

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

HIGOR EDUARDO PAVONI

**CLASSIFICAÇÃO DE SINAIS VOCAIS EM PARÂMETROS NÃO
ACÚSTICOS UTILIZANDO REDES NEURAS ARTIFICIAIS**

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO

2017

HIGOR EDUARDO PAVONI

**CLASSIFICAÇÃO DE SINAIS VOCAIS EM PARÂMETROS NÃO
ACÚSTICOS UTILIZANDO REDES NEURAS ARTIFICIAIS**

Proposta de Trabalho de Conclusão de Curso
apresentada como requisito parcial à obtenção
do título de Bacharel em Engenharia Elétrica
pela Universidade Tecnológica Federal do
Paraná - UTFPR.

Orientadora: Profa. Dra. María Eugenia Dajer
Coorientador: Prof. Dr. Danilo Hernane Spatti

CORNÉLIO PROCÓPIO

2017



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



FOLHA DE APROVAÇÃO

Higor Eduardo Pavoni

Classificação de sinais vocais em parâmetros não acústicos utilizando redes neurais artificiais

Trabalho de conclusão de curso apresentado às 8:20hs do dia 22/06/2017 como requisito parcial para a obtenção do título de Engenheiro Eletricista no programa de Graduação em Engenharia Elétrica 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). Bruno Catarino Bispo - (Membro)

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

Dedico este trabalho à minha família e à minha
namorada, pelos momentos de ausência
necessários para a elaboração do mesmo.

AGRADECIMENTOS

Agradeço a Deus por ter me dado saúde e força para superar todas as dificuldades.

Agradeço a minha mãe Cássia, meu pai José e ao meu irmão Fernando por todo o incentivo e compreensão durante esta jornada.

Agradeço também a minha fiel companheira Bianca, que utilizou sabiamente toda sua paciência para me acalmar e incentivar em todos os meus momentos de dificuldade.

Agradeço meus orientadores, Profa. Dra. María Eugenia Dajer e o Prof. Dr. Danilo Hernani Spatti, pela sabedoria com que me guiaram por esta trajetória e pelo carinho que tiveram durante todas as explicações e conversas.

Agradeço a meu grande amigo Murillo por trabalhar comigo durante toda a graduação e principalmente durante a elaboração deste TCC.

Agradeço meus demais amigos, colegas de graduação, professores e familiares, que de alguma forma contribuíram com esta conquista. A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado!

**O sucesso consiste em ir de fracasso em
fracasso sem perder o entusiasmo.**

Winston Churchill

RESUMO

PAVONI, Higor Eduardo. **Classificação de sinais vocais em parâmetros não acústicos utilizando redes neurais artificiais**. 2017. 72 f. Trabalho de Conclusão de Curso – Engenharia Elétrica – Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2017.

Este trabalho propõe um método de auxílio para profissionais da voz, em âmbitos clínico, científico e pedagógico, uma vez que a classificação de amostras vocais em parâmetros subjetivos é de extrema importância para diagnosticar patologias e muito utilizada no cotidiano destes profissionais. Utilizando-se amostras previamente gravadas, a voz do paciente é analisada computacionalmente por meio de parâmetros acústicos – energia e entropia logarítmica – extraídos dos coeficientes do terceiro nível da Transformada Wavelet Packet. Posteriormente estas amostras são classificadas por um conjunto de redes neurais artificiais em rugosidade, soprosidade ou tensão, que são parâmetros subjetivos da voz. Para obtenção de maior taxa de acerto, o trabalho propõe a utilização de seis redes neurais artificiais, sendo cada uma especialista na identificação de um dos parâmetros subjetivos, divididos também em feminino e masculino. Os resultados experimentais demonstram que a metodologia proposta pode generalizar adequadamente amostras de voz com percentual médio de acerto de 96,33%.

Palavras-chave: Classificação Vocal. Redes Neurais Artificiais. Transformada Wavelet Packet. Parâmetros Subjetivos da Voz. Parâmetros Acústicos.

ABSTRACT

PAVONI, Higor Eduardo. **Classification of speech signs in non-acoustic parameters using artificial neural networks**. 2017. 72 f. Trabalho de Conclusão de Curso – Engenharia Elétrica – Federal Technology University – Parana. Cornélio Procópio, 2017.

This work proposes an aid method for speech professionals, in clinical, scientific and pedagogical environments, since the classification of vocal samples into subjective parameters is extremely important for the diagnosis of pathologies and much used in the daily life of these professionals. Using previously recorded samples, the patient's voice is analyzed by means of acoustic parameters - energy and logarithmic entropy - extracted from the third level coefficients of the Wavelet Packet Transform. Subsequently these samples are classified by a set of artificial neural networks in roughness, breathiness or strain, all subjective parameters of the voice. In order to obtain a higher rate of accuracy, the study proposes the use of six artificial neural networks, each one specialized in the identification of one of the subjective parameters, also divided into female and male. The experimental results demonstrate that the proposed methodology can adequately generalize speech samples with a mean percentage of correctness of 96.33%.

Keywords: Speech Classification. Artificial Neural Network. Wavelet Packet Transform. Subjective Speech. Parameters. Acoustic Parameters.

LISTA DE FIGURAS

Figura 1 – Representação esquemática do trato vocal humano.....	16
Figura 2 – Exemplos de funções wavelet. a) Haar. b) Daubechies “db2”. c) Symlets “sym4”. d) Ricker ou Chapéu Mexicano.....	22
Figura 3 – Exemplo de decomposição em aproximação e detalhe da transformada wavelet. .	23
Figura 4 – Decomposição terceiro nível árvore wavelet.	23
Figura 5 – Árvore de decomposição transformada wavelet packet, contendo os coeficientes de aproximação (A) e detalhe (D).	24
Figura 6 – Representação neurônio artificial.....	27
Figura 7 – Representação PMC com duas camadas ocultas.....	28
Figura 8 – Fluxograma contendo os procedimentos adotados.	30
Figura 9 – Exemplo de segmentação sobreposta.....	32
Figura 10 – Procedimento para extração de característica acústica de um segmento de áudio.	36
Figura 11 – Algoritmo utilizado durante o procedimento de treinamento das RNAs.....	43
Figura 12 – Exemplo de aplicação para o sistema RNA proposto.	44
Figura 13 – Coeficientes acústicos utilizados durante a etapa de treinamento de cada uma das RNAs propostas.....	46
Figura 14 – Percentual de eficiência para cada uma das topologias testadas para a rede F-R [N1 = 7; N2 = 17]	47
Figura 15 – Percentual de eficiência para cada uma das topologias testadas para a rede F-S [N1 = 5; N2 = 3]	47
Figura 16 – Percentual de eficiência para cada uma das topologias testadas para a rede F-T [N1 = 9; N2 = 5]	48
Figura 17 – Percentual de eficiência para cada uma das topologias testadas para a rede M-R [N1 = 3; N2 = 13]	48
Figura 18 – Percentual de eficiência para cada uma das topologias testadas para a rede M-S [N1 = 7; N2 = 15]	49
Figura 19 – Percentual de eficiência para cada uma das topologias testadas para a rede M-T [N1 = 3; N2 = 15]	49
Figura 20 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede F-R.	51
Figura 21 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede F-S.....	52
Figura 22 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede F-T.	52
Figura 23 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede M-R.....	53
Figura 24 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede M-S.	53
Figura 25 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede M-T.	54

Figura 26 – Histograma de desvio percentual para a rede F-R. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.	57
Figura 27 – Histograma de desvio percentual para a rede F-S. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.	57
Figura 28 – Histograma de desvio percentual para a rede F-T. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.	58
Figura 29 – Histograma de desvio percentual para a rede M-R. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.	58
Figura 30 – Histograma de desvio percentual para a rede M-S. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.	59
Figura 31 – Histograma de desvio percentual para a rede M-T. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.	59

LISTA DE TABELAS

Tabela 1 – Relação Sistema / Função / Estrutura da produção vocal.....	16
Tabela 2 – Exemplos de parâmetros acústicos utilizados por profissionais da voz	19
Tabela 3 – Algumas Famílias Wavelets	21
Tabela 4 – Distribuição dos arquivos que compõem o banco de dados de acordo com suas características.....	31
Tabela 5 – Características dos arquivos do banco de dados – Predominância: Rugosidade....	33
Tabela 6 – Características dos arquivos do banco de dados – Predominância: Soprosidade...	34
Tabela 7 – Características dos arquivos do banco de dados – Predominância: Tensão	34
Tabela 8 – Distribuição dos segmentos formados de acordo com a característica subjetiva dos arquivos de voz.....	35
Tabela 9 – Grau de correlação obtido entre os parâmetros acústicos extraídos dos arquivos femininos e a condição subjetiva Rugosidade.....	37
Tabela 10 – Grau de correlação obtido entre os parâmetros acústicos extraídos dos arquivos femininos e a condição subjetiva Soprosidade.....	38
Tabela 11 – Grau de correlação obtido entre os parâmetros acústicos extraídos dos arquivos femininos e a condição subjetiva Tensão.....	39
Tabela 12 – Grau de correlação obtido entre os parâmetros acústicos extraídos dos arquivos masculinos e a condição subjetiva Rugosidade.....	39
Tabela 13 – Grau de correlação obtido entre os parâmetros acústicos extraídos dos arquivos masculinos e a condição subjetiva Soprosidade.....	40
Tabela 14 – Grau de correlação obtido entre os parâmetros acústicos extraídos dos arquivos masculinos e a condição subjetiva Tensão.....	41
Tabela 15 – Percentual de acerto da RNA mais apta à resolução do problema estudado e quantidade de neurônios pertencentes as camadas escondidas das mesmas.....	46
Tabela 16 – Percentual de Indefinição para cada RNA proposta	51
Tabela 17 – Comparativo entre procedimentos e resultados do artigo e trabalho de conclusão de curso.....	54
Tabela 17 – Erro percentual médio para cada RNA proposta	56
Tabela 18 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio femininos – Etapa 01.....	60
Tabela 19 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio masculinos – Etapa 01.....	61
Tabela 20 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio femininos – Etapa 02.....	61
Tabela 21 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio masculinos – Etapa 02.....	62
Tabela 22 – Percentual de confiabilidade geral para as 6 RNAs mais aptas à resolução do problema proposto	63

SUMÁRIO

1	INTRODUÇÃO	10
2	OBJETIVOS	14
2.1	OBJETIVO GERAL.....	14
2.2	OBJETIVOS ESPECÍFICOS	14
3	REVISÃO BIBLIOGRÁFICA.....	15
3.1	VOZ.....	15
3.1.1	O processo de produção da voz.....	15
3.1.2	Avaliação perceptivo-auditiva da voz	17
3.2	PROCESSAMENTO ACÚSTICO DA VOZ	18
3.2.1	Parâmetros acústicos do sinal de voz	19
3.3	TRANSFORMADA WAVELET	20
3.3.1	Decomposição wavelet para sinais discretos.....	22
3.3.2	Transformada wavelet packet.....	24
3.4	SISTEMAS INTELIGENTES	25
3.4.1	Redes neurais artificiais.....	26
3.4.1.1	Estrutura básica: o neurônio artificial	26
3.4.1.2	Redes Perceptron Multicamadas	27
3.4.1.3	Treinamento e generalização da PMC.....	29
4	MATERIAIS UTILIZADOS E PROCEDIMENTOS ADOTADOS.....	30
4.1	BANCO DE DADOS	30
4.2	VERIFICAÇÃO DAS CONDIÇÕES DO BANCO DE DADOS	31
4.3	SEGMENTAÇÃO DOS ARQUIVOS DE ÁUDIO	32
4.4	EXTRAÇÃO DE CARACTERÍSTICA ACÚSTICA.....	35
4.5	CORRELAÇÃO ENTRE OS COEFICIENTES EXTRAÍDOS E A PREDOMINÂNCIA SUBJETIVA DOS ARQUIVOS DE ÁUDIO	37
4.6	FORMAÇÃO DO CONJUNTO DE TREINAMENTO E TESTE.....	42
4.7	OBTENÇÃO DA TOPOLOGIA RNA MAIS APTA.....	42
5	RESULTADOS OBTIDOS.....	45
5.1	COEFICIENTES DA TWP UTILIZADOS	45
5.2	EFICIÊNCIA DAS TOPOLOGIAS TESTADAS	46
5.3	RESPOSTA AO PROCEDIMENTO DE VALIDAÇÃO.....	50
5.4	TAXA DE ACERTO E COMPARATIVO COM RESULTADOS DO ARTIGO PUBLICADO	54
5.5	DESVIO PERCENTUAL DO PROCEDIMENTO DE TESTES DAS RNAS	55
5.6	MULTIPLICIDADE DE CLASSIFICAÇÃO DOS SEGMENTOS	60
5.6.1	Taxa de confiabilidade geral de classificação das RNAs.....	62
6	CONSIDERAÇÕES FINAIS.....	64
	REFERÊNCIAS	66

1 INTRODUÇÃO

A comunicação é um fenômeno fundamentalmente interativo e a voz figura entre os modos mais utilizados de comunicação entre as pessoas. É por meio dela que expressões e informações são transmitidas a outras pessoas, necessitando de ordem e clareza para o devido entendimento (SCOTT-PHILLIPS, 2010). Problemas vocais suprimem a qualidade comunicativa dos indivíduos, influenciando diretamente em sua relação interpessoal e trazendo restrições nos campos emocionais, sociais e psicológicos (HOGIKYAN e SETHURAMAN, 1999, p.559), impedindo a esperada interatividade entre os indivíduos, muitas vezes por vergonha ou receio de repreensão.

Além das questões sociais, problemas vocais também afetam diretamente o setor financeiro. Segundo o Comitê Brasileiro Multidisciplinar da Voz Ocupacional, cerca de 33% dos trabalhadores brasileiros necessitam da voz para o exercício de suas profissões (BRASIL, 2010) e o afastamento desses profissionais, em decorrência de problemas vocais, traz transtornos para empresas privadas e setores públicos uma vez que se faz necessário realocar funcionários e custear tratamentos.

De acordo com profissionais da clínica vocal, problemas vocais muitas vezes são oriundos de hábitos prejudiciais – como tabagismo, etilismo, hidratação insuficiente, sedentarismo e automedicação – e/ou condições ambientais desfavoráveis em que os indivíduos estão inseridos (ARAÚJO, 2014, p.14). Tomando como exemplo um professor, observa-se que estes profissionais utilizam a voz de modo prolongado e em níveis elevados de intensidade, sofrem com salas de aula com grande quantidade de estudantes, estruturas precárias, ambientes ruidosos, além de todo o estresse associado à profissão (GUIMARÃES, 2004, p.35). Todas estas condições severas refletem diretamente na saúde vocal, sendo a classe de professores a que mais possui profissionais afastados de suas atribuições profissionais em decorrência de problemas vocais (GUIMARÃES, 2004, p.33).

Sintomas como cansaço vocal, garganta seca e rouquidão são comumente considerados indícios de problemas vocais (MUSIAL et al., 2011, p.337). Ao perceber tais alterações, os indivíduos devem procurar um profissional capacitado para realizar uma avaliação detalhada de todo seu trato vocal, sendo tais avaliações de caráter invasivo, perceptivo-auditivo ou acústico.

Avaliações invasivas são aquelas que necessitam da introdução de instrumentos específicos nas vias orais dos pacientes para que o avaliador visualize o trato vocal do mesmo, causando dor e grande desconforto (WANG et al, 2011, p.38). A avaliação perceptiva-auditiva

depende exclusivamente do conhecimento e das percepções do profissional responsável pelo exame para identificar parâmetros subjetivos atrelados às condições anormais da voz. De modo simplificado, o avaliador escuta a voz do paciente e a partir de suas percepções e experiência identifica disfunções e patologias (SANTOS, 2015, p.8). Por fim, a avaliação acústica é aquela em que o sinal de voz do paciente é processado computacionalmente, retornando parâmetros acústicos que devem ser previamente estudados e mapeados, auxiliando o avaliador em seu diagnóstico.

Algumas condições fizeram com que a avaliação perceptiva-auditiva se tornasse a mais empregada nas décadas passadas para diagnóstico de disfunções vocais, uma vez que avaliações invasivas desgastam os pacientes e são procedimentos ainda mais difíceis de serem executados quando realizados em crianças. As avaliações acústicas demandam grande capacidade de processamento computacional, inviabilizando esta modalidade de análise nas décadas passadas (TEIXEIRA et al., 2011, p.3). Por sua vez, a avaliação perceptiva-auditiva traz grande desgaste ao profissional. Além da vasta experiência e habilidade exigida, o profissional deve se preparar para a avaliação “calibrando” seu ouvido antes de iniciar a avaliação propriamente dita. Todo o procedimento chega a durar horas e como a análise realizada é subjetiva, não é possível quantificar um resultado nessas condições, resultando em grandes dificuldades para a padronização e comparação dos resultados (ROSA, 1998, p.1)

Atualmente, técnicas utilizando análise acústica dos sinais de voz na identificação de patologias passaram a ser comumente encontradas em pesquisas, devido ao desenvolvimento tecnológico ocorrido nos últimos anos, com isso, surgiu a possibilidade de construir-se uma relação quantitativa entre os parâmetros acústicos extraídos computacionalmente e patologias vocais (DAJER, 2006, p.32). O interesse de engenheiros e outros profissionais desenvolvedores de tecnologias aplicadas a sistemas biológicos resultou em pesquisas e publicações de trabalhos que apresentaram bons resultados para a análise acústica da voz, conforme estudos de Markaki (MARKAKI e STYLIANOU, 2011) e Arias (ARIAS-LONDONO et al., 2011).

Considerando esta mesma linha de desenvolvimento, este trabalho busca implementar um método capaz de classificar amostras de voz dentro de três parâmetros subjetivos muito utilizados por fonoaudiólogos para a identificação de patologias vocais, sendo eles: rugosidade, sopro e tensão (VILLA-CANAS et al., 2013, p.1). A identificação e posterior análise de tais parâmetros possibilita ao profissional da voz averiguar as condições vocais do paciente, auxiliá-lo na identificação de patologias e acompanhar a evolução do paciente durante tratamentos (VILLA-CANAS et al., 2013, p.1).

Como descrito anteriormente, a análise acústica consiste em procedimentos computacionais capazes de extrair das amostras de áudio coeficientes quantitativos referentes às amostras analisadas. Tais coeficientes podem ser estudados e posteriormente relacionados a condições conhecidas das amostras, tornando-se “sintoma” daquela condição. Análises relacionando parâmetros acústicos com doenças vocais e parâmetros subjetivos – como os três aqui estudados – estão presentes na literatura, apresentando coeficientes tais como: *Jitter* e *Shimmer* (FARRÚS e HERNANDO, 2009) (TEIXEIRA et al., 2011), Energia extraída da transformada *Wavelet* (SANTOS, 2015), Coeficientes Mel-Ceptrais (MFCC), modulação espectral e parâmetros de ruído (VILLA-CANAS et al, 2013).

A extração de coeficientes acústicos abre possibilidade para utilização de classificadores artificiais, suprimindo a necessidade de um especialista para classificar manualmente as amostras. Tomando as semelhanças presentes nos coeficientes, as amostras podem ser agrupadas de acordo com critérios estabelecidos previamente, de acordo com a ferramenta classificadora utilizada. Tradicionalmente, Redes Neurais Artificiais (RNA) são utilizadas como ferramentas classificadoras em múltiplas áreas da ciência, não sendo diferente na bioengenharia, onde são facilmente encontradas em trabalhos científicos, como nos trabalhos de Souissi (SOUISSI e CHERIF, 2016), Holi (HOLI, 2013) e Crovato (CROVATO e SCHUCK, 2007).

Unindo as etapas de extração de características acústicas das amostras de áudio e classificação das mesmas em grupos de acordo com suas semelhanças e diferenças, pode-se obter uma ferramenta suficientemente robusta para auxílio de profissionais da voz em suas avaliações. Neste trabalho utilizou-se como classificador uma RNA *Perceptron* Multicamadas (PMC) por se tratar de uma arquitetura robusta, empregada amplamente em diversas áreas do conhecimento, lidando com os mais variados problemas. Após procedimentos práticos realizados durante o período de elaboração deste trabalho, concluiu-se que o extrator de características mais apto, dentre os investigados, para resolução da proposta é energia e entropia logarítmica extraídos dos coeficientes da Transformada *Wavelet Packet* (TWP).

