: p. 418-432, 2008.

GROSSBERG, S. **Adaptive pattern classification and universal recoding**, Parallel development and coding of neural feature detectors. Biological Cybernetics, vol. 23, p.121-134, 1976.

HALLIDAY & RESNICK. **Fundamentos de Física**. 10^a ed. vol.2. Editora LTC, 2016.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2^a. ed. Hamilton, Ontario, Canadá. Bookman, 2001.

HAYKIN, S. VEEN, B. V. **Sinais e Sistemas**. Porto Alegre: Bookman, 2001.

LIMA, A. A. M. et al. **Classification of Hand Movements from EMG Signals using Optimized MLP**. International Joint Conference in Neural Network- IJCNN. Rio de Janeiro - RJ, 2018.

LIMA, R. R. **Classificação de sinais de variabilidade da frequência cardíaca por meio da análise de componentes principais e redes neurais**. Trabalho de Conclusão de Curso em Engenharia Elétrica - Faculdade de Tecnologia, Universidade de Brasília, DF, 2012.

LU, G., HANKINSON, T. **A technique towards automatic audio classification and retrieval**. 4th International Conference on Signal Processing Proceedings, Beijing, China. 1998.

MARTINS, A. A. et al. **Reconhecimento de padrões de movimentos da mão a partir de sinais mioelétricos do antebraço utilizando redes neurais artificiais e algoritmos genéticos**. V Congresso Brasileiro de Eletromiografia e Cinesiologia e X Simpósio de Engenharia Biomédica. Uberlândia, MG, 2018.

MCCULLOCH W. & PITTS W. **A Logical Calculus of the Ideas Immanent in Nervous Activity**, Bulletin of Mathematical Biophysics, 1943.

OPPENHEIM, A. V., SCHAFER, R. W.; BUCK, J. R. **Discrete-Time Signal Processing**. 2^a ed. New Jersey: Prentice Hall, 1998.

PARRAGA, A. **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. Dissertação de Mestrado em Engenharia Elétrica, Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2002.

PAVONI, H. E. et al **Identificação de Parâmetros Subjetivos da Voz Utilizando Redes Neurais Artificiais**. 12th IEEE/IAS International Conference on Industry Applications – INDUSCON. Curitiba, PR, 2016.

PIMENTA, R. A. **Uso da Avaliação Multidimensional da Voz na Caracterização Vocal de Pacientes com Paralisia Unilateral de Pregas Vocais**. Tese de Doutorado em Ciências - Programa de Pós-Graduação Interunidades Bioengenharia, Escola de Engenharia de São Carlos, São Carlos, SP, 2016.

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

SATALOFF, R. T. **Voice Science**. Plural Publishing Inc, San Diego, 2005.

SCALASSARA, P. R. **Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias da Laringe**. Tese de Doutorado em engenharia Elétrica - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 2009.

SILVA G. M. A., et al. **Teoria wavelet na análise de sinais de ECG como ferramenta de auxílio na detecção de arritmias cardíacas**. XII Simpósio Brasileiro de Automação Inteligente (SBAI) Natal – RN, 2015

SILVA, I. da; SPATTI, D. H., FLAUZINO, R. A. **Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas**. São Paulo: Artliber Editora Ltda., 2010.

SIVANANDAM, S. N., SUMATHI, S., DEEPA, S. N. **Introduction to Neural Networks Using MATLAB 6.0**. New Delhi: Tata McGraw-Hill Publishing Company Limited, 2006.

SOUZA, R., FREITAS, S. V., FERREIRA, A. **Acoustic evaluation of voice in the practice of the portuguese speech therapists**. Revista Portuguesa de otorrinolaringologia e cirurgia cérvico-facial, vol. 49, p. 11-18, 2011.

SODRÉ, B. R. **Reconhecimento de Padrões Aplicados à Identificação de Patologias de Laringe**. Mestrado em Engenharia e Informática Industrial - Universidade Tecnológica Federal do Paraná, Curitiba, PR, 2016.

STERNBERG, R. J., CONWAY, B. E., KETRON, J. L., BERNSTEIN, M. **People's conceptions of intelligence**. Journal of Personality and Social Psychology, vol.41, p. 37-55,1981.

TSUJI, D. H., et al. **Improvement of Vocal Pathologies Diagnosis Using High-Speed Videolaryngoscopy**. International Archives of Otorhinolaryngology, vol. 18, p. 294-302, Rio de Janeiro, RJ, 2014.

YOSHIZUMI, V. H. et al. **Sistema de otimização de parâmetros internos de redes neurais artificiais**. Seminário de Iniciação Científica e Tecnológica da UTFPR, Apucarana, 2018.

WEEKS, M.. **Digital Signal Processing Using MATLAB & Wavelets**. Massachusetts, US: Jones & Bartlett Learning, 2010.

ZAMBON, F. C. **Estratégias de enfrentamento em professores com queixa de voz**. 2011. 13-19 p. Dissertação Mestrado em Ciências - Escola Paulista de Medicina, Universidade de São Paulo, São Paulo, SP, 2011.