O banco de dados utilizado neste trabalho foi gentilmente cedido pelo Centro de Estudos da Voz (CEV), contendo apenas arquivos com a vogal sustentada /e/, com predominância rugosa, ou sopro ou tensa, gravados em ambiente controlado e supervisionado por profissionais da área da fonoaudiologia. Grande parte das avaliações vocais são realizadas com o paciente sustentando as vogais /a/, /e/ ou /i/, uma vez que, com esta ação, o trato vocal comporta-se de modo a evidenciar características marcantes relacionadas às condições vocais do indivíduo (BEHLAU, 2001, p.27).

O desenvolvimento deste trabalho resultou no último ano em uma publicação para 12th IEEE/IAS *International Conference on Industry Applications* – INDUSCON 2016, intitulada “Identificação de parâmetros subjetivos da voz utilizando redes neurais artificiais”, em que os arquivos de áudio analisados foram classificados, utilizando-se TWP e RNA, nos mesmos parâmetros subjetivos aqui estudados. Alguns novos procedimentos foram incluídos neste trabalho a fim de conseguir melhores resultados que aqueles obtidos no artigo publicado. Tais procedimentos e resultados serão melhor explicados e discutidos ao transcorrer deste trabalho.

Tal porcentagem demonstra a robustez e eficiência da ferramenta proposta, podendo ser utilizada não só no auxílio de fonoaudiólogos em suas avaliações, como também podem contribuir com a formação de novos profissionais ou apoiar qualquer outro indivíduo envolvido em processos que necessitem classificar vozes dentro dos três parâmetros subjetivos analisados.

2 OBJETIVOS

2.1 OBJETIVO GERAL

O objetivo deste trabalho é desenvolver uma ferramenta classificadora, utilizando redes neurais artificiais, capaz de identificar a predominância dos parâmetros perceptivo-auditivos de rugosidade, soproidade e tensão em amostras gravadas da vogal sustentada /e/ de um indivíduo, por meio da análise de valores de energia e entropia logarítmica extraídas dos coeficientes do terceiro nível da transformada *Wavelet Packet*.

2.2 OBJETIVOS ESPECÍFICOS

De maneira mais detalhada, os objetivos podem ser descritos como:

- Verificar a integridade do banco de dados, assegurando que o mesmo contenha exclusivamente gravações da vogal /e/ sustentada.
- Segmentar os arquivos que compõem o banco de dados em amostras menores, com a finalidade de tornar o processamento mais eficaz.
- Processar e extrair parâmetros acústicos dos segmentos de áudio.
- Obter por meio de treinamentos da rede neural artificial uma topologia suficientemente robusta para a resolução da proposta.
- Validar a capacidade de classificação da rede neural obtida.

3 REVISÃO BIBLIOGRÁFICA

3.1 VOZ

A voz é a forma principal de comunicação entre as pessoas. Sua formação depende da interação de vários sistemas, resumindo-se a um sinal acústico que provém da passagem do ar pelas pregas vocais e demais estruturas do corpo (DAJER, 2006, p.14). Muitos fazem da voz seu instrumento de trabalho e é também por intermédio dela que as pessoas se relacionam, expressam seus sentimentos, pensamentos e vontades. Pode-se dizer que a voz possui um caráter único para cada indivíduo, podendo ser tão pessoal quanto as impressões digitais (SANTOS, 2015, p.5).

Clinicamente falando, são passíveis de identificação pequenas anomalias na voz decorrentes de inúmeros fatores. Isso se justifica pelo fato de que a laringe (local onde as pregas vocais se encontram) está alocada em uma região do pescoço cercada por vasos sanguíneos, nervos e glândulas. Quaisquer ocorrências nos arredores das pregas vocais influenciam diretamente a geração vocal. Como exemplo, durante a realização de ações como respirar profundamente, esforço para levantar algum peso, deglutir, tossir ou até mesmo em uma situação de afogamento, verifica-se a alteração ou até mesmo o impedimento em emitir voz (TITZE, 2013, p.14).

3.1.1 O processo de produção da voz

O mecanismo responsável pela produção vocal não deve ser classificado como um mecanismo simples. Alguns autores descrevem a produção vocal como um encadeamento composto por quatro fases: respiração, fonação, articulação e ressonância e não consideram importantes os mecanismos de audição e outras vias responsáveis pelo *feedback* dos sons produzidos. Problemas na fonação são facilmente encontrados em portadores de deficiência auditiva e em indivíduos com distúrbios neurológicos, uma vez que a produção vocal está também diretamente relacionada aos impulsos neurais, afetando diretamente a qualidade da voz produzida (TORTORA, 2000, p.48).

Segundo Behlau (2001, p.26), para que seja produzida voz em condições adequadas é necessário o trabalho síncrono entre sistema respiratório, sistema fonatório, sistema de ressonância, sistema articulatorio e sistema nervoso central. Cada um desses sistemas possui fundamental relação com a produção vocal, sendo cada um destes sistemas melhor discriminado na Tabela 1:

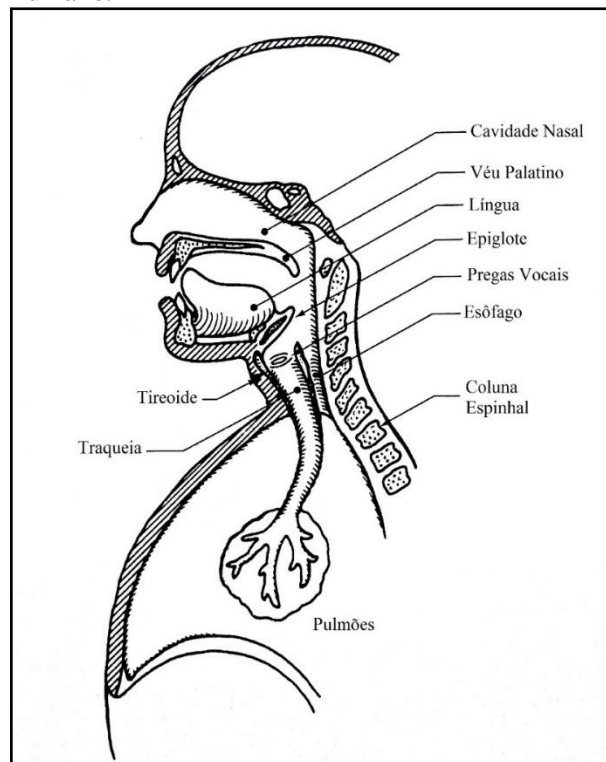
Tabela 1 – Relação Sistema / Função / Estrutura da produção vocal

Sistema	Responsabilidade	Estruturas relacionadas
Respiratório	Fonte de energia	Pulmões
Fonatório	Fonte Vibratória	Pregas Vocais
Ressonância	Fonte de ressonância	Cavidade oral e nasal
Articulatório	Articulação	Língua, lábios, mandíbula, palato e dentes
Nervoso	Coordenação	Córtex Cerebral

Fonte: Adaptado de Behlau (2001, p.26).

Os órgãos e estruturas corporais relacionadas com maior intimidade com a geração da fala são os pulmões, a traqueia, a laringe, as cavidades nasais e a cavidade oral. Estas estruturas tratadas como principais e outras tais como língua, dentes e véu palatino formam um sistema versátil e complexo de produção da voz (RABINER e SCHAFER, 2011, p.68). A Figura 1 exemplifica o sistema em questão.

Figura 1 – Representação esquemática do trato vocal humano.



Fonte: Adaptado de RABINER e SCHAFER (2011, p.71)

3.1.2 Avaliação perceptivo-auditiva da voz

Assim como a voz, patologias vocais são complexas e não obedecem a nenhum parâmetro pré-determinado de fácil observação. Então, para que o profissional identifique sinais de anomalia, se faz necessária uma análise mais ampla e profunda de características audíveis dos pacientes. Tal análise se relaciona diretamente com as condições do especialista responsável pelo diagnóstico, pois a identificação das características primordiais para o reconhecimento de alguma anomalia depende exclusivamente da atuação dos seus sentidos e de toda sua experiência profissional (ROSA, 1998, p.2).

Esta avaliação tem como base o julgamento subjetivo das características presentes na voz, resultando de uma série de comparações internas do avaliador. Tais concordâncias estão sujeitas a diversos fatores e se modificam de profissional para profissional. Este tipo de análise ainda é comum e muito utilizado por fonoaudiólogos e otorrinolaringologistas, sendo também objeto de estudo e aprimoramento, uma vez que, por conta de sua análise subjetiva, sempre se levantam dúvidas quanto ao seu grau de confiabilidade (CARDING et al, 2000, p.134).

A avaliação perceptiva-auditiva, utiliza escalas e protocolos capazes de fornecer parâmetros apropriados para identificação de características vocais. Algumas escalas e protocolos possuem reconhecimento internacional, como o caso das escalas GRBAS e CAPE-V (SIMBERG, 2000, p.235), porém nem todas são mundialmente reconhecidas. Como não há recomendações específicas ou restrições para a utilização de escalas locais (adaptadas às condições de uma comunidade local, como por exemplo uma escala para falantes do Português do Brasil), tais avaliações se distanciam cada vez mais umas das outras, tornando impraticáveis comparações das avaliações realizadas por outros profissionais (DAJER, 2010, p.42).

Os parâmetros subjetivos mais utilizados nas diversas escalas de avaliação perceptivo-auditiva para analisar desvios vocais são: rugosidade, soprosidade e tensão (OATES, 2009, p.50). Uma breve definição de cada um destes parâmetros é abordada adiante (DAJER, 2010, p.43):

- **Rugosidade**: Consiste em uma irregularidade vibratória das pregas vocais durante a fonação, englobando conceitos de rouquidão, crepitação e bitonalidade. Percebe-se a voz com ruídos presentes em baixa frequência, produzindo um som similar ao de uma bandeira tremulando ao vento forte. Uma voz rugosa pode estar relacionada a nódulos vocais, edemas e hiperemias.

- **Soprosidade:** Caracterizado pelo escape de ar provocado pela falta de coaptação glótica, a voz soprosa pode ser identificada pela presença de ruído de fundo audível.
- **Tensão:** Associada ao esforço vocal por aumento da rigidez sobre a musculatura glótica. Em uma avaliação acústica, a voz tensa apresenta ruídos de alta frequência. Sua característica mais perceptiva é uma voz sem projeção e seca. Relacionados a esta característica vocal estão problemas como pontes mucosas, cistos e sulco vocais.

3.2 PROCESSAMENTO ACÚSTICO DA VOZ

Métodos não-invasivos de avaliação da função fonatória são ferramentas auxiliares de fonoaudiólogos e otorrinolaringologistas que propiciam medidas objetivas, isto é, dados quantitativos extraídos utilizando-se processamento computacional adequado, realizado muitas vezes por instrumentos relativamente simples e baratos (PINHO, TSUJI e BOHADANA, 2006, p.33).

Segundo Dajer (2006, p.32), o desenvolvimento tecnológico possibilitou a implementação da análise acústica no dia a dia dos profissionais da voz, permitindo a definição de alguns parâmetros e medidas acústicas, além da possibilidade de relacionar tais parâmetros e medidas com determinadas patologias.

Pinho, Tsuji e Bohadana (2006, p.33) citam algumas motivações para a utilização de análises quantitativas:

- Podem servir de apoio para observações subjetivas, como por exemplo, em imagens videolaringoestroboscópicas durante um diagnóstico.
- São capazes de auxiliar o monitoramento terapêutico, comparando as análises realizadas antes e depois de procedimentos cirúrgicos ou tratamentos medicamentosos.
- A utilização de imagens e gráficos das análises acústicas podem auxiliar a compreensão do paciente a respeito do problema vocal enfrentado.
- Registros de medidas acústicas podem amparar a defesa dos profissionais em possíveis disputas judiciais.

A utilização de avaliações acústicas precisa de alguns cuidados. Estes vão desde a obtenção e tratamento dos sinais até o conhecimento prévio sobre as capacidades, limitações e fragilidades dos algoritmos computacionais envolvidos. (PINHO, TSUJI e BOHADANA, 2006, p.34).

3.2.1 Parâmetros acústicos do sinal de voz

Utilizando-se *softwares* e *hardwares* projetados para extrair parâmetros acústicos de sinais de voz, profissionais conseguem parâmetros objetivos proveniente dos arquivos analisados. O diagnóstico é então amparado pela análise dos coeficientes obtidos e conhecimento específico para classificá-los. De modo geral, é necessário ao profissional da voz conhecer quais parâmetros acústicos – e em quais níveis – estão relacionados com patologias vocais (MENDES et al., 2012, p.422).

Parâmetros acústicos vocais, em sua grande maioria, estão subdivididos em 5 classes conforme apresentado na Tabela 2. Estes parâmetros são utilizados por profissionais da voz não só no âmbito clínico, mas também em áreas pedagógicas e científicas, uma vez que tais avaliações estão cada vez mais presentes no trabalho de tais profissionais. Complementam a Tabela 2 alguns parâmetros acústicos comumente utilizados por profissionais da voz, bem como um breve comentário acerca de cada um deles, evidenciando faixas de normalidade vocal ou algum outro aspecto de interesse.

Tabela 2 – Exemplos de parâmetros acústicos utilizados por profissionais da voz

Classe	Parâmetro Acústico	Comentários diversos
	Extensão máxima de frequência vocal	Intervalo entre a frequência mais grave e a mais aguda que o indivíduo consegue produzir.
Espectrais	Frequência fundamental vocal	Masculino variando de 100 à 150Hz e feminino de 180 à 250Hz.
	Quebras de Frequência	Modificações abruptas da frequência da voz.
	Variabilidade de frequência fundamental	Desvio ocorridos durante a fala do indivíduo.

Tabela 2 – Exemplos de parâmetros acústicos utilizados por profissionais da voz

Classe	Parâmetro Acústico	Comentários diversos
Intensidade	Extensão dinâmica	Variação da intensidade vocal. Do mais suave ao mais forte.
	Índice de fonação suave	Razão entre o grau de energia média entre baixas e altas frequências.
	Variabilidade da amplitude	Desvio padrão da variação de amplitude vocal.
Perturbação	<i>Jitter</i>	Percentual de perturbação da frequência fundamental vocal.
	<i>Shimmer</i>	Percentual de perturbação da intensidade vocal.
Ruído	Razão Sinal-Ruído	Razão entre o nível de voz e ruído de fundo.
	Razão Harmônico-Ruído	Razão entre a amplitude do sinal pela amplitude do ruído.
Temporais	Taxa de elocução	Número de determinadas palavras que o indivíduo consegue falar em determinado período de tempo.
	Diadococinésia	Capacidade de realizar movimentos rápidos e sequenciais: /tátátá por exemplo.
	Índice S/Z	Razão entre o tempo máximo de sustentação das consoantes /S/ e /Z/.
	Tempo máximo de fonação	Tempo máximo de sustentação de uma vogal ou som específico.

Fonte: Adaptado de Mendes (2012, p.423)

3.3 TRANSFORMADA WAVELET

Wavelets são funções com características matemáticas próprias, capazes de representar dados e outras funções que variam no tempo (OLIVEIRA, 2007, p.172). A nomenclatura *wavelet* foi utilizada pela primeira vez na década de 70, em um estudo de Jean Morlet, onde o autor propôs modificações na Transformada de Fourier buscando melhorar os resultados de seus estudos geofísicos (OLIVEIRA, 2007, p.175).

Funções que satisfaçam as propriedades de uma função *wavelet* são denominadas *wavelet*-mãe e podem ser condicionadas (ampliada, reduzida, deslocada) por meio de operações matemáticas de acordo com as necessidades exigidas pelos sinais estudados (CASTELANO, 2006, p.49). Diante deste fato, inúmeras podem ser as funções eleitas como *wavelet*-mãe, existindo também variações pequenas nestas *wavelets*-mães que resultam em famílias de

wavelets (OLIVEIRA, 2007, p.180). Para ilustrar a variedade de famílias *wavelets* apresenta-se algumas delas na Tabela 3 e Figura 2.

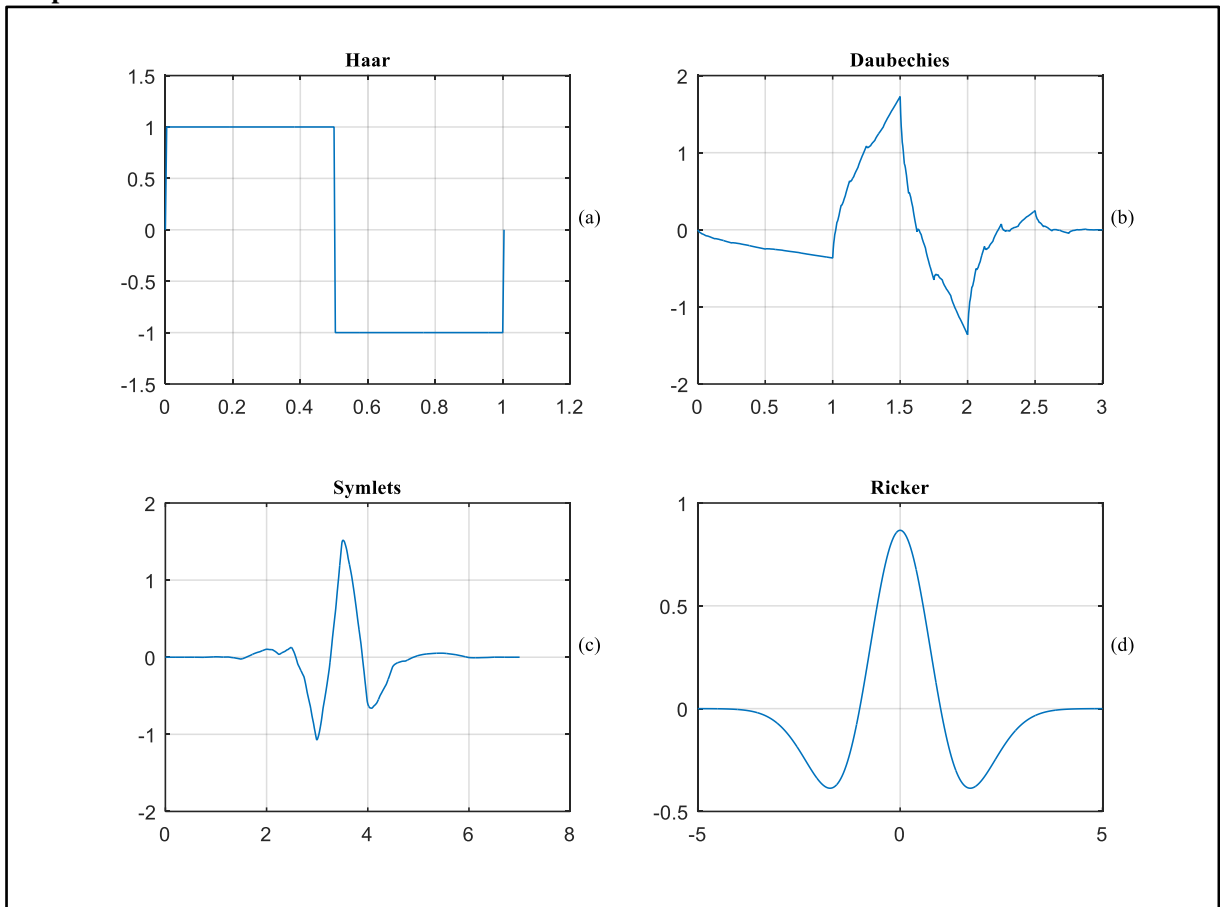
O estudo empregando uma transformada *wavelet* pode ser de caráter contínuo, onde se utiliza a Transformada *Wavelet* Contínua (CWT – *Continuous Wavelet Transform*), ou discreto por meio da Transformada *Wavelet* Discreta (DWT – *Discrete Wavelet Transform*). Em ambos os casos são gerados coeficientes provenientes dos produtos entre a função estudada e a *wavelet*-mãe condicionada (após a *wavelet*-mãe ser ampliada, reduzida, deslocada), sendo estes coeficientes função de um dado parâmetro de escala e posição (CASTELANO, 2006, p.55-57).

Tabela 3 – Algumas Famílias Wavelets

Nome	Abreviação
Haar wavelet	‘haar’
Daubechies wavelets	‘db’
Symlets	‘sym’
Coiflets	‘coif’
Mexican hat wavelet	‘mexh’

Fonte: Adaptado de Oliveira (2007, p.180)

Figura 2 – Exemplos de funções wavelet. a) Haar. b) Daubechies “db2”. c) Symlets “sym4”. d) Ricker ou Chapéu Mexicano.



Fonte: Adaptado de Oliveira (2007, p.181)

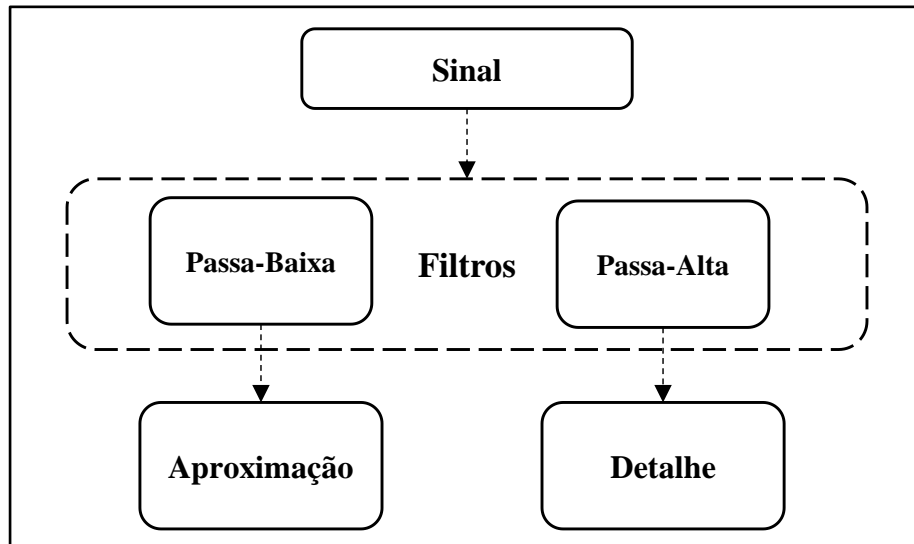
3.3.1 Decomposição *wavelet* para sinais discretos

O processo de decomposição de um sinal discreto utilizando uma transformada *wavelet* é implementado por meio de filtragens digitais, realizadas de maneira semelhante à utilização de filtros passa-baixa e passa-alta, em que metade da banda do sinal original é condicionada para cada um dos lados. A componente de baixa frequência é denominada coeficiente de aproximação (A), enquanto detalhe (D) refere-se à componente de alta frequência, conforme ilustrado na Figura 3 (CASTELANO, 2006, p.50; OLIVEIRA, 2007, p.200).

Para análises mais profundas e detalhadas de um sinal, a decomposição vai ganhando novos níveis de análise. Para isso, aplica-se o conjunto de filtros ao coeficiente de aproximação resultante da última decomposição, gerando dois novos coeficientes. Este procedimento de decomposição dá origem à árvore de decomposição *wavelet*, conforme apresentado na Figura 4 (CASTELANO, 2006, p.50).

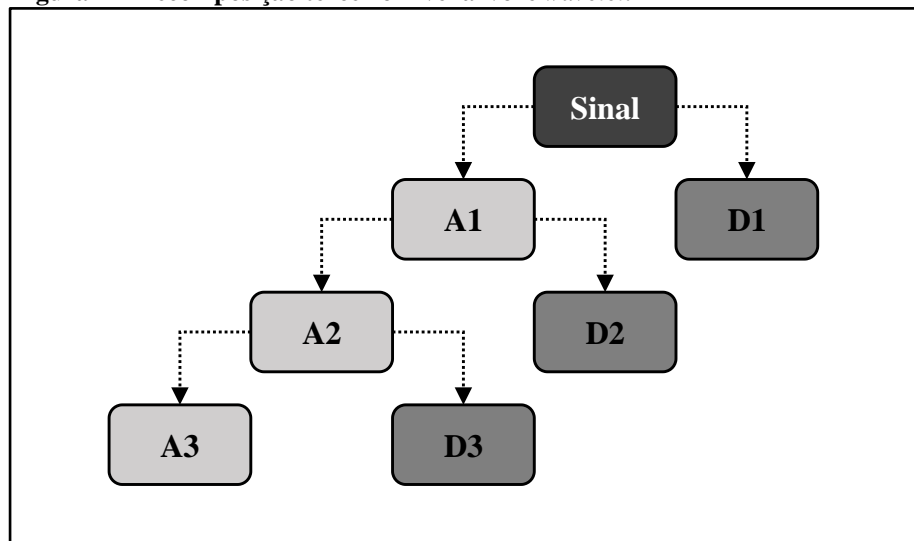
Conforme se avança nos níveis, obtém-se maiores informações (detalhes) sobre a porção de sinal analisado e perde-se o “contexto geral” referente ao sinal total. De modo análogo, pode-se exemplificar este comportamento por meio do manuseio de um mapa eletrônico. Quando o usuário deseja obter uma visualização global do território, ajusta-se o *zoom* até que o território desejado esteja apresentado em sua tela de visualização. Conforme a necessidade de obter informações mais precisas, o usuário ajusta o *zoom* para dar enfoque a apenas uma região daquele território, porém não é possível ter uma definição exata de onde a região em análise se encontra dentro do território (OLIVEIRA, 2007, p.169).

Figura 3 – Exemplo de decomposição em aproximação e detalhe da transformada *wavelet*.



Fonte: Adaptado de Castelano (2006, p.50)

Figura 4 – Decomposição terceiro nível árvore *wavelet*.

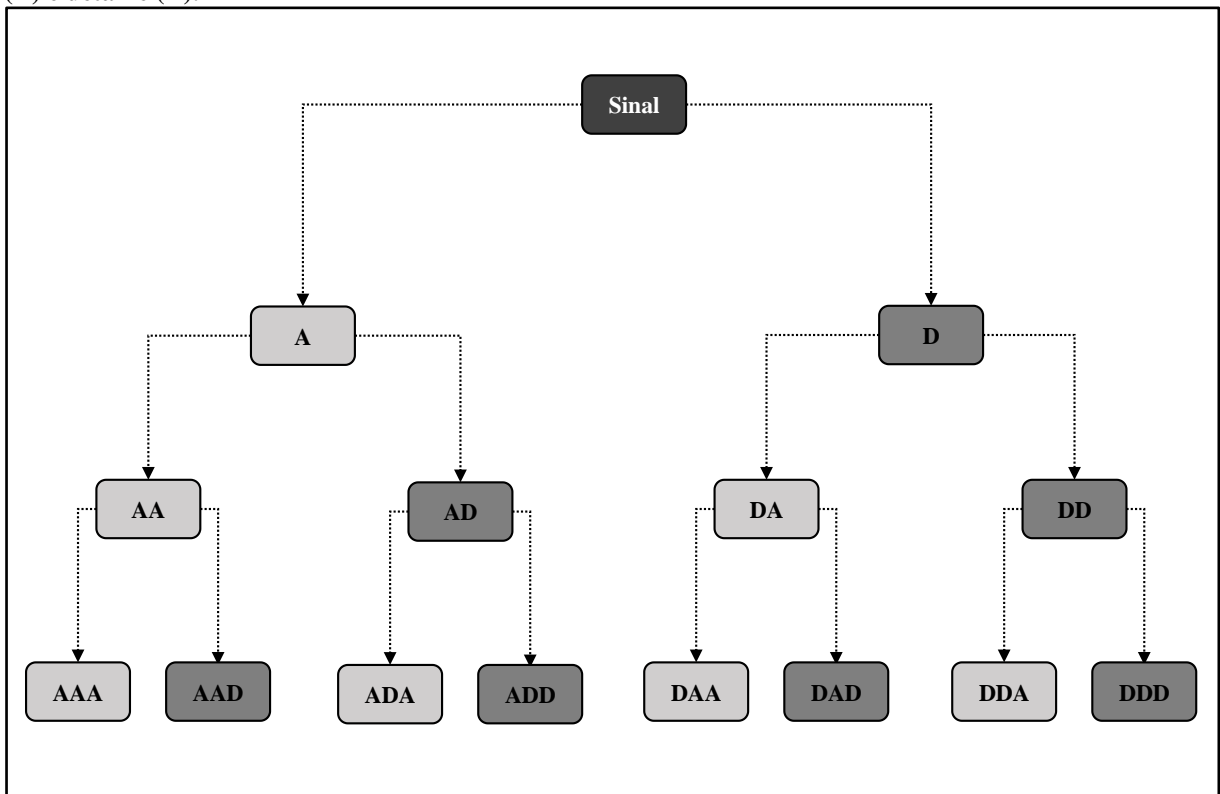


Fonte: Adaptado de Castelano (2006, p.50)

3.3.2 Transformada *wavelet packet*

A transformada *wavelet packet* (TWP) é considerada uma generalização dos conceitos da DWT, diferenciando-se na quantidade de coeficientes gerados após cada decomposição. A utilização desta transformada em decomposições enriquece a análise com detalhes não levados em consideração ao utilizar-se a DWT. A decomposição utilizando-se a TWP não ignora as informações presentes no coeficiente de detalhe ao iniciar um novo nível, ou seja, a partir de uma porção de sinal são gerados dois conjuntos de coeficientes: aproximação e detalhe. O procedimento é realizado como na DTW, porém, ao realizar-se a próxima decomposição para o nível seguinte são gerados mais dois coeficientes para cada um dos coeficientes gerados na decomposição anterior. A Figura 5 apresenta um exemplo de decomposição em 3 níveis utilizando TWP.

Figura 5 – Árvore de decomposição transformada *wavelet packet*, contendo os coeficientes de aproximação (A) e detalhe (D).



Fonte: Adaptado de Castelano (2006, p.59)

3.4 SISTEMAS INTELIGENTES

Sistemas inteligentes estão intimamente relacionados com o processo humano de raciocínio, dedução, aprendizado, percepção, evolução e adaptação. Basicamente um sistema inteligente é aquele que simula o pensamento racional do ser humano e toma decisões próprias (GOLDSCHMIDT, 2010, p.07-08).

Diversas ramificações compõem o vasto cenário de ferramentas inteligentes, podendo-se citar: redes neurais artificiais (RNA), sistemas *fuzzy*, computação evolutiva, inteligência coletiva, sinais imunológicos artificiais e agentes inteligentes, cada qual com suas características e limitações particulares (SILVA, SPATTI, FLUZINO, 2010, p.24).

Com a utilização destas ferramentas exploraram-se áreas desconhecidas e a solução de problemas extremamente complexos tornaram-se viáveis. As possibilidades aumentaram com a inserção dos chamados sistemas híbridos, compostos por duas ou mais ferramentas inteligentes (GOLDSCHMIDT, 2010, p.10).

Considerando as inúmeras possibilidades de aplicações e arranjos destas ferramentas, a escolha de qual utilizar requer análise refinada. O problema em questão necessita classificar os sinais de voz proveniente dos pacientes, de acordo com a predominância subjetiva de cada arquivo de áudio, utilizando-se de coeficientes acústicos extraídos dos mesmos sinais. Segundo Bishop (2016, p.226) e Hornik (1989, p.226), as redes *Perceptron* Multicamadas (PMC) são altamente indicadas para resolução de problemas envolvendo entradas múltiplas de sinais complexos e de difícil análise. Encontram-se também na literatura trabalhos com objetivos semelhantes e bons resultados que utilizaram redes PMC, como em Holi (2013) e Almaadeed (2015).

As redes PMC se destacam das demais arquiteturas de redes neurais pelo fato de serem ferramentas versáteis quanto à sua utilização, estando presentes em diversas aplicações nas mais diferentes áreas do conhecimento. Por possuir processamento paralelo e distribuído por toda a sua estrutura, a rede PMC apresenta resultados robustos e confiáveis, tornando-se um procedimento compacto após o processo de treinamento. A rede também é considerada um aproximador universal de funções, conseguindo assim distinguir padrões presentes nas amostras de entrada, separando-as em classes com desempenho superior a outras ferramentas, como por exemplo as redes *Perceptron* simples e Adaline (SILVA, SPATTI e FLAUZINO, 2010, p.120-121; BISHOP, 2006, p.226; HORNIK, STINCHCOMBE e WHITE, 1989, p.359).

3.4.1 Redes neurais artificiais

As RNAs são encadeamentos de operações de multiplicação e soma, inspiradas em sistemas biológicos que buscam, por intermédio de processamento de informações, solucionar problemas de diversas áreas da engenharia e ciências (BISHOP, 2007, p. 226). Tais redes possuem alta capacidade de generalização após seu período de treinamento, correlacionando dados de entrada com saídas previamente esperadas. Em suma, as RNAs são conhecidas por sua capacidade de aprendizado, adaptação por experiência, tolerância a falhas, facilidade de prototipagem, entre outros. (HAYKIN, 2000, p.24)

Nas subseções seguintes, abordam-se em maiores detalhes alguns aspectos desta ferramenta.

3.4.1.1 Estrutura básica: o neurônio artificial

Um neurônio artificial é uma estrutura computacional baseada em modelos de neurônios existentes no sistema nervoso. Apesar de possuir certa simplicidade, tal estrutura pode ser utilizada para classificar padrões linearmente separáveis (DUDA; HART; STORK, 2012, p.285).

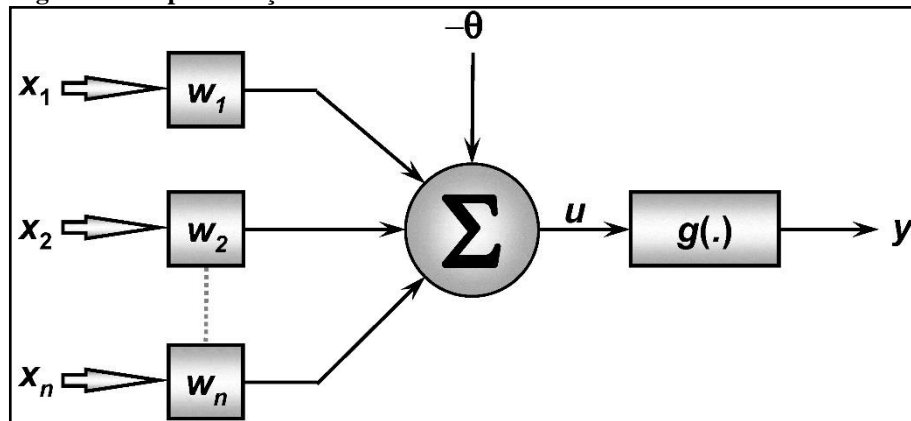
Um neurônio é constituído, segundo Silva, Spatti e Flauzino (2010, p.34-35), por sete elementos básicos listados nos tópicos a seguir e ilustrados na Figura 6:

- **Sinais de entrada (X_n)**: Sinais que representam os valores de variáveis em aplicações específicas. Usualmente são normalizados entre -1 e +1, visando incrementar a eficiência computacional dos algoritmos.
- **Pesos Sinápticos (W_n)**: Valores responsáveis por ponderar os sinais de entrada, permitindo quantificar as relevâncias de cada uma delas.
- **Combinador linear (Σ)**: Agrega todos os sinais de entrada já ponderados pelos pesos sinápticos, produzindo o potencial de ativação.
- **Limiar de Ativação (θ)**: Variável responsável por limitar a ativação do neurônio, pois, para que o neurônio seja ativado (transmitindo a informação

recebida), é necessário que a somatória do sinal de entrada seja maior que o valor do limiar de ativação.

- **Potencial de ativação (u)**: Diferença entre o valor proveniente do combinador linear e o valor do limiar de ativação.
- **Função de ativação (g(.))**: Limita a saída do neurônio dentro de um intervalo de valores.
- **Sinal de saída (y)**: Valor final produzido pelo neurônio em relação à determinados valores de entrada.

Figura 6 – Representação neurônio artificial.



Fonte: Silva, Spatti e Flauzino (2010, p.34)

As funções de ativação desenvolvem um importante papel na análise de saída. De acordo com sua característica, estas funções podem condicionar o valor de saída a saturar rapidamente, como no caso da função logística e tangente hiperbólica. Podem apresentar comportamento discreto, forçando a saída ser 0 ou 1 no caso da função degrau e -1 ou 1 no caso da função degrau bipolar, ou podem resultar uma saída contínua dentro do universo de discurso, como no caso da rampa linear.

3.4.1.2 Redes *Perceptron* Multicamadas

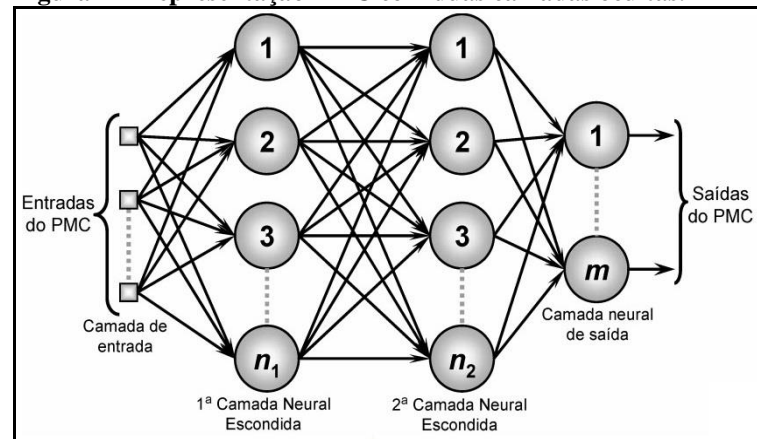
As redes PMC são caracterizadas pelas inúmeras possibilidades de aplicação nos mais diversos tipos de problemas e áreas do conhecimento, sendo considerada uma arquitetura

versátil quanto à sua aplicabilidade. Podem-se citar como potenciais áreas de atuação desta rede problemas que envolvem classificação de padrões, aproximações de funções e sistemas dinâmicos (SILVA; SPATTI; FLAUZINO, 2010, p.120-121).

A rede pertence à arquitetura *feedforward* (para frente), sem qualquer tipo de realimentação, sendo composta por múltiplas camadas e treinamento executado de forma supervisionada (SILVA; SPATTI; FLAUZINO, 2010, p.91-92).

Conforme pode ser visualizado na Figura 7, uma PMC é composta por uma camada de entrada, uma ou mais camadas neurais escondidas e camadas neurais de saída. O número de neurônios presentes em cada camada não obedece a nenhuma regra, podendo ser ajustado de acordo com a modelagem de cada problema.

Figura 7 – Representação PMC com duas camadas ocultas.



Fonte: SILVA; SPATTI; FLAUZINO (2010, p.92)

O funcionamento desta topologia de rede se assemelha ao funcionamento de um único neurônio artificial, dado que a PMC nada mais é do que um grande encadeamento de neurônios artificiais conectados uns aos outros, onde um neurônio de uma determinada camada está conectado a todos os neurônios da camada seguinte. O resultado deste encadeamento possibilita à rede ser tolerante a possíveis falhas, pois seu processamento é distribuído por todo o arranjo neural (HAYKIN, 2000, p.30).

De acordo com Haykin (2000, p.186) e Silva, Spatti e Flauzino (2010, p. 93), o fluxo de sinal em uma PMC progride para frente, da esquerda para direita e indo de camada para camada após o procedimento de treinamento. Os sinais resultantes de uma camada neural serão as entradas para a camada seguinte, mantendo este comportamento até que o sinal alcance a camada neural de saída.

3.4.1.3 Treinamento e generalização da PMC

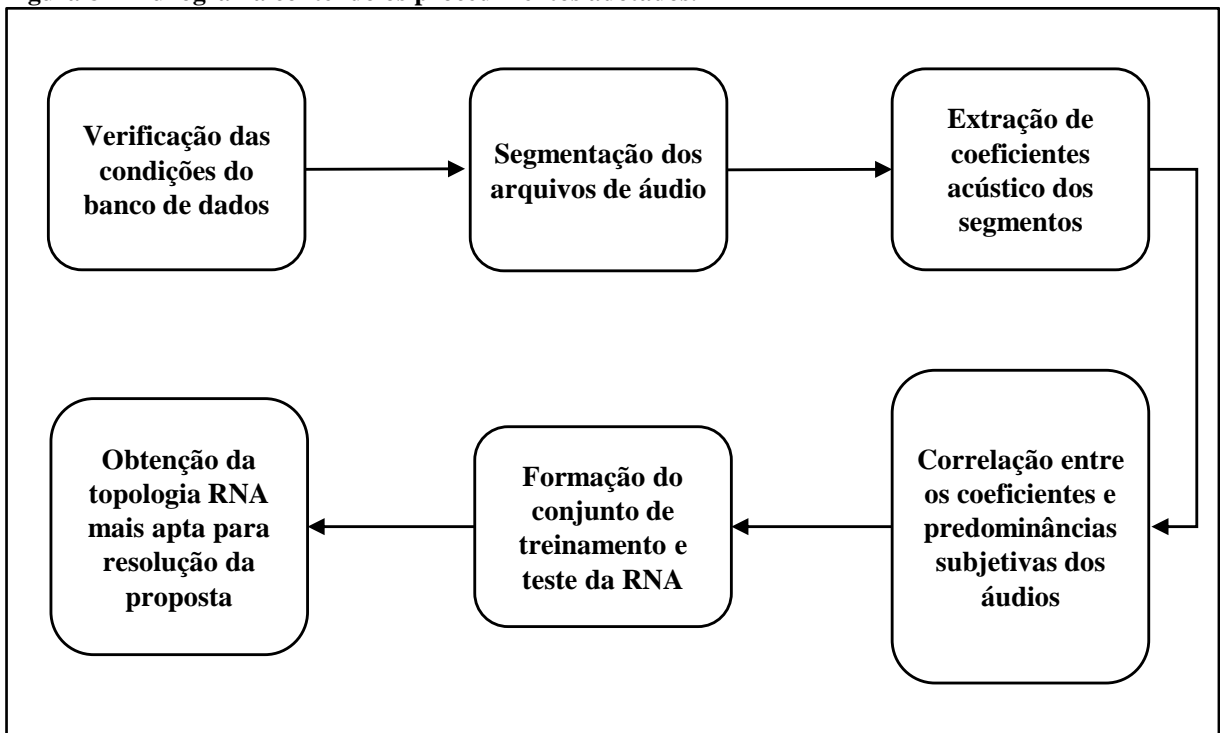
O processo de treinamento utilizando o algoritmo *backpropagation* de uma rede PMC pode ser dividido em duas etapas, uma etapa “para frente” (*forward*) e outra “para trás” (*backward*). Durante o processo “para frente”, o sinal percorre todas as camadas neurais, no sentido entrada-saída, visando somente obter as respostas da rede. O procedimento “para trás” consiste em comparar as saídas obtidas no passo anterior com as saídas desejadas, gerando valores de erro para cada neurônio de saída. O sinal percorre então o sentido contrário, saída-entrada, onde os pesos sinápticos e limiares de ativação são ajustados (SILVA; SPATTI; FLAUZINO, 2010, p. 94-95).

Realizando-se sucessivas repetições, os pesos sinápticos e limiares de ativação vão sendo ajustados, diminuindo o erro entre a resposta da rede e o valor desejado, até que se atinja um critério pré-estabelecido de parada. Assim, pode-se considerar esta etapa como um “ajuste de curva” entre os valores de entrada e saída. O objetivo deste ajuste é proporcionar à rede grande capacidade de generalização para que, após o processo de treinamento, ela seja capaz de relacionar, de alguma maneira, sinais nunca vistos com aqueles presentes no processo de treinamento (HAYKIN, 2000, p.232).

4 MATERIAIS UTILIZADOS E PROCEDIMENTOS ADOTADOS

Este capítulo contém informações a respeito do banco de dados utilizado, bem como a descrição de todos os procedimentos adotados durante a realização deste trabalho. Pode-se considerar como único material utilizado o banco de arquivos de áudio, uma vez que é sobre ele que todas as análises são realizadas. Todos os procedimentos adotados estão dispostos no fluxograma apresentado na Figura 8 e a explanação de cada um deles encontra-se nos subcapítulos seguintes.

Figura 8 – Fluxograma contendo os procedimentos adotados.



Fonte: Autoria Própria

4.1 BANCO DE DADOS

O banco de dados utilizado na elaboração deste trabalho foi cedido pelo Centro de Estudos da Voz – CEV, sendo este composto por trinta arquivos de trinta pacientes diferentes, possuindo como predominância subjetiva rugosidade, ou soprosidade, ou tensão. Todos os arquivos consistem em gravações da vogal sustentada /e/ registradas diretamente em um computador. Além dos arquivos de áudio, forneceu-se um parecer médico, atestando a condição vocal de cada uma das amostras (informando a predominância subjetiva da voz e gênero do

paciente), além de uma breve explicação dos procedimentos adotados durante o processo de gravação e avaliação.

Para o registro das vozes, os pacientes permaneceram em ambiente silencioso em uma sala de gravação. Os indivíduos ficam sentados durante todo o procedimento e a captura do áudio realizou-se por meio de um microfone de cabeça, marca Plantronics (modelo Audio 20), unidirecional, com faixa reta de resposta, posicionado cerca de 1cm do canto da boca do indivíduo, em posição diagonal. O microfone estava conectado a placa de som Sound Blaster Audigy 7.1, que por sua vez conectava-se a um computador. O registro do áudio foi feito utilizando-se o programa Sound Forge, versão 4.5, com taxa de amostragem de 44.100 Hz e quantização de 16 bits.

Cada amostra gravada foi editada no próprio software, desprezando-se o início da emissão, pois apresenta instabilidade típica da fase de ataque vocal. Os arquivos foram então avaliados e classificados por três fonoaudiólogas especialistas em voz, com mais de cinco anos de experiência em avaliações perceptiva-auditiva. Os arquivos utilizados estão igualmente distribuídos entre os três parâmetros e gênero dos pacientes, conforme observa-se na Tabela 4.

Tabela 4 – Distribuição dos arquivos que compõem o banco de dados de acordo com suas características.

Predominância Subjetiva	Masculino	Feminino
Rugosidade	5	5
Soprosidade	5	5
Tensão	5	5
Total	15	15

Fonte: Autoria Própria.

4.2 VERIFICAÇÃO DAS CONDIÇÕES DO BANCO DE DADOS

Por tratar-se de um conjunto de dados colhido, organizado e cedido por outro grupo de pesquisadores, alguns cuidados foram tomados como passos iniciais durante a realização deste trabalho. Algumas das características dos arquivos presentes no banco de dados puderam ser verificados, como por exemplo a vogal sustentada e o gênero do paciente de cada arquivo.

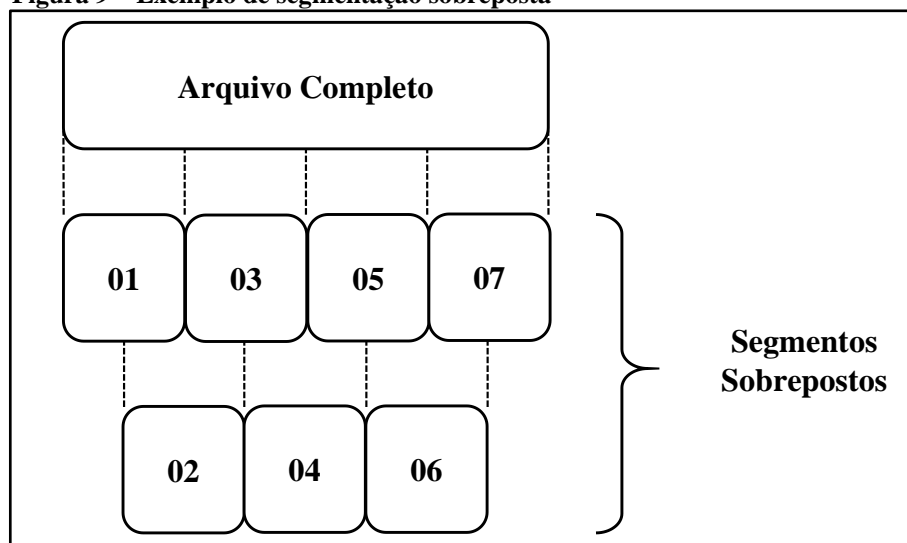
A predominância subjetiva de cada amostra não necessitou de tal verificação, uma vez que há um laudo médico atrelado à cada arquivo do banco de dados.

Para realização de tal procedimento, ouviu-se todos os arquivos identificando a vogal sustentada /e/ e o gênero do paciente, de acordo com os dados fornecidos para cada arquivo. Estando os 30 arquivos dentro das condições desejadas, o estudo prosseguiu sem a necessidade de retenção de nenhuma amostra do banco de dados.

4.3 SEGMENTAÇÃO DOS ARQUIVOS DE ÁUDIO

Buscando diminuir a complexidade computacional da análise, cada arquivo do banco de dados foi segmentado em porções menores. Este procedimento faz com que, de cada amostra do banco de dados, surjam inúmeros segmentos, todos sendo pequenas janelas do arquivo-mãe. Desta maneira, afim de suprimir a variação entre cada segmento e obter quantidade ainda maior de segmentos, o processo de segmentação foi realizado com sobreposição de 50% entre as janelas, conforme apresentado na Figura 9.

Figura 9 – Exemplo de segmentação sobreposta



Fonte: Autoria Própria

Encontram-se na literatura sugestões para que análises acústicas sejam realizadas em janelas que apresentam estacionaridade no sinal vocal. Tal condição é obtida em segmentos de áudio com duração de 20 a 30ms segundo Behlau (2001, p.133). Inicialmente utilizou-se janelas com duração aproximada de 23ms (1024 pontos), porém os resultados obtidos com este tamanho de janela não foram satisfatórios. Decidiu-se então testar outro tamanho de janela, com

duração aproximada de 46ms (2048 pontos), que resultou em uma taxa de acerto da RNA maior que o obtido com a janela de 23ms. Diante deste fato, decidiu-se aumentar novamente o tamanho da janela comparando os resultados obtidos. Para uma janela de aproximadamente 92ms (4096 pontos) obteve-se novamente uma taxa de acerto maior que as obtidas anteriormente.

Por questões estatísticas, descartou-se a possibilidade de um novo aumento no tamanho dos segmentos, pois isto colocaria a quantidade de segmentos abaixo do valor desejado para o estudo – pelo menos 1.000 padrões masculinos e 1.000 femininos. Assim, o procedimento de segmentação utilizou janelas de aproximadamente 92ms, resultando em número de segmentos por arquivo do banco de dados conforme apresentado nas Tabelas 5-8.

**Tabela 5 – Características dos arquivos do banco de dados –
Predominância: Rugosidade**

Arquivo	Gênero	Duração (s)	Segmentos Formados
02	Masculino	3,87	82
05	Masculino	4,41	94
07	Masculino	3,18	67
08	Feminino	3,56	75
09	Masculino	3,56	75
16	Feminino	3,34	71
17	Masculino	3,30	70
21	Feminino	3,18	67
28	Feminino	3,15	67
29	Feminino	3,15	67
Total			735

Fonte: Autoria Própria

**Tabela 6 – Características dos arquivos do banco de dados –
Predominância: Soproside**

Arquivo	Gênero	Duração (s)	Segmentos Formados
01	Feminino	5,27	112
03	Feminino	3,74	79
10	Masculino	3,40	72
11	Feminino	3,37	71
14	Feminino	2,98	63
20	Masculino	3,21	68
23	Masculino	3,18	67
24	Feminino	3,18	67
25	Masculino	3,18	67
26	Masculino	3,12	66
Total			732

Fonte: Autoria Própria

**Tabela 7 – Características dos arquivos do banco de dados –
Predominância: Tensão**

Arquivo	Gênero	Duração (s)	Segmentos Formados
04	Feminino	3,18	67
06	Masculino	3,61	76
12	Feminino	3,18	67
13	Masculino	3,56	75
15	Masculino	3,59	76
18	Feminino	3,34	71
19	Feminino	3,19	67
22	Feminino	3,18	67
27	Masculino	3,32	71
30	Masculino	3,15	67
Total			704

Fonte: Autoria Própria

Tabela 8 – Distribuição dos segmentos formados de acordo com a característica subjetiva dos arquivos de voz.

Predominância Subjetiva	Gênero	Número de segmentos formados
Rugosidade	Feminino	347
	Masculino	388
Soprosidade	Feminino	392
	Masculino	340
Tensão	Feminino	339
	Masculino	365
Total		2171

Fonte: Autoria Própria

4.4 EXTRAÇÃO DE CARACTERÍSTICA ACÚSTICA

A classificação por meio de RNAs necessita de entradas que representem o problema em estudo. Para o trabalho proposto se faz necessário apresentar à RNA coeficientes que possuam traços específicos, para que esta seja capaz de diferenciar uma condição vocal de outra, realizando assim a classificação propriamente dita.

O “extrator de características” necessário para implementação de ferramentas classificadoras não possui regra específica para sua formação, assim, para a elaboração deste trabalho realizaram-se testes¹ com alguns coeficientes utilizados em trabalhos semelhantes, escolhendo-se aquele com melhor desempenho. Dentre os vários testes realizados pode-se citar a utilização de transformadas como as de Fourier e Wavelets, além de coeficientes tais como entropia e energia do sinal, entropia de Shannon, entre outras. Optou-se por utilizar apenas coeficientes presentes nas *toolboxes* do *software* Matlab®, com o intuito de evitar possíveis problemas ocasionados pela implementação errada de qualquer coeficiente. Diante de tais testes, o melhor resultado foi obtido utilizando-se energia e entropia logarítmica, extraídas dos coeficientes do terceiro nível da transformada TWP.

O procedimento para extração de característica acústica dos segmentos de áudio, apresentado na Figura 10, inicia-se com a decomposição de cada um dos segmentos de áudio ao terceiro nível da TWP, resultando em 8 coeficientes. De cada um dos 8 coeficientes

¹ Tais testes consistem em pequenos procedimentos de treinamento e teste da RNA, utilizando-se em cada um dos testes topologias pré-definidas e comparando-se os resultados ao final do processo.

provenientes da TWP, extraem-se a energia e a energia logarítmica do sinal de áudio, conforme apresentado nas equações (1) e (2) (MATHEWORKS).

$$\mathbf{Energia} = \sum_{i=1}^n x_i^2 \quad (1)$$

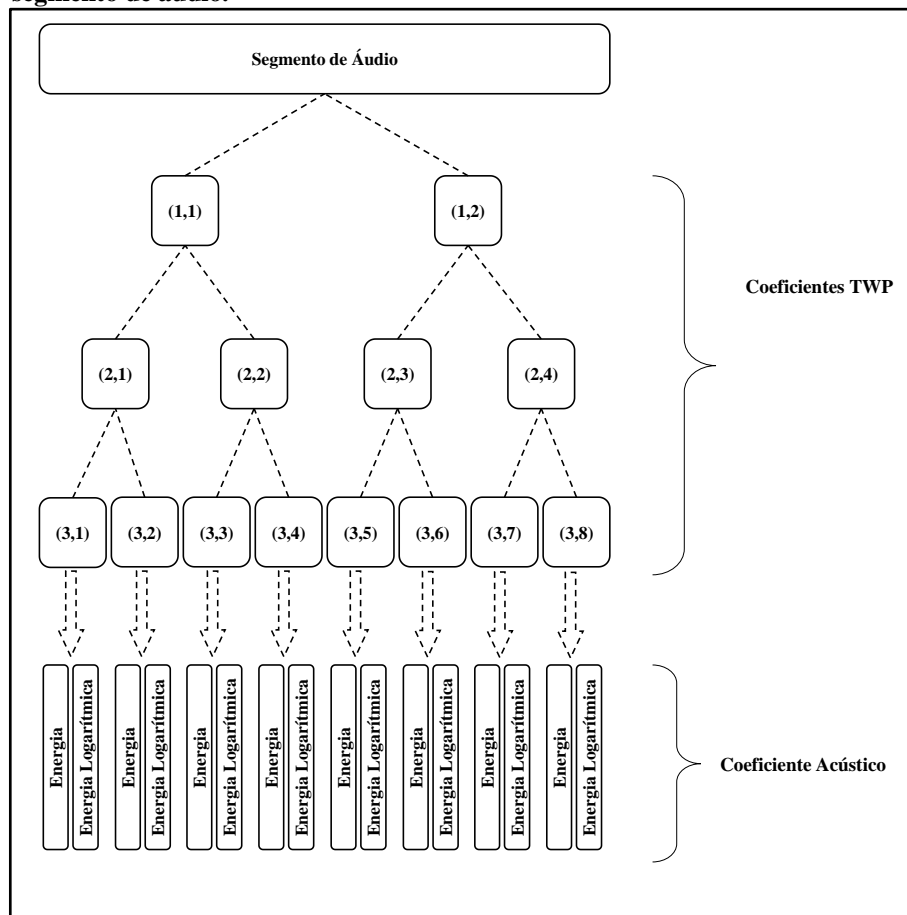
$$\mathbf{Energia\ Logarítmica} = \sum_{i=1}^n \log(x_i^2) \quad (2)$$

Em que:

n : Número de pontos existentes em cada um dos coeficientes da TWP;

x : Valor do ponto;

Figura 10 – Procedimento para extração de característica acústica de um segmento de áudio.



Fonte: Autoria Própria

4.5 CORRELAÇÃO ENTRE OS COEFICIENTES EXTRAÍDOS E A PREDOMINÂNCIA SUBJETIVA DOS ARQUIVOS DE ÁUDIO

O procedimento de extração de características acústicas resultou em 8 coeficientes de energia e outros 8 de energia logarítmica para cada segmento de áudio. A análise de correlação buscou mapear quais são os 8 coeficientes que possuem maior grau de informação pertinente à classificação desejada. Utilizou-se o coeficiente linear de Pearson para determinar o grau de correlação entre os coeficientes acústicos extraídos da TWP em relação à cada uma das predominâncias subjetivas. As análises foram feitas para cada uma das redes propostas, resultando assim em 6 análises.

O coeficiente de correlação de Pearson resulta em um coeficiente que assume valores entre -1 e +1, indicando correlação direta quando é positivo e inversa quando é negativo. Pode-se interpretar da seguinte maneira tais coeficientes: quanto mais próximo o valor de ρ estiver de zero, menor é o grau de correlação entre as grandezas. Quanto mais próximo de +1 ou -1, maior é o grau de correlação, podendo ser ele direto ou inverso.

Para avaliar o grau de correlação entre os coeficientes extraídos com uma das 3 predominâncias estudadas, utilizou-se a função '*corr*' presente na versão R2015a do *software* MATLAB[®]. Assim sendo, avaliaram-se o grau de correlação de todos os coeficientes extraídos com a predominância subjetiva própria de cada um dos segmentos. Apresenta-se nas Tabelas 9-14 os coeficientes de correlação obtido para cada um dos coeficientes analisados.

Tabela 9 – Coeficientes de correlação obtido entre os parâmetros acústicos extraídos dos arquivos femininos e a condição subjetiva Rugosidade.

Atribuição Genérica	Coefficiente acústico	Coefficiente de correlação obtido
01	Energia Coeficiente (3,1) WTP	0,2115
02	Energia Coeficiente (3,2) WTP	-0,2535
03	Energia Coeficiente (3,3) WTP	-0,1016
04	Energia Coeficiente (3,4) WTP	-0,1489
05	Energia Coeficiente (3,5) WTP	-0,2830
06	Energia Coeficiente (3,6) WTP	-0,1671
07	Energia Coeficiente (3,7) WTP	0,3246
08	Energia Coeficiente (3,8) WTP	-0,0588
09	Energia Logarítmica Coeficiente (3,1) WTP	0,1141
10	Energia Logarítmica Coeficiente (3,2) WTP	-0,1439
11	Energia Logarítmica Coeficiente (3,3) WTP	-0,1090
12	Energia Logarítmica Coeficiente (3,4) WTP	0,0379

Tabela 9 – Coeficientes de correlação obtido entre os parâmetros acústicos extraídos dos arquivos femininos e a condição subjetiva Rugosidade.

Atribuição Genérica	Coeficiente acústico	Coeficiente de correlação obtido
13	Energia Logarítmica Coeficiente (3,5) WTP	-0,1662
14	Energia Logarítmica Coeficiente (3,6) WTP	0,0222
15	Energia Logarítmica Coeficiente (3,7) WTP	0,2369
16	Energia Logarítmica Coeficiente (3,8) WTP	0,1255

Fonte: Autoria Própria

Tabela 10 – Coeficientes de correlação obtido entre os parâmetros acústicos extraídos dos arquivos femininos e a condição subjetiva Soproidade.

Atribuição Genérica	Coeficiente acústico	Coeficiente de correlação obtido
01	Energia Coeficiente (3,1) WTP	0,4700
02	Energia Coeficiente (3,2) WTP	-0,4657
03	Energia Coeficiente (3,3) WTP	-0,4906
04	Energia Coeficiente (3,4) WTP	-0,3458
05	Energia Coeficiente (3,5) WTP	-0,4746
06	Energia Coeficiente (3,6) WTP	-0,3677
07	Energia Coeficiente (3,7) WTP	-0,1538
08	Energia Coeficiente (3,8) WTP	-0,2060
09	Energia Logarítmica Coeficiente (3,1) WTP	-0,0416
10	Energia Logarítmica Coeficiente (3,2) WTP	-0,3799
11	Energia Logarítmica Coeficiente (3,3) WTP	-0,3706
12	Energia Logarítmica Coeficiente (3,4) WTP	-0,4668
13	Energia Logarítmica Coeficiente (3,5) WTP	-0,3439
14	Energia Logarítmica Coeficiente (3,6) WTP	-0,4712
15	Energia Logarítmica Coeficiente (3,7) WTP	-0,3462
16	Energia Logarítmica Coeficiente (3,8) WTP	-0,4077

Fonte: Autoria Própria

Tabela 11 – Coeficientes de correlação obtido entre os parâmetros acústicos extraídos dos arquivos femininos e a condição subjetiva Tensão.

Atribuição Genérica	Coefficiente acústico	Coefficiente de correlação obtido
01	Energia Coeficiente (3,1) WTP	-0,6987
02	Energia Coeficiente (3,2) WTP	0,7365
03	Energia Coeficiente (3,3) WTP	0,6096
04	Energia Coeficiente (3,4) WTP	0,5074
05	Energia Coeficiente (3,5) WTP	0,7753
06	Energia Coeficiente (3,6) WTP	0,5483
07	Energia Coeficiente (3,7) WTP	-0,1672
08	Energia Coeficiente (3,8) WTP	0,2723
09	Energia Logarítmica Coeficiente (3,1) WTP	0,1577
10	Energia Logarítmica Coeficiente (3,2) WTP	0,5376
11	Energia Logarítmica Coeficiente (3,3) WTP	0,4929
12	Energia Logarítmica Coeficiente (3,4) WTP	0,4447
13	Energia Logarítmica Coeficiente (3,5) WTP	0,5228
14	Energia Logarítmica Coeficiente (3,6) WTP	0,4651
15	Energia Logarítmica Coeficiente (3,7) WTP	0,1199
16	Energia Logarítmica Coeficiente (3,8) WTP	0,2955

Fonte: Autoria Própria

Tabela 12 – Coeficiente de correlação obtido entre os parâmetros acústicos extraídos dos arquivos masculinos e a condição subjetiva Rugosidade.

Atribuição Genérica	Coefficiente acústico	Coefficiente de correlação obtido
01	Energia Coeficiente (3,1) WTP	0,3118
02	Energia Coeficiente (3,2) WTP	-0,3562
03	Energia Coeficiente (3,3) WTP	-0,3925
04	Energia Coeficiente (3,4) WTP	-0,1593
05	Energia Coeficiente (3,5) WTP	-0,5006
06	Energia Coeficiente (3,6) WTP	-0,1671
07	Energia Coeficiente (3,7) WTP	-0,1945
08	Energia Coeficiente (3,8) WTP	-0,1439
09	Energia Logarítmica Coeficiente (3,1) WTP	-0,0486
10	Energia Logarítmica Coeficiente (3,2) WTP	-0,4864
11	Energia Logarítmica Coeficiente (3,3) WTP	-0,5418

Tabela 12 – Coeficiente de correlação obtido entre os parâmetros acústicos extraídos dos arquivos masculinos e a condição subjetiva Rugosidade.

Atribuição Genérica	Coeficiente acústico	Coeficiente de correlação obtido
12	Energia Logarítmica Coeficiente (3,4) WTP	-0,4231
13	Energia Logarítmica Coeficiente (3,5) WTP	-0,5207
14	Energia Logarítmica Coeficiente (3,6) WTP	-0,4147
15	Energia Logarítmica Coeficiente (3,7) WTP	-0,4586
16	Energia Logarítmica Coeficiente (3,8) WTP	-0,4347

Fonte: Autoria Própria

Tabela 13 – Coeficiente de correlação obtido entre os parâmetros acústicos extraídos dos arquivos masculinos e a condição subjetiva Soprosidade.

Atribuição Genérica	Coeficiente acústico	Coeficiente de correlação obtido
01	Energia Coeficiente (3,1) WTP	-0,6057
02	Energia Coeficiente (3,2) WTP	0,6078
03	Energia Coeficiente (3,3) WTP	0,4834
04	Energia Coeficiente (3,4) WTP	0,5416
05	Energia Coeficiente (3,5) WTP	0,5104
06	Energia Coeficiente (3,6) WTP	0,5637
07	Energia Coeficiente (3,7) WTP	0,3129
08	Energia Coeficiente (3,8) WTP	0,4572
09	Energia Logarítmica Coeficiente (3,1) WTP	-0,3701
10	Energia Logarítmica Coeficiente (3,2) WTP	0,2922
11	Energia Logarítmica Coeficiente (3,3) WTP	0,2751
12	Energia Logarítmica Coeficiente (3,4) WTP	0,4631
13	Energia Logarítmica Coeficiente (3,5) WTP	0,1800
14	Energia Logarítmica Coeficiente (3,6) WTP	0,4554
15	Energia Logarítmica Coeficiente (3,7) WTP	0,4032
16	Energia Logarítmica Coeficiente (3,8) WTP	0,4708

Fonte: Autoria Própria

Tabela 14 – Coeficiente de correlação obtido entre os parâmetros acústicos extraídos dos arquivos masculinos e a condição subjetiva Tensão.

Atribuição Genérica	Coeficiente acústico	Coeficiente de correlação obtido
01	Energia Coeficiente (3,1) WTP	0,2784
02	Energia Coeficiente (3,2) WTP	-0,2354
03	Energia Coeficiente (3,3) WTP	-0,0762
04	Energia Coeficiente (3,4) WTP	-0,3702
05	Energia Coeficiente (3,5) WTP	0,0069
06	Energia Coeficiente (3,6) WTP	-0,3839
07	Energia Coeficiente (3,7) WTP	-0,1098
08	Energia Coeficiente (3,8) WTP	-0,3029
09	Energia Logarítmica Coeficiente (3,1) WTP	0,4127
10	Energia Logarítmica Coeficiente (3,2) WTP	0,2067
11	Energia Logarítmica Coeficiente (3,3) WTP	0,2798
12	Energia Logarítmica Coeficiente (3,4) WTP	-0,0252
13	Energia Logarítmica Coeficiente (3,5) WTP	0,3518
14	Energia Logarítmica Coeficiente (3,6) WTP	-0,0262
15	Energia Logarítmica Coeficiente (3,7) WTP	0,0699
16	Energia Logarítmica Coeficiente (3,8) WTP	-0,0211

Fonte: Autoria Própria

4.6 FORMAÇÃO DO CONJUNTO DE TREINAMENTO E TESTE

Inicialmente, devem-se condicionar os dados utilizados dentro dos valores limites da função de ativação a ser utilizada, pois assim, estando todos os valores em um mesmo universo de discurso, evita-se saturação dos dados e erros nas operações realizadas entre os neurônios. Normalizaram-se os 16 valores provenientes dos coeficientes da TWP de cada amostra de áudio, de -1 a +1. Este procedimento foi realizado utilizando-se a função ‘*mapminmax*’ presente na versão R2015a do *software* MATLAB[®], operando da seguinte maneira: dispendo-se todos os coeficientes em uma tabela, tem-se 16 colunas por ‘*n*’ linhas. A normalização é realizada por colunas, onde o maior coeficiente recebe valor +1, o menor -1 e os valores intermediários assumem valores proporcionais entre estes dois extremos, existindo assim um valor máximo e um valor mínimo em cada uma das colunas da tabela.

Com o intuito de exigir mais da capacidade de generalização da RNA, foram dispostos de maneira aleatória todos os coeficientes normalizados, buscando evitar que dois segmentos seguidos fossem oriundos de um mesmo arquivo de áudio. Para um treinamento bem-sucedido também se faz necessário apresentar os valores máximos e mínimos dos coeficientes para a RNA durante o processo treinamento. Assim sendo, moveu-se tais valores para a parte inicial do banco de coeficientes.

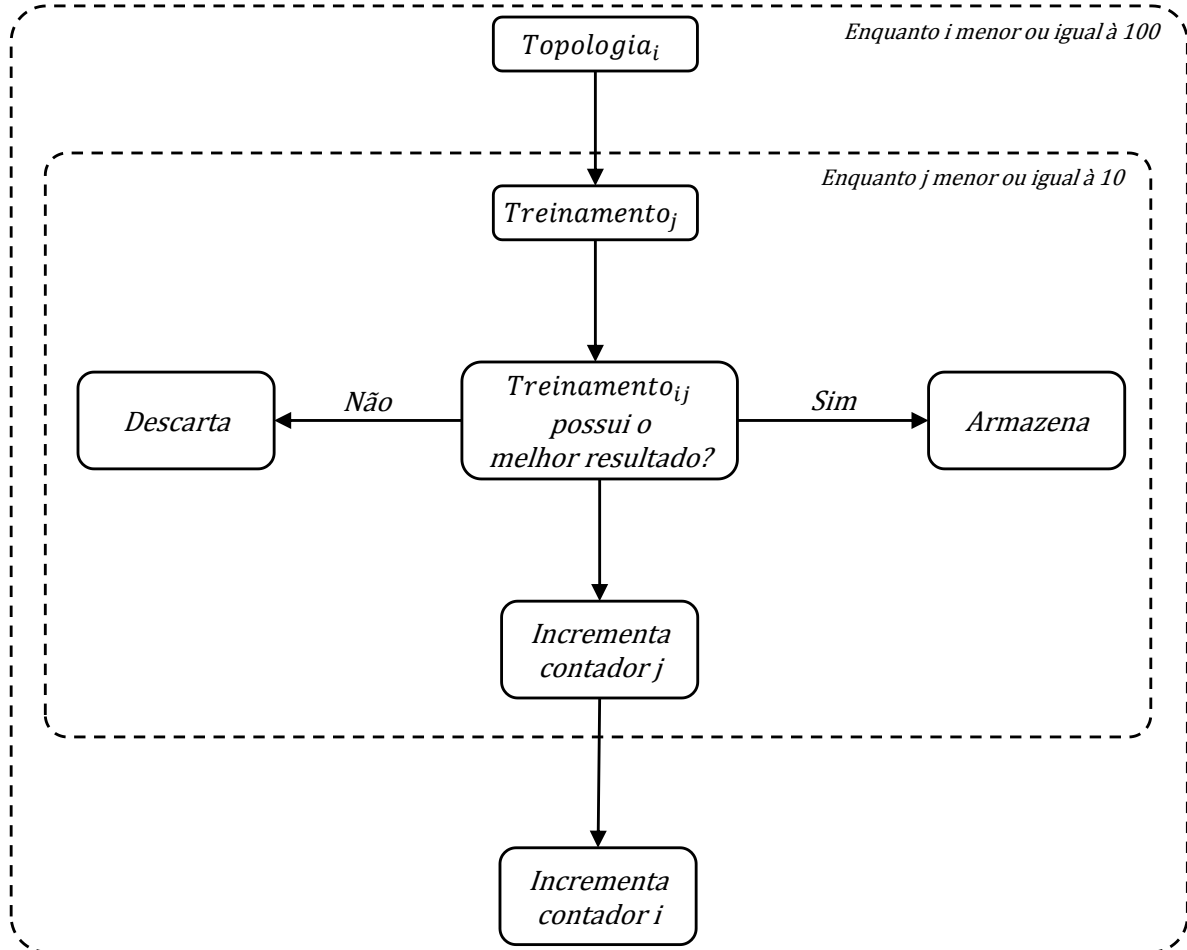
Após tomar-se os cuidados descritos nos parágrafos anteriores, pode-se separar o banco de coeficientes em duas porções, uma para treinamento e outra para teste da RNA. Não há definição clara quanto a proporção dos dados a ser utilizada em cada um destes conjuntos, optou-se pela utilização proporcional de dados em cada um dos conjuntos, utilizando-se 50% das amostras para o processo de treinamento e outros 50% para a etapa de teste.

4.7 OBTENÇÃO DA TOPOLOGIA RNA MAIS APTA

Para determinação da topologia RNA capaz de classificar adequadamente as amostras de áudio alguns procedimentos empíricos foram necessários. Com recursos computacionais e tempo para desenvolvimento deste trabalho limitados, estipulou-se efetuar 100 variações de topologia em duas camadas neurais, tomando aquela com maior taxa de acerto como rede mais apta. As topologias testadas foram permutações entre os números ímpares presentes entre 3 e 21, incluindo os dois numerais extremos. Para cada uma das topologias efetuavam-se 10 repetições, armazenando-se o melhor resultado obtido para posterior comparação com resultados obtidos pelas demais topologias propostas. Tais repetições dentro de uma mesma

topologia buscam evitar que o algoritmo de treinamento da RNA fique estacionado em um possível mínimo local. A Figura 11 apresenta com maiores detalhes os passos para desenvolvimento do procedimento de treinamento.

Figura 11 – Algoritmo utilizado durante o procedimento de treinamento das RNAs.



Fonte: Autoria Própria

Em um primeiro momento optou-se por realizar a classificação utilizando três redes distintas, cada uma responsável pela identificação de uma das características subjetivas da voz estudadas. Tal procedimento obteve bons resultados, conforme publicado no artigo descrito na introdução deste trabalho². Realizou-se outros procedimentos e testes e melhores resultados foram obtidos, considerando seis redes, sendo estas separadas por gênero e pelas características subjetivas a serem classificadas.

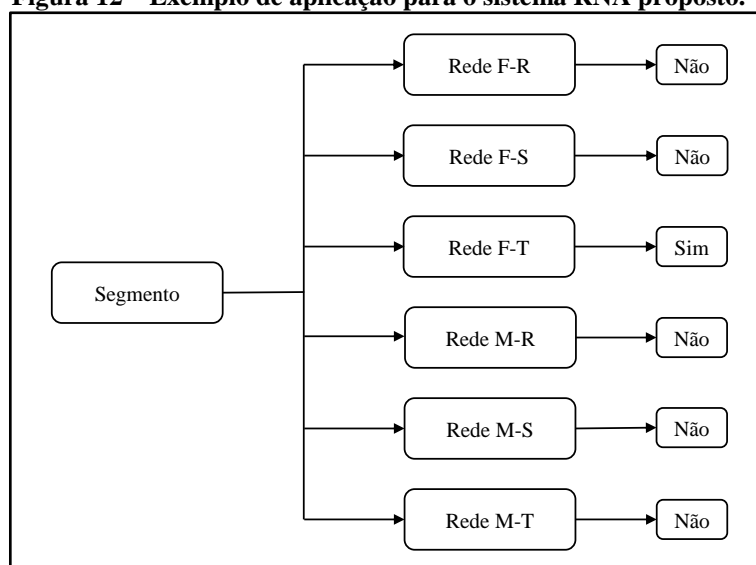
A Figura 12 representa o sistema RNA proposto, em que, em seu funcionamento ideal, apenas uma das redes do sistema deve classificar positivamente uma amostra de áudio. Como

² Apresenta-se e compara-se os resultados publicados no artigo em subseção própria dentro do capítulo de Resultados.

para a resolução do problema estudado, abordou-se separadamente cada parâmetro subjetivo e gênero dos pacientes, apresenta-se separadamente resultados para todas as condições estudadas. Buscando-se simplificar as informações aqui apresentadas, adotou-se as seguintes abreviações:

- **F-R:** Rede relacionada aos arquivos femininos do parâmetro subjetivo rugosidade;
- **F-S:** Rede relacionada aos arquivos femininos do parâmetro subjetivo soproosidade;
- **F-T:** Rede relacionada aos arquivos femininos do parâmetro subjetivo tensão;
- **M-R:** Rede relacionada aos arquivos masculinos do parâmetro subjetivo rugosidade;
- **M-S:** Rede relacionada aos arquivos masculinos do parâmetro subjetivo soproosidade;
- **M-T:** Rede relacionada aos arquivos masculinos do parâmetro subjetivo tensão;

Figura 12 – Exemplo de aplicação para o sistema RNA proposto.



Fonte: Autoria Própria

5 RESULTADOS OBTIDOS

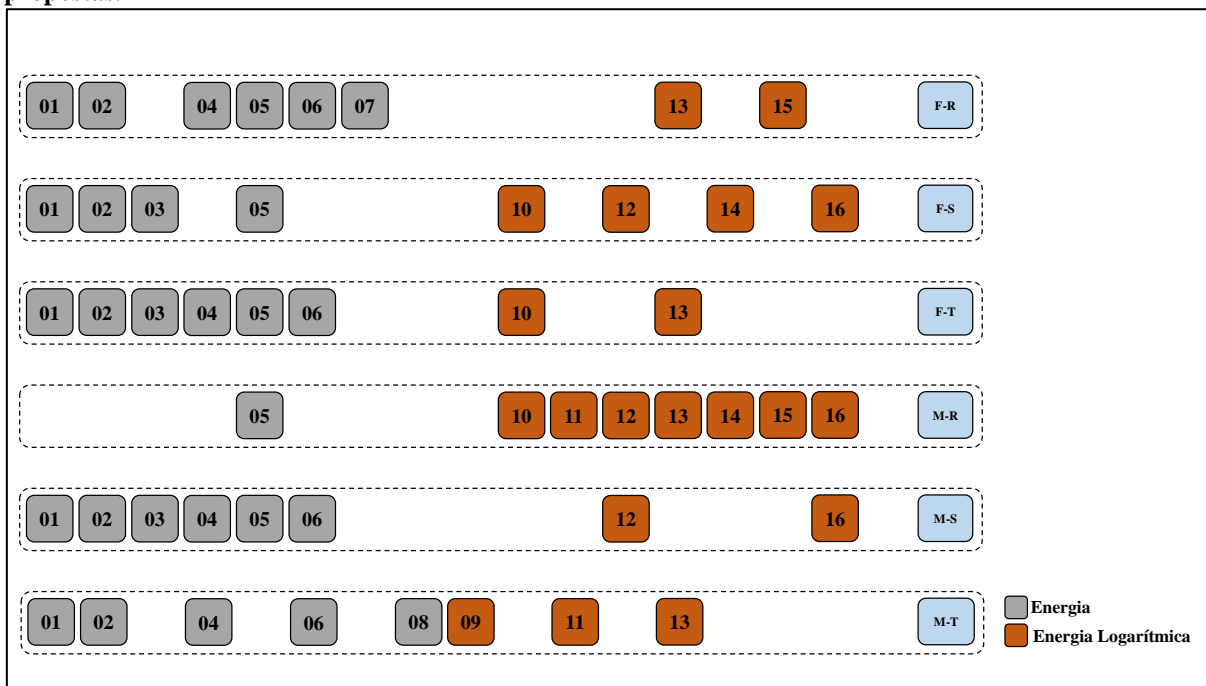
A busca por uma topologia compacta, capaz de resolver o problema em estudo culminou em um sistema RNA composto por 6 redes individuais, cada uma especialista na identificação de uma das características subjetivas. Assim sendo, a diferença existente entre cada rede é apenas o número de neurônios presentes nas camadas escondidas e quais coeficientes provenientes da TWP foram apresentados para as mesmas. Tais diferenças são resultado das análises de correlação para os coeficientes da TWP e dos procedimentos de treinamento e teste para as RNAs. Considerou-se como mais apta a rede que apresentou maior taxa de acerto com topologia mais compacta.

5.1 COEFICIENTES DA TWP UTILIZADOS

Conforme apresentado no subcapítulo 4.5, neste trabalho não foram utilizados todos os coeficientes extraídos dos segmentos de áudio por meio da TWP. Empregou-se aqueles coeficientes que possuíam maior correlação com cada um dos parâmetros estudados, ou seja, cada RNA é alimentada por coeficientes diferentes de acordo com a resposta do procedimento de correlação.

Apresenta-se na Figura 13 os 8 melhores coeficientes para cada uma das redes estudadas, não estando estes coeficientes em ordem crescente quanto seu grau de correlação e estando os mesmos com suas mesmas atribuições genéricas apresentadas nas tabelas do subcapítulo 4.5. Observa-se também a disparidade entre os coeficientes utilizados em cada uma das classes estudadas.

Figura 13 – Coeficientes acústicos utilizados durante a etapa de treinamento de cada uma das RNAs propostas.



Fonte: Autoria Própria

5.2 EFICIÊNCIA DAS TOPOLOGIAS TESTADAS

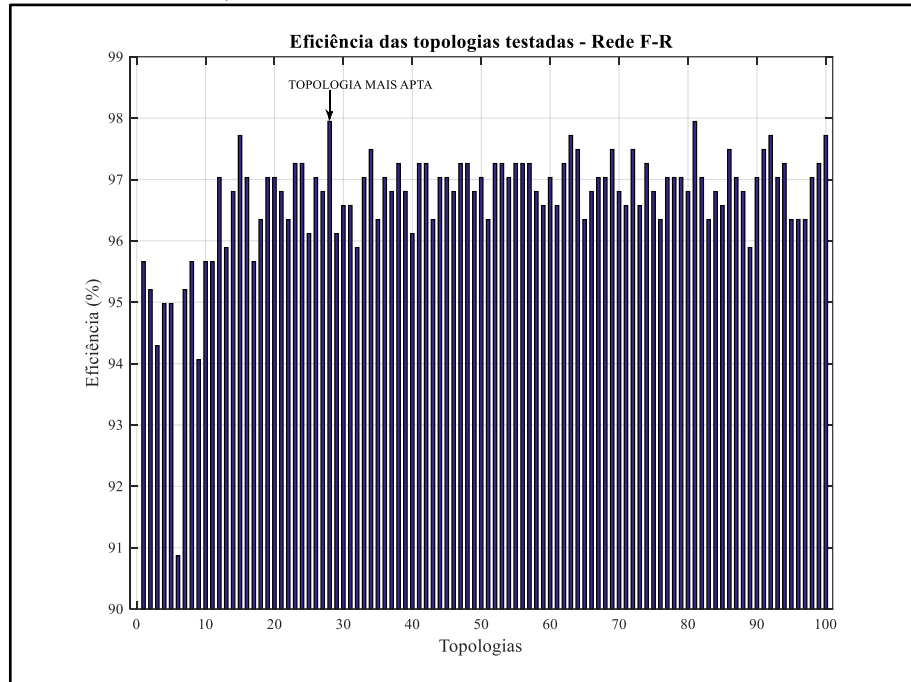
O parâmetro para escolha da topologia mais apta para resolução do problema de classificação é a eficiência da rede durante o teste de validação. Apresenta-se na Tabela 16 informações quanto o percentual de acerto para cada uma das RNAs propostas bem como a topologia mais apta. Ilustra-se nas Figuras 13 – 18 uma visão geral da eficiência obtida durante o treinamento.

Tabela 15 – Percentual de acerto da RNA mais apta à resolução do problema estudado e quantidade de neurônios pertencentes as camadas escondidas das mesmas.

Rede	Taxa de acerto (%)	Neurônios 1ª Camada	Neurônios 2ª Camada
F-R	97,95	7	17
F-S	97,95	5	3
F-T	99,09	9	5
M-R	98,77	3	13
M-S	99,54	7	15
M-T	99,54	3	15

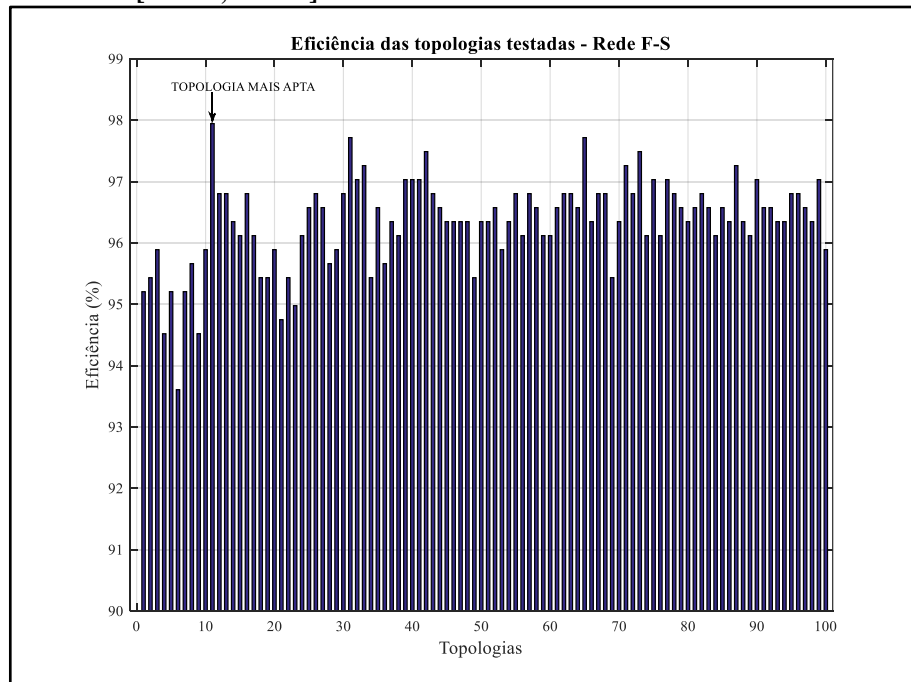
Fonte: Autoria Própria

Figura 14 – Percentual de eficiência para cada uma das topologias testadas para a rede F-R [N1 = 7; N2 = 17]



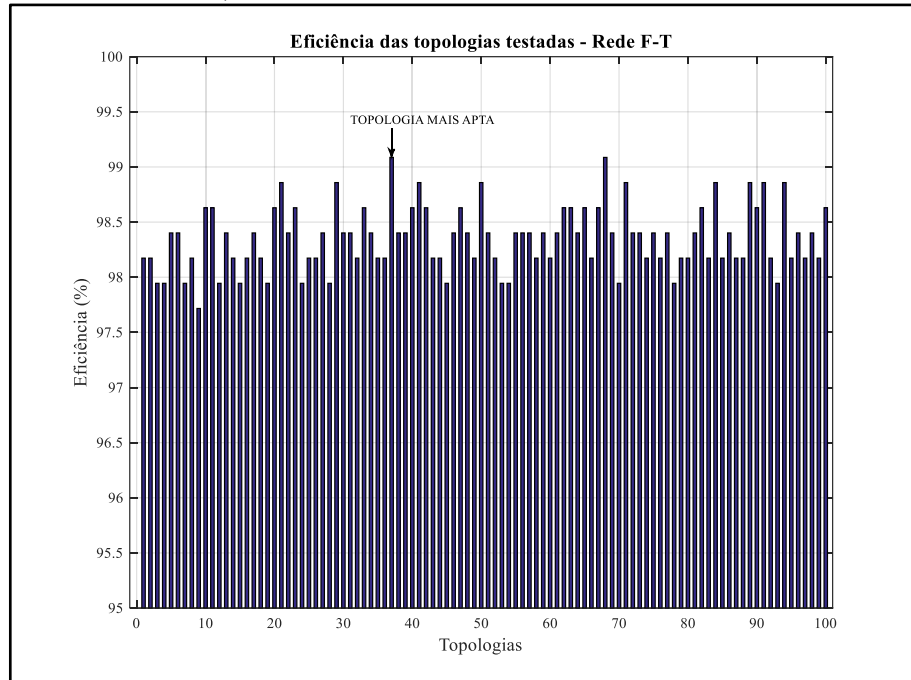
Fonte: Autoria Própria

Figura 15 – Percentual de eficiência para cada uma das topologias testadas para a rede F-S [N1 = 5; N2 = 3]



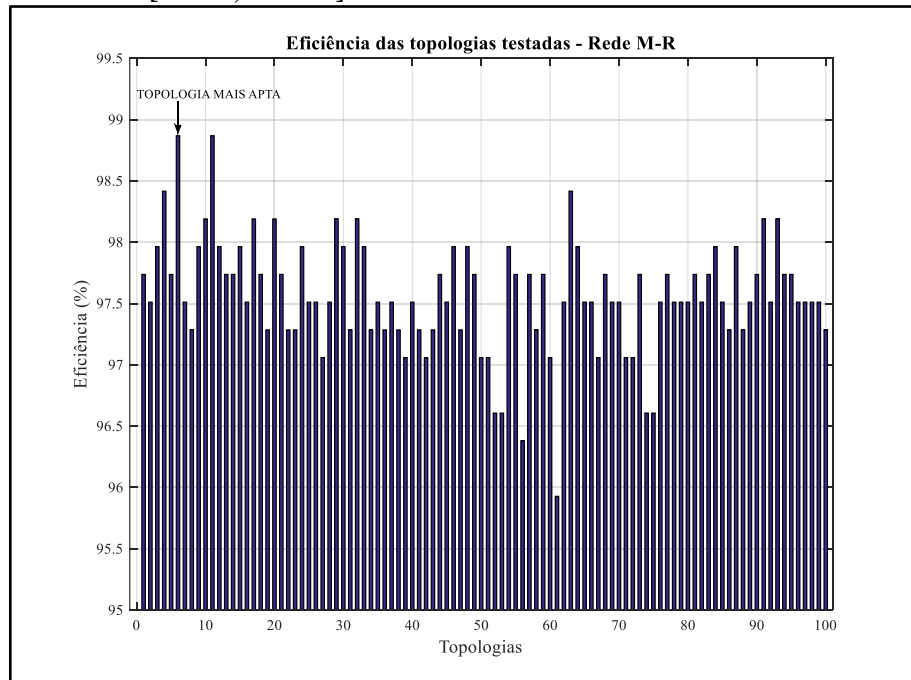
Fonte: Autoria Própria

Figura 16 – Percentual de eficiência para cada uma das topologias testadas para a rede F-T [N1 = 9; N2 = 5]



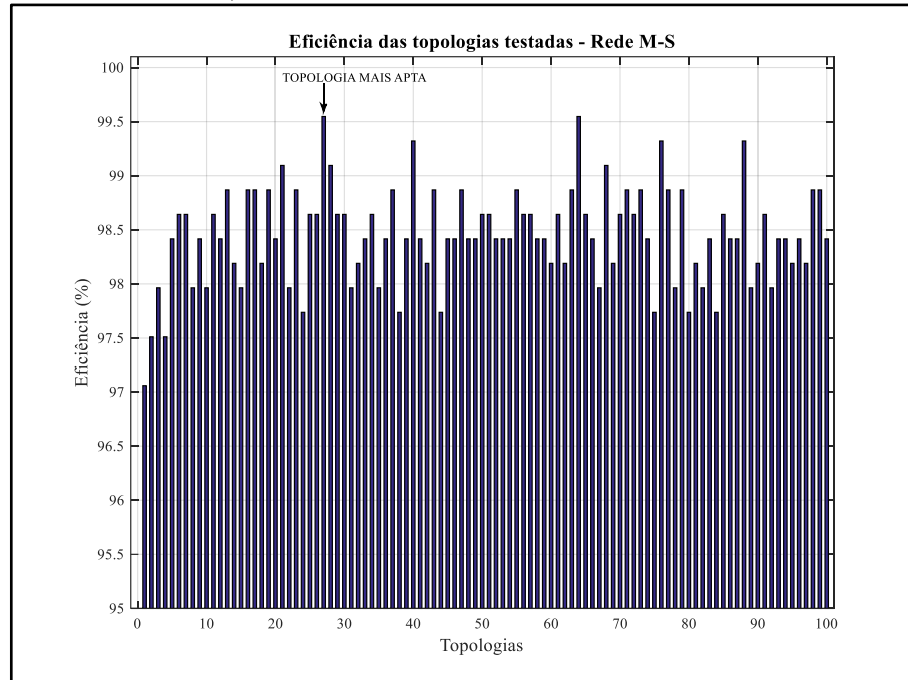
Fonte: Autoria Própria

Figura 17 – Percentual de eficiência para cada uma das topologias testadas para a rede M-R [N1 = 3; N2 = 13]



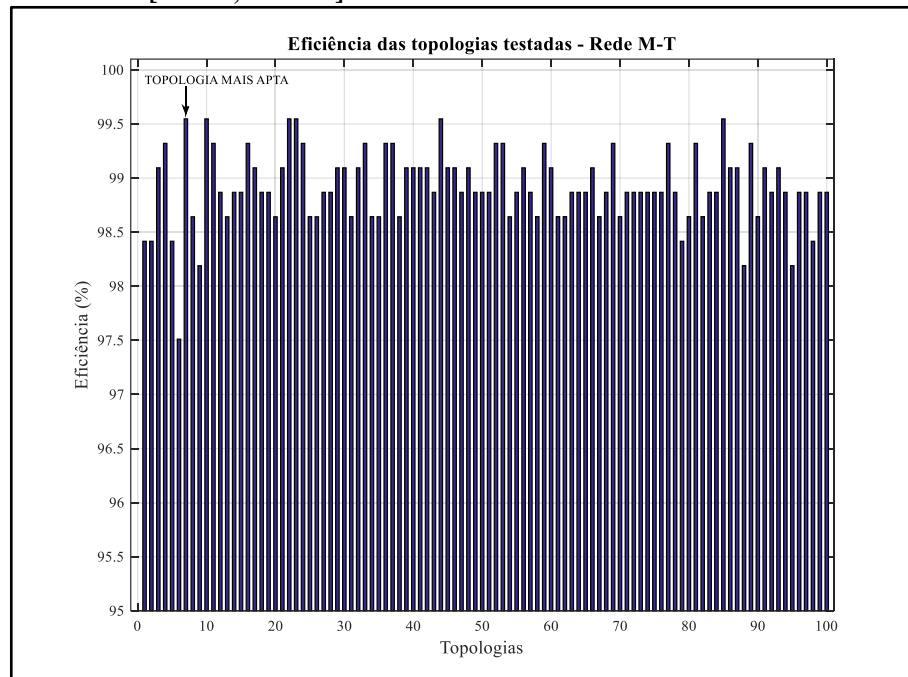
Fonte: Autoria Própria

Figura 18 – Percentual de eficiência para cada uma das topologias testadas para a rede M-S [N1 = 7; N2 = 15]



Fonte: Autoria Própria

Figura 19 – Percentual de eficiência para cada uma das topologias testadas para a rede M-T [N1 = 3; N2 = 15]



Fonte: Autoria Própria

5.3 RESPOSTA AO PROCEDIMENTO DE VALIDAÇÃO

Após obtenção da topologia mais apta para resolução do problema, apresentou-se a esta topologia a porção de segmentos dos arquivos do banco de dados destinada a teste – aqueles que não foram apresentados à RNA. Cada segmento foi então estimado dentro do intervalo de -1 a +1, onde, para este trabalho, -1 significa “não pertence à classe de especialidade da RNA” e +1, “pertence à classe de especialidade da RNA”. Pela utilização de uma rampa linear como função de ativação, esperava-se valores de saída dispersos por toda a faixa de definição da rampa, com maior concentração de resultados próximos aos extremos.

Levando em conta esta questão, considerou-se um limitante de pós processamento de 10% para simplificar a análise final, contendo 5% para cada extremo. Este limitante é responsável por definir a resposta da RNA como *classificado* – quando o valor estimado fica para além da linha de pós processamento e *indefinido*, quando o mesmo está entre estes limitantes. Esta ideia está apresentada matematicamente pela Equação 3, onde ‘*a*’ representa o valor de saída da RNA.

$$\begin{cases} a > 0,95 \rightarrow 1 \\ -0,95 \leq a \leq 0,95 \rightarrow 0 \\ a < -0,95 \rightarrow -1 \end{cases} \quad (3)$$

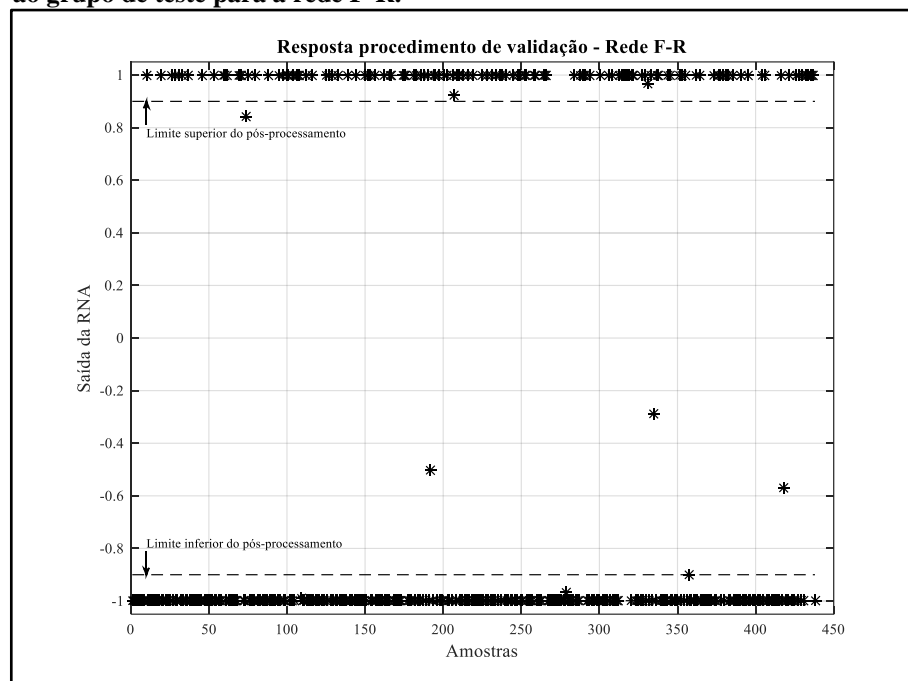
Esta etapa tem como objetivo atestar a capacidade da RNA treinada classificar um segmento de áudio entre pertencente ou não-pertencente de acordo com a especialidade daquela rede. Posteriormente outros mecanismos serão utilizados para avaliar o nível de acerto destas respostas. Neste trabalho, o percentual de indefinição é contado como erro de classificação, porém verifica-se na Tabela 16 índices bem baixos para esta condição durante a classificação. Apresenta-se nas Figuras 20, 21, 22, 23, 24 e 25 a resposta das RNAs sem o pós-processamento a fim de evidenciar a localização de tais imprecisões.

Tabela 16 – Percentual de Indefinição para cada RNA proposta

Rede	Indefinição (%)
F-R	0,9132
F-S	0,2283
F-T	0
M-R	0,4525
M-S	0
M-T	0,4525

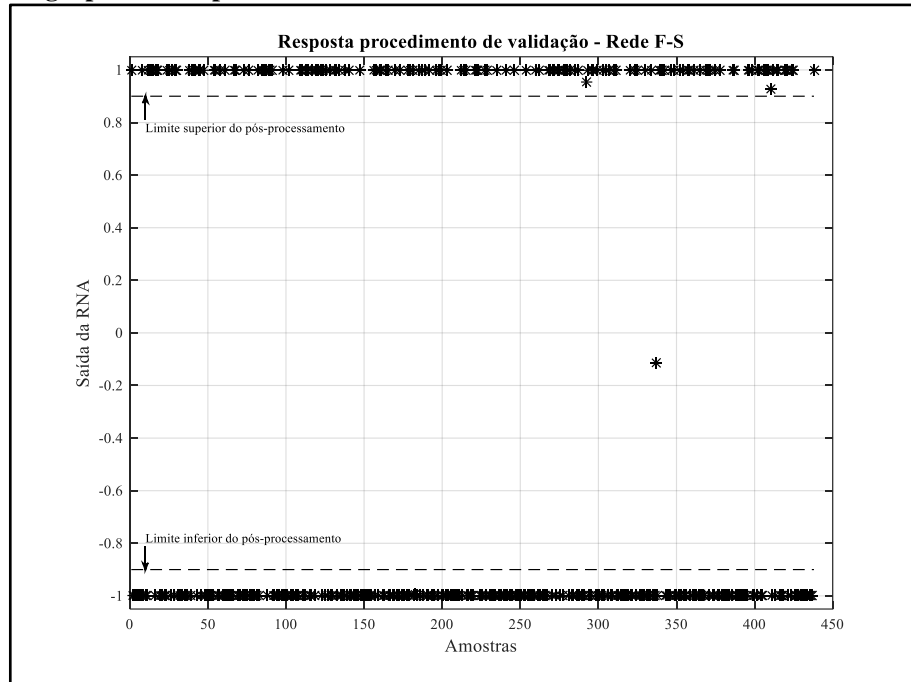
Fonte: Aatoria Própria

Figura 20 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede F-R.



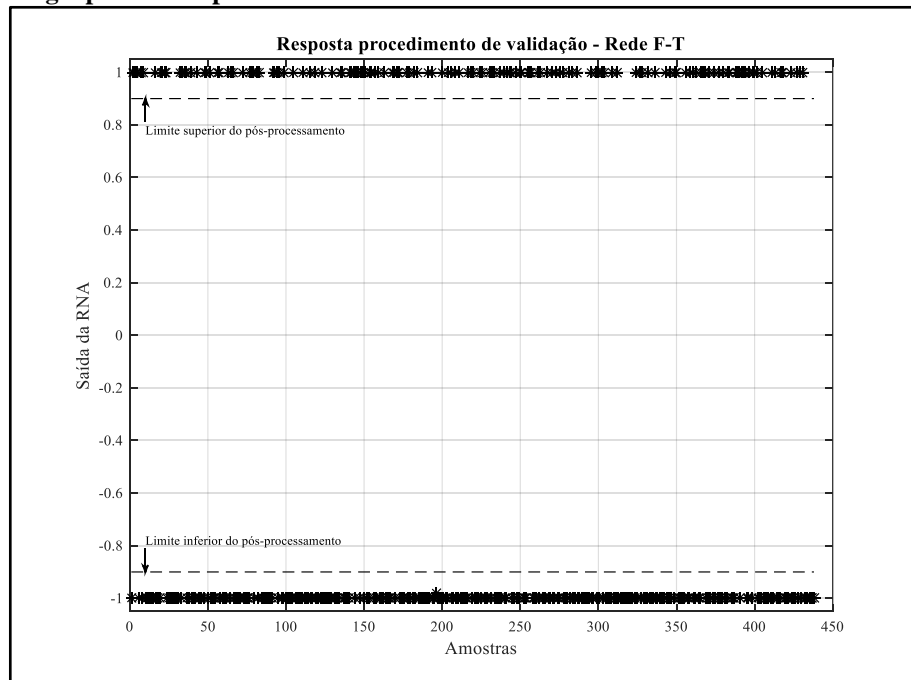
Fonte: Aatoria Própria

Figura 21 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede F-S.



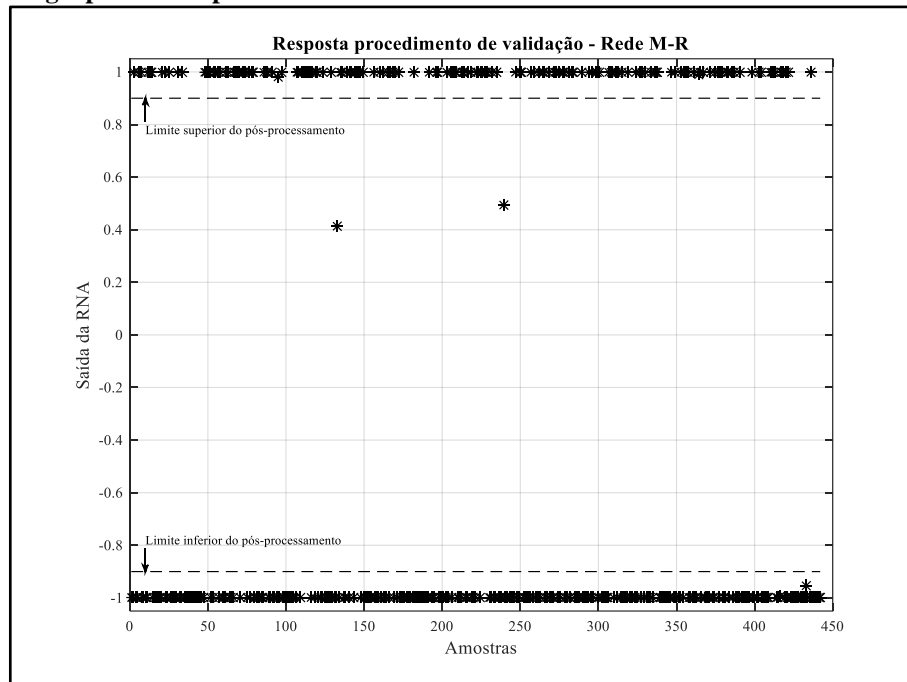
Fonte: Autoria Própria

Figura 22 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede F-T.



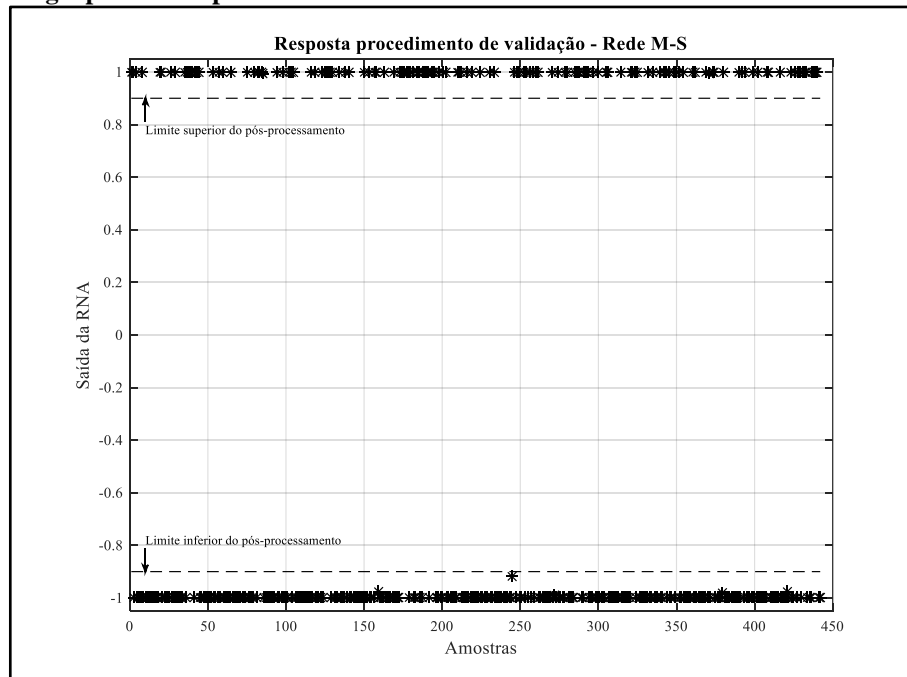
Fonte: Autoria Própria

Figura 23 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede M-R.



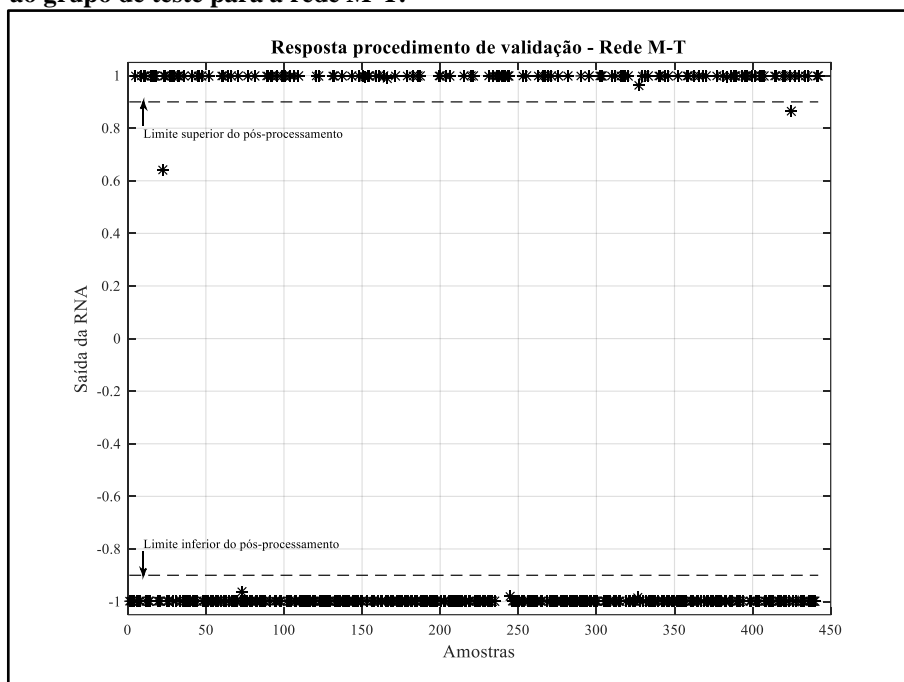
Fonte: Autoria Própria

Figura 24 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede M-S.



Fonte: Autoria Própria

Figura 25 – Resposta ao procedimento de validação dos segmentos pertencentes ao grupo de teste para a rede M-T.



Fonte: Autoria Própria

5.4 TAXA DE ACERTO E COMPARATIVO COM RESULTADOS DO ARTIGO PUBLICADO

Para obtenção da taxa de acertos, utilizou-se uma regra proporcional simples: número de segmentos classificados corretamente divididos pelo total de segmentos analisados, considerando os segmentos com classificação *indeterminada* como erros. Os resultados já foram apresentados na Tabela 15 no início deste capítulo.

A fim de comparar os resultados obtidos com o desenvolver deste trabalho e os resultados alcançados durante a elaboração do artigo, apresenta-se a Tabela X:

Tabela 17 – Comparativo entre procedimentos e resultados do artigo e trabalho de conclusão de curso

	Artigo	TCC
Banco de dados	30 arquivos contendo predominâncias subjetivas (rugosidade, soproisidade e tensão) cedido pelo CEV	30 arquivos contendo predominâncias subjetivas (rugosidade, soproisidade e tensão) cedido pelo CEV
Segmentação dos arquivos	Segmentação sobreposta com janelas de 4096 pontos.	Segmentação sobreposta com janelas de 4096 pontos.

Tabela 17 – Comparativo entre procedimentos e resultados do artigo e trabalho de conclusão de curso

	Artigo	TCC
Coefficientes Extraídos	Energia de cada um dos 8 coeficientes do 3º nível da WPT	Energia e energia logarítmica de cada um dos 8 coeficientes do 3º nível da WPT
Análise de Correlação	Não realizada	Busca obter os 8 coeficientes com maior nível de informação útil para classificação das amostras
RNAs	3 redes independentes, uma especialista para cada um dos parâmetros subjetivos estudados	6 redes independentes, uma especialista para cada um dos parâmetros subjetivos estudados divididos por gênero.
Taxa de acerto – Rugosidade	96,84%	Feminino: 97,95%
		Masculino: 98,77%
Taxa de acerto – Soproisidade	94,13%	Feminino: 97,95%
		Masculino: 99,54%
Taxa de acerto – Tensão	95,58%	Feminino: 99,09%
		Masculino: 99,54%
Taxa de acerto média	95,51%	98,80%
Análise de multiplicidade de classificação	Não realizada	Leva em consideração as possibilidades de falha durante a operação do sistema proposto. Possui taxa média de confiabilidade de 96,33% ³

Diante das informações apresentadas na Tabela 17 verifica-se maior robustez para o sistema proposto, bem como melhoria dos resultados obtidos ocasionados pela inserção de novos procedimentos.

5.5 DESVIO PERCENTUAL DO PROCEDIMENTO DE TESTES DAS RNAs

Um parâmetro capaz de averiguar a qualidade de classificação da RNA é o desvio percentual. Este percentual demonstra o quão longe os resultados estimados pela rede estão de seus respectivos valores reais, sendo 0% quando o valor estimado é idêntico ao valor real e 100% quando ocorre o maior erro admitido pelo sistema.

³ Conforme procedimento descrito na seção 5.6.1

Para este estudo os valores de saída das RNA são sempre valores dispostos entre -1 e +1 (resultado da utilização de uma rampa linear como função de ativação) e os valores reais atribuídos a cada um dos segmentos estudados são -1 e +1, que representam respectivamente: não pertence à classe e pertence à classe. Para cálculo do desvio percentual não foram levadas em consideração as regras de pós-processamento utilizadas em etapas anteriores, uma vez que o pós-processamento mascararia informações a serem demonstradas.

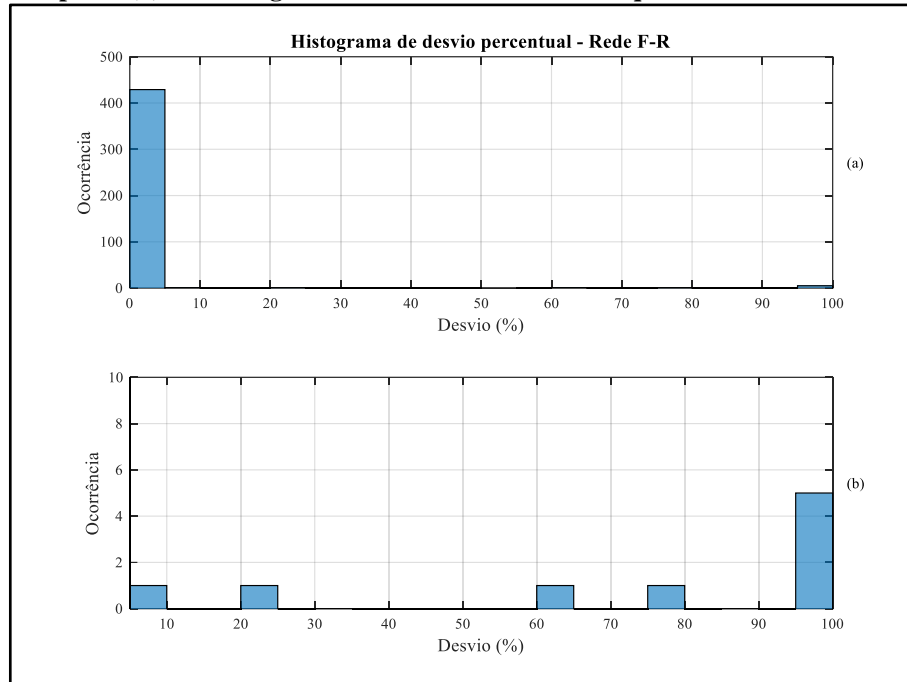
Para realização dos cálculos utilizou-se a Equação (4) apresentada a seguir, sendo esta aplicada para cada uma das amostras utilizadas no procedimento de teste. Tomando a média dos resultados obtidos em cada uma das amostras obtém-se o percentual médio de desvio para cada uma das redes que compõe o sistema RNA aqui proposto. Apresenta-se na Tabela 17 os percentuais obtidos durante os cálculos e nas Figuras 32 a 37, histogramas evidenciando o número de ocorrências para o percentual de desvio calculado.

$$Desvio\ Percentual = \frac{|Valor_{RNA} - Valor_{Real}|}{Limite_{Sup} - Limite_{Inf}} * 100(\%) \quad (4)$$

Tabela 18 – Erro percentual médio para cada RNA proposta	
Rede	Erro Percentual Médio (%)
F-R	1,5716
F-S	1,9422
F-T	0,9160
M-R	0,9256
M-S	0,4771
M-T	0,2165

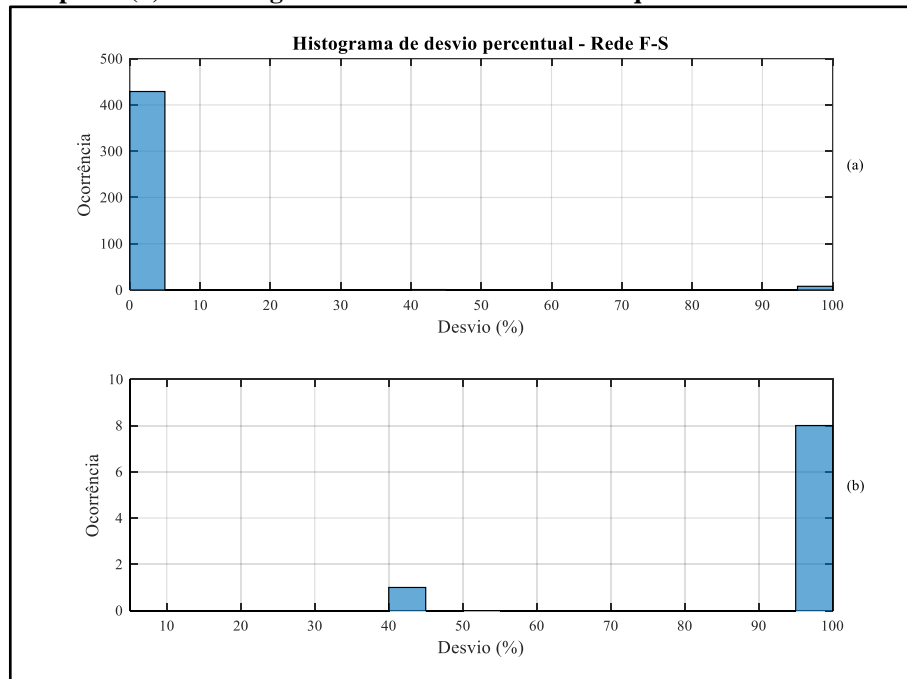
Fonte: Autoria Própria

Figura 26 – Histograma de desvio percentual para a rede F-R. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.



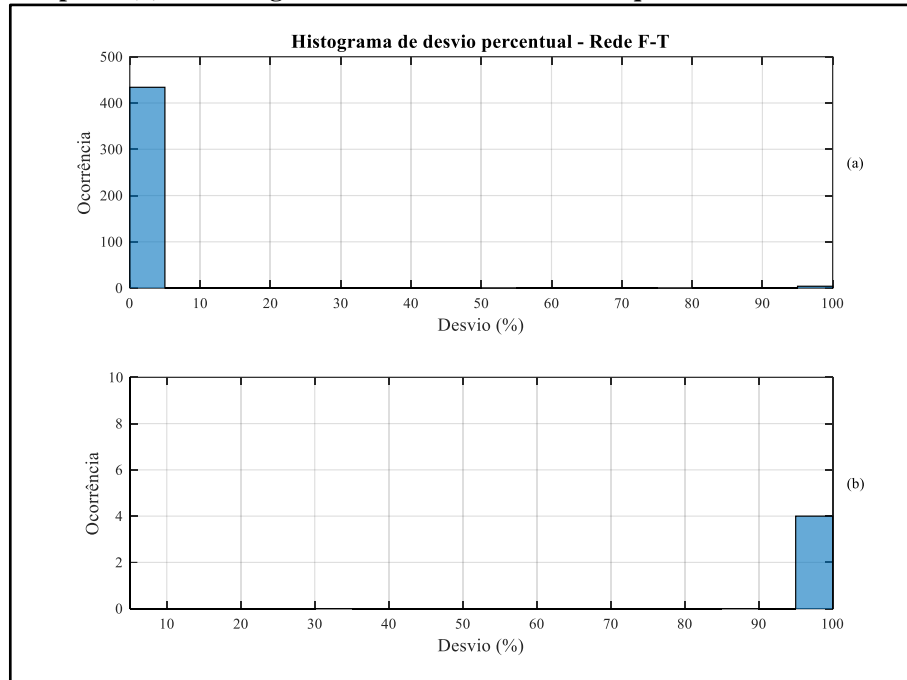
Fonte: Autoria Própria

Figura 27 – Histograma de desvio percentual para a rede F-S. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.



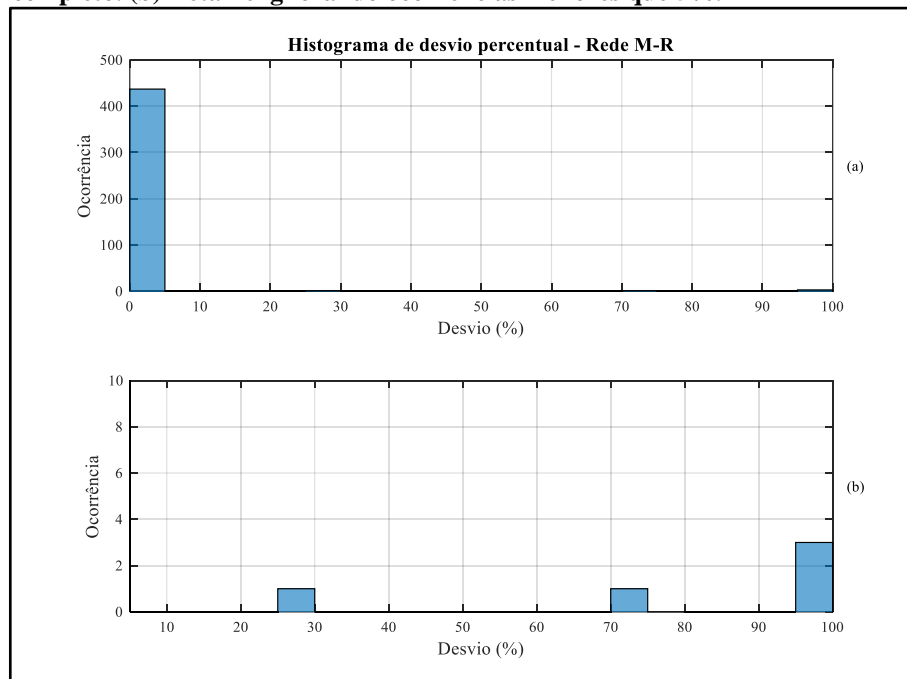
Fonte: Autoria Própria

Figura 28 – Histograma de desvio percentual para a rede F-T. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.



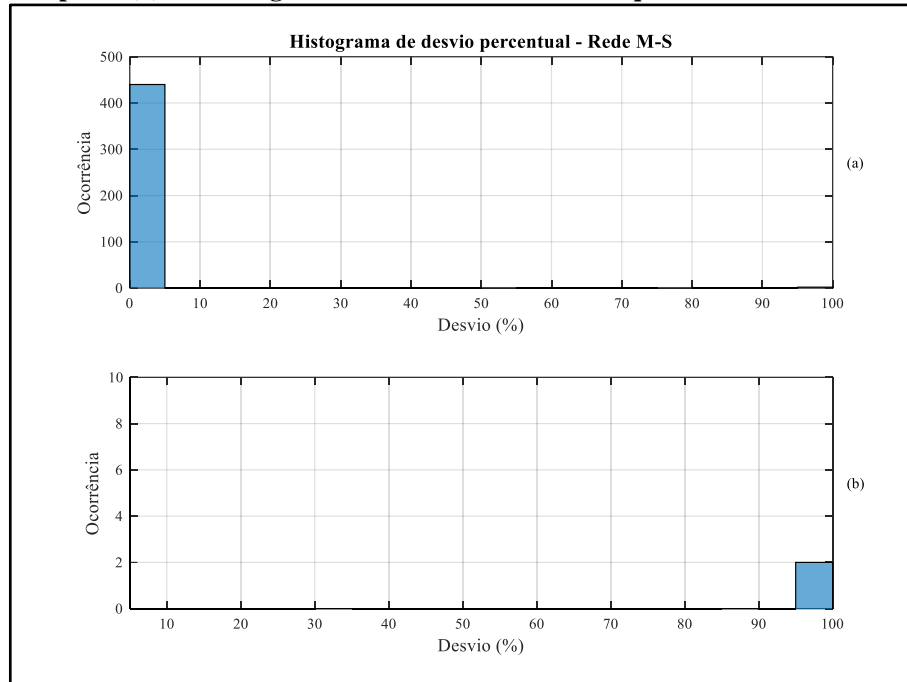
Fonte: Autoria Própria

Figura 29 – Histograma de desvio percentual para a rede M-R. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.



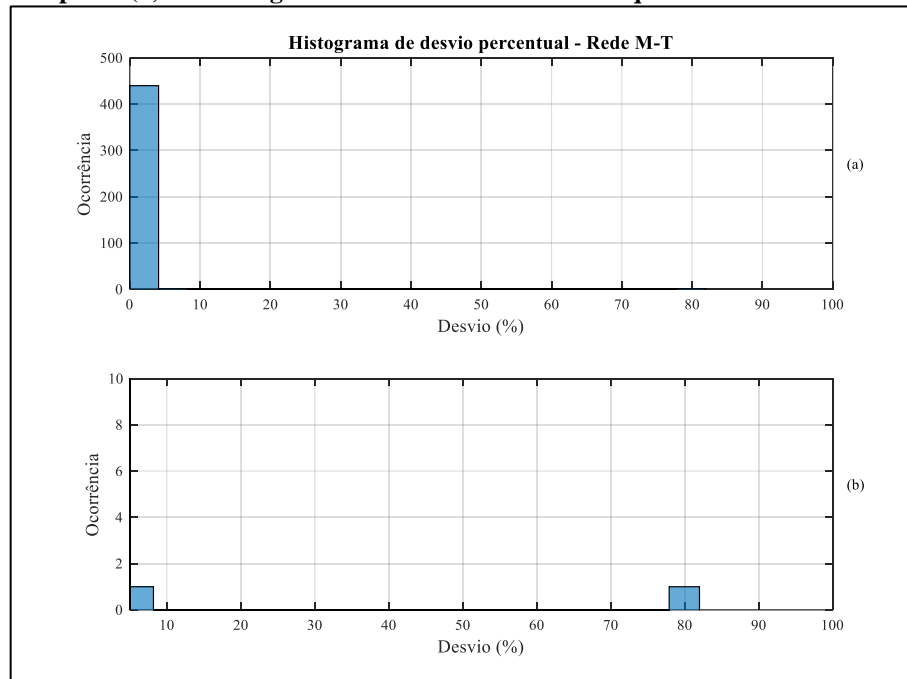
Fonte: Autoria Própria

Figura 30 – Histograma de desvio percentual para a rede M-S. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.



Fonte: Autoria Própria

Figura 31 – Histograma de desvio percentual para a rede M-T. (a) Histograma completo. (b) Detalhe ignorando ocorrências menores que 5%.



Fonte: Autoria Própria

5.6 MULTIPLICIDADE DE CLASSIFICAÇÃO DOS SEGMENTOS

A análise de multiplicidade de classificação traz grande contexto à análise geral do trabalho. Até o momento, os demais parâmetros foram estabelecidos por meio de considerações individuais para cada uma das redes. Há necessidade de verificar-se a integridade da capacidade de classificação das RNA em um contexto global, ou seja, verificar se um mesmo segmento recebe classificação positiva em mais de uma RNA dentro do sistema proposto, ou é classificada como não pertencente à nenhuma das classes em todas as RNAs.

Para tal análise, todos os segmentos de áudio foram apresentados a cada uma das RNAs, mapeando-se a classificação resultante e comparando posteriormente seus resultados. Cada RNA pode apresentar três possíveis respostas, sendo elas *sim(+1)* quando o segmento pertence à classe de especialidade da RNA, *não(-1)* para quando o segmento não pertence à classe de especialidade e *indefinido (valores intermediários)* para quando a resposta da RNA não permite classificar o segmento de maneira satisfatória.

Esta análise desenvolveu-se em duas etapas, na primeira apresentou-se arquivos de pacientes do sexo masculino para redes especialistas em arquivos do sexo masculino e arquivos femininos para redes femininas, monitorando a resposta de cada uma das 3 RNAs em cada um dos gêneros avaliados. Nesta primeira etapa há 10 possibilidades de classificação para um único segmento de áudio conforme apresentado na primeira coluna das Tabelas 18-21. A segunda etapa consiste em apresentar arquivos de gêneros trocados para as RNAs, ou seja, apresenta-se arquivos masculinos para RNAs especialistas em arquivos femininos e vice-versa, monitorando as classificações também de acordo com as possibilidades avaliadas na primeira etapa. Apresenta-se nas Tabelas 18-21 os resultados obtidos em cada uma das etapas realizadas:

Tabela 19 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio femininos – Etapa 01

Possibilidade	Percentual de incidência (%)
3 Sim	0
2 Sim / 1 Não	1,1416
2 Sim / 1 Indefinido	0
1 Sim / 1 Não / 1 Indefinido	0,6849
1 Sim / 2 Não	96,1187
1 Sim / 2 Indefinido	0
3 Não	1,5982
2 Não / 1 Indefinido	0,4566

Tabela 19 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio femininos – Etapa 01

Possibilidade	Percentual de incidência (%)
1 Não / 2 Indefinido	0
3 Indefinido	0

Fonte: Autoria Própria

Tabela 20 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio masculinos – Etapa 01

Possibilidade	Percentual de incidência (%)
3 Sim	0
2 Sim / 1 Não	0
2 Sim / 1 Indefinido	0
1 Sim / 1 Não / 1 Indefinido	0,4525
1 Sim / 2 Não	98,8688
1 Sim / 2 Indefinido	0
3 Não	0,2262
2 Não / 1 Indefinido	0,4525
1 Não / 2 Indefinido	0
3 Indefinido	0

Fonte: Autoria Própria

Tabela 21 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio femininos – Etapa 02

Possibilidade	Percentual de incidência (%)
3 Sim	0
2 Sim / 1 Não	0
2 Sim / 1 Indefinido	0
1 Sim / 1 Não / 1 Indefinido	0
1 Sim / 2 Não	0
1 Sim / 2 Indefinido	0
3 Não	100
2 Não / 1 Indefinido	0
1 Não / 2 Indefinido	0
3 Indefinido	0

Fonte: Autoria Própria

Tabela 22 – Percentuais de incidência conforme possibilidade de resposta para as RNAs dos segmentos de áudio masculinos – Etapa 02

Possibilidade	Percentual de incidência (%)
3 Sim	0
2 Sim / 1 Não	0
2 Sim / 1 Indefinido	0
1 Sim / 1 Não / 1 Indefinido	0
1 Sim / 2 Não	0
1 Sim / 2 Indefinido	0
3 Não	100
2 Não / 1 Indefinido	0
1 Não / 2 Indefinido	0
3 Indefinido	0

Fonte: Autoria Própria

Obteve-se para o procedimento realizado alto percentual para a situação ideal de operação do sistema RNA proposto, em que apenas uma das RNAs classifica positivamente um segmento de áudio.

5.6.1 Taxa de confiabilidade geral de classificação das RNAs

Como parâmetro final para este trabalho, considera-se uma taxa de confiabilidade geral para cada RNA, sendo tal taxa formada pelo produto entre o percentual de acerto de cada umas das RNAs durante o procedimento de teste e pelo percentual de classificação proveniente do sistema RNA, ou seja, leva-se em conta a probabilidade de um segmento de áudio receber apenas uma classificação positiva ao ser apresentado ao sistema com as 6 RNAs.

Matematicamente, a taxa de confiabilidade geral de classificação pode ser calculada conforme regra da multiplicação para probabilidades (MONTGOMERY e RUNGER, 2010, p.45) apresentada na Equação (5) e os resultados obtidos se encontram na Tabela 22:

$$T_{\text{confiabilidade Geral}} = T_{\text{acertos}} * T_{\text{Sistema_RNA}} \quad (5)$$

Tabela 23 – Percentual de confiabilidade geral para as 6 RNAs mais aptas à resolução do problema proposto

Rede	Taxa de Acertos	Sistema RNA	Confiabilidade Geral (%)
F-R	97,95	96,12	94,14
F-S	97,95	96,12	94,14
F-T	99,09	96,12	95,24
M-R	98,77	98,87	97,65
M-S	99,54	98,87	98,41
M-T	99,54	98,87	98,41
Média Geral	98,74	-	96,33

Fonte: Autoria Própria

6 CONSIDERAÇÕES FINAIS

A busca por um sistema capaz de classificar, por meio de parâmetros acústicos condições clínicas em sinais de voz, pode ser considerada uma grande jornada, composta por etapas que exigem grande dedicação e atenção durante todos os procedimentos. O objetivo central desta busca é desenvolver uma ferramenta capaz de auxiliar profissionais da voz em seus diagnósticos dando suporte para as observações subjetivas. Outras práticas podem ser beneficiadas pela utilização da ferramenta em questão, uma vez que ela também é capaz de: auxiliar o monitoramento terapêutico durante o desenvolvimento de tratamentos, facilitar o entendimento dos pacientes quanto a sua condição de saúde e contribuir para uma sólida formação de novos profissionais da voz.

De modo particular, este trabalho pode ser visto como os movimentos iniciais da jornada descrita anteriormente, uma vez que, por meio dos métodos aqui propostos é possível indicar a predominância subjetiva dos arquivos de áudio analisados, sendo esta a primeira triagem em muitas avaliações realizadas por profissionais da voz. Neste mesmo contexto, trabalhos pertencentes a esta mesma linha de pesquisa podem se utilizar da classificação resultante do método proposto neste trabalho para alimentar algum outro sistema estimador/classificador mais avançado. Deste modo a movimentação é mais efetiva e os resultados são desenvolvidos com maior facilidade.

O desenvolvimento deste trabalho contou com inúmeros procedimentos computacionais em que possibilidades de estudo levantadas foram testadas e seus resultados comparados, escolhendo-se aquelas com maior aptidão para resolução da proposta. A ideia inicial de trabalhar com RNAs compactas direcionou o estudo a buscar uma RNA especialista na identificação de cada um dos três parâmetros subjetivos dos arquivos de voz. Deste modo, trabalhando-se com um sistema de redes especialistas qualquer segmento de áudio que contenha predominâncias de um dos parâmetros subjetivos estudado, poderia ser classificada.

A separação, de acordo com o gênero, dos arquivos que compunham o banco de dados durante todos os procedimentos realizados, foram um dos responsáveis pelo aumento na taxa de acerto das RNAs. A diferença existente entre a frequência fundamental masculina e feminina certamente estava limitando a capacidade de generalização das RNAs e separá-las possibilitou aumentar a taxa de acerto com redes compactas, conforme utilizadas anteriormente. A identificação dos coeficientes com maior grau de correlação também possibilitou melhorias no resultado, uma vez que os treinamentos eram realizados de maneira mais ágil e eficiente.

Observando-se o ponto chave para elaboração deste trabalho – desenvolvimento de uma ferramenta capaz de auxiliar profissionais da voz em seus diagnósticos – e os resultados nele obtidos, torna-se oportuna a implementação do sistema RNA obtido em um meio de fácil acesso e manuseio. Assim, profissionais da voz poderão ser auxiliados por um sistema que possui média de confiabilidade de classificação maior que 96%.

Como mencionado em parágrafos anteriores, o desenvolvimento de uma ferramenta contempla várias etapas, sendo este trabalho apenas uma delas. Como passos futuros, podem-se citar a utilização dos resultados obtidos por meio deste trabalho para elaboração de um sistema mais avançado de classificação, desenvolvimento de *hardware* ou *software* capaz de tornar a utilização desta ferramenta mais intuitiva e busca de outros coeficientes acústicos que descrevem melhor as condições subjetivas dos arquivos estudados, a fim de aumentar a taxa de acerto obtida.

REFERÊNCIAS

ALMAADEED, Noor e AGGOUN, Amar e AMIRA, Abbas. **Speaker identification using multimodal neural networks and wavelet analysis**. IET Biometrics, v. 4, n. 1, p. 18–28, 2015.

ARAÚJO, Roberta de Paula. **Perfil dos beneficiários do INSS em auxílio-doença por distúrbios benignos da voz**. Revista Brasileira de Medicina do Trabalho, v. 12, n. 1, p. 13–19, 2014.

ARIAS-LONDONO, Julian D e colab. **Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients**. IEEE Transactions on Biomedical Engineering, v. 58, n. 2, p. 370–379, 2011.

BEHLAU, Mara. **Voz: o livro do especialista. v. 1**. Rio de Janeiro: Revinter, 2001.

BISHOP, Christopher M. **Pattern Recognition**. Machine Learning, 2006.

BRASIL. **Comitê Brasileiro Multidisciplinar de Voz Ocupacional**. . [S.l.: s.n.]. Disponível em: <http://www.sbfa.org.br/portal/pdf/boletimn1_COMVOZ.pdf>. , 2010

CARDING, Paul e colab. **Formal perceptual evaluation of voice quality in the United Kingdom**. Logopedics Phoniatrics Vocology, v. 25, n. 3, p. 133–138, 2000.

CASTELANO, Célio Ricardo. **Estudo comparativo da Transformada Wavelet no Reconhecimento de Padrões da íris humana**. . [S.l.]: Universidade de São Paulo. , 2006

CROVATO, César David Paredes e SCHUCK, Adalberto. **The use of wavelet packet transform and artificial neural networks in analysis and classification of dysphonic voices**. IEEE Transactions on Biomedical Engineering, v. 54, n. 10, p. 1898–1900, 2007.

DAJER, María Eugenia. **Análise de sinais de voz por padrões visuais de dinâmica vocal**. 2010. 154 f. Universidade de São Paulo, 2010.

DAJER, María Eugenia. **Padrões Visuais de Sinais de Voz Através de Técnica de Análise Não Linear**. 2006. 98 f. Universidade de São Paulo, São Carlos, 2006.

DUDA, Richard O e HART, Peter E e STORK, David G. **Pattern classification**. 2. ed. New Jersey: John Wiley & Sons, 2012.

FARRÚS, M. e HERNANDO, J. **Using Jitter and Shimmer in speaker verification**. IET Signal Processing, v. 3, n. November 2008, p. 247, 2009.

GOLDSCHMIDT, Ronaldo Ribeiro. **Uma Introdução à Inteligência Computacional: Fundamentos, ferramentas e aplicações**. Rio de Janeiro: IST-Rio, 2010. v. 53.

GUIMARÃES, Isabel. **Os problemas de voz nos professores: prevalência, causas, efeitos e formas de prevenção**. Revista Portuguesa de saúde publica, v. 22, n. 2, 2004.

HAYKIN, Simon S. **Redes neurais artificiais: princípio e prática**. 2. ed. São Paulo: Bookman, 2000.

HOLI, Mallikarjun S. Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks. 2013, [S.l.]: IEEE, 2013. p. 76–79.

HORNIK, Kurt e STINCHCOMBE, Maxwell e WHITE, Halbert. **Multilayer feedforward networks are universal approximators**. Neural networks, v. 2, n. 5, p. 359–366, 1989.

MARKAKI, Maria e STYLIANOU, Yannis. **Voice pathology detection and discrimination based on modulation spectral features**. IEEE Transactions on Audio, Speech, and Language Processing, v. 19, n. 7, p. 1938–1948, 2011.

MATHWORKS. **wentropy**. Disponível em:

<<https://www.mathworks.com/help/wavelet/ref/wentropy.html>>. Acesso em: 1 jun 2017.

MENDES, Ana P e FERREIRA, Lénia J L e CASTRO, Eugénia. **Softwares e hardwares de análise acústica da voz e da fala**. Distúrbios da Comunicação. ISSN 2176-2724, v. 24, n. 3, 2012.

MONTGOMERY, Douglas C e RUNGER, George C. **Applied statistics and probability for engineers**. [S.l.]: John Wiley & Sons, 2010.

MUSIAL, Patrícia L e colab. **Interferência dos sintomas vocais na atuação profissional de professores**. *Distúrbios da Comunicação*. ISSN 2176-2724, v. 23, n. 3, 2011.

OATES, Jennifer. **Auditory-perceptual evaluation of disordered voice quality**: *Folia Phoniatica et Logopaedica*, v. 61, n. 1, p. 49–56, 2009.

OLIVEIRA, H M. **Análise de fourier e wavelets: sinais estacionários e não estacionários**. Editora Universitária da UFPE, 2007.

PINHO, Sílvia M. Rabelo e TSUJI, Domingos Hiroshi e BOHADANA, Saramira C. **Fundamentos em Laringologia e Voz**. Rio de Janeiro: Revinter, 2006. v. 1.

RABINER, Lawrence R. e SCHAFER, Ronald W. **Theory and application of digital signal processing**. 1. ed. Englewood Cliffs: Prentice-Hall, 2011. v. 1.

ROSA, Marcelo de Oliveira. **Análise Acústica da Voz para Pré-Diagnóstico de Patologias da Laringe**. 1998. 261 f. 1998.

SANTOS, Mikaelle Oliveira. **Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet**. 2015.

SCOTT-PHILLIPS, Thomas C. **The evolution of communication: Humans may be exceptional**. *Interaction Studies*, v. 11, n. 1, p. 78–99, 2010.

SILVA, I N Da e SPATTI, Danilo Hernane e FLAUZINO, Rogério Andrade. **Redes neurais artificiais para engenharia e ciências aplicadas**. [S.l.: s.n.], 2010.

SIMBERG, Susanna. **Prevalence of Voice Disorders Among Future Teachers**. v. 14, n. 2, p. 231–235, 2000.

SOUISSI, Nawel e CHERIF, Adnane. Speech recognition system based on short-term cepstral parameters, feature reduction method and Artificial Neural Networks. 2016, [S.l.]: IEEE, 2016. p. 667–671.

TEIXEIRA, João Paulo e FERREIRA, Débora e CARNEIRO, Susana Moreira. Análise acústica vocal-determinação do Jitter e Shimmer para diagnóstico de patologias da fala. 2011, [S.l.]: INEGI, 2011.

TITZE, Ingo R. **Princípios Da Produção Vocal**. [S.l.]: National Center for Voice & Speech, 2013. Disponível em: <<https://books.google.com.br/books?id=papFnwEACAAJ>>.

TORTORA, Gerard. **Princípios de anatomia e fisiologia**. 4. ed. Porto Alegre: Artmed, 2000.

VILLA-CANAS, T e colab. **Automatic assessment of voice signals according to the GRBAS scale using modulation spectra, Mel frequency Cepstral Coefficients and Noise parameters**. 2013, [S.l.]: IEEE, 2013. p. 1–5.

WANG, Xiang e ZHANG, Jianping e YAN, Yonghong. **Discrimination between pathological and normal voices using GMM-SVM approach**. *Journal of Voice*, v. 25, n. 1, p. 38–43, 2011. Disponível em: <<http://dx.doi.org/10.1016/j.jvoice.2009.08.002>>.