

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

ANDRIÉLI HILÁRIO BARIZÃO

**ESTIMAÇÃO DO GRAU DE PARÂMETROS SUBJETIVOS
VOCAIS APLICANDO REDES NEURAS ARTIFICIAIS**

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO
2017

ANDRIÉLI HILÁRIO BARIZÃO

**ESTIMAÇÃO DO GRAU DE PARÂMETROS SUBJETIVOS
VOCAIS APLICANDO REDES NEURAS ARTIFICIAIS**

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

Orientadora: Prof. 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

Andriéli Hilário Barizão

Estimação do grau de parâmetros subjetivos aplicando redes neurais artificiais

Trabalho de conclusão de curso apresentado às 17:30hs do dia 20/06/2017 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). Maria Eugenia Dajer - Presidente (Orientador)

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

Prof(a). Dr(a). Bruno Catarino Bispo - (Membro)

Prof(a). Dr(a). Vander Teixeira Prado - (Membro)

A folha de aprovação assinada encontra-se na coordenação do curso.

Dedico este trabalho à minha família e a todos que me ajudaram nesta jornada e acreditaram em mim.

AGRADECIMENTOS

Obrigada a Deus e Nossa Senhora em primeiro lugar, de onde provém minha fé, combustível que me alimenta, para conseguir seguir por toda trajetória de vida um dia sonhada por mim.

Agradeço aos meus queridos orientadores, María Eugenia Dajer e Danilo Hernane Spatti, por toda a dedicação ao me conduzir para a luz da sabedoria.

Minha gratidão também a toda minha família, em especial, meus avôs e avós, a minha mãe Aparecida Fátima Hilário Barizão, meu pai Aparecido Barizão, que já não está entre nós, minha irmã Andressa Hilário Barizão, minha tia Nilza Hilário, meu grande amigo Cezar Botelho, por toda a base que possibilita meu crescimento como pessoa.

Ademais, não posso deixar de mencionar meu grande amigo Murillo Augusto Fermino, que contribuiu de maneira tão significativa para a conclusão desta fase de minha vida e todos os outros, em especial Gabriel Rodrigues de Amorim Neto.

Por fim, todo o respeito e reconhecimento para Bruno Catarino Bispo e Vander Teixeira Prado, por fomentar o avanço científico.

“A razão é o passo, o aumento da ciência o caminho, e o benefício da humanidade é o fim.”

Thomas Hobbes

RESUMO

BARIZÃO, Andriéli Hilário. **Estimação do grau de parâmetros subjetivos vocais aplicando redes neurais artificiais**. 2017. 65 f. Trabalho de Conclusão de Curso – Engenharia Elétrica. Universidade Tecnológica Federal do Paraná. Cornélio Procopio, 2017.

Sob a visão psicossocial e cultural, a voz traz consigo a necessidade da comunicação humana, carregando a identidade de cada indivíduo, podendo ser tão pessoal quanto as impressões digitais, sendo ela uma das expressões mais fortes da personalidade humana. Portanto, a eminência de qualquer distúrbio vocal, pode ocasionar vários agravantes para a pessoa em questão. Este trabalho engloba tanto a área da engenharia quanto fonoaudiologia, trazendo uma ferramenta computacional, elaborada por meio de redes neurais artificiais e transformada *wavelet packet* capaz de auxiliar os profissionais que estudam os distúrbios vocais por meio de um método não invasivo, como a avaliação perceptivo-auditiva, que utiliza da percepção dos avaliadores para classificarem parâmetros subjetivos da voz, tais como a rugosidade, sopro e tensão. A ferramenta computacional desenvolvida auxiliará tanto os avaliadores se autoanalisarem em suas percepções audíveis, como também diminuirá o tempo hábil das avaliações, trazendo a classificação da predominância dos parâmetros subjetivos da escala visual analógica, assim como a estimativa desses parâmetros em porcentagens. Tal sistema computacional elaborado mostrou-se eficiente na análise dos resultados, apresentando taxas superiores a 95% de acertos na classificação e estimação dos parâmetros subjetivos abordados no estudo realizado.

Palavras-chave: Distúrbio vocal. Parâmetros subjetivos de Voz. Ferramenta computacional. Redes neurais artificiais. Transformada *wavelet packet*.

ABSTRACT

BARIZÃO, Andriéli Hilário. **Estimation of the degree of subjective vocal parameters applying artificial neural networks.** 2017. 65 f. Trabalho de Conclusão de Curso – Engenharia Elétrica. Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2017.

Under the psychosocial and cultural view, the voice brings with it the need for human communication, carrying the identity of each individual, being as personal as fingerprints, being one of the strongest expressions of the human personality. Therefore, the eminence of any vocal disorder can cause several aggravating for the person in question. This work encompasses both the field of engineering and speech therapy, bringing a computational tool, developed through artificial neural networks and wavelet packet transform that is capable of assisting professionals who study vocal disorders through a non-invasive method, such as perceptual-auditory, which uses the perception of the evaluators to classify subjective parameters of the voice, such as roughness, breathiness and tension. The computational tool developed will help both the evaluators to analyze themselves in their audible perceptions, as well as to shorten the evaluation time, bringing the classification of the predominance of the subjective parameters of the analogue visual scale, as well as the estimation of these parameters in percentages. Such elaborated computational model proved to be efficient in the analysis of the results, presenting precision rates above 95% in the classification and estimation of the subjective parameters addressed in the study.

Keywords: Vocal disturbance. Subjective Parameters of Voice. Computational tool. Artificial neural networks. Wavelet packet transform.

LISTA DE FIGURAS

Figura 1 – Videolaringoestroboscopia.	17
Figura 2 – Escala CAPE-V.	21
Figura 3 – Diagrama de blocos da produção da voz humana.	24
Figura 4 – Decomposição wavelet discreta multiresolução.	27
Figura 5 – Decomposição da transformada wavelet packet em árvore binária.	28
Figura 6 – Representação do neurônio artificial.	30
Figura 7 – Exemplo de rede feedforward de camada múltiplas.	31
Figura 8 – Sinal vocal predominantemente rugoso – H R.wav.	35
Figura 9 – Sinal vocal predominantemente soproso – H S.wav.	36
Figura 10 – Sinal vocal predominantemente tenso – H T.wav.	36
Figura 11 – Exemplificação da matriz de janelamento.	38
Figura 12 – Exemplo de sinal original reamostrado e normalizado.	38
Figura 13 – Janelamento simples do sinal.	39
Figura 14 – Matriz dos dados de entradas das RNAs.	40
Figura 15 – RNA especialista em rugosidade para áudio rugoso.	51
Figura 16 – RNA especialista em soproso para áudio rugoso.	52
Figura 17 – RNA especialista em tensão para áudio rugoso.	53
Figura 18 – RNA especialista em soproso para áudio soproso.	54
Figura 19 – RNA especialista em rugosidade para áudio soproso.	55
Figura 20 – RNA especialista em tensão para áudio soproso.	56
Figura 21 – RNA especialista em tensão para áudio tenso.	57
Figura 22 – RNA especialista em soproso para áudio tenso.	58
Figura 23 – RNA especialista em rugosidade para áudio tenso.	59

LISTA DE QUADROS

Quadro 1 - Principais funções de ativação.....	32
Quadro 2 – Amostragem do banco de dados.....	34
Quadro 3 – Relação dos arquivos independentes do gênero.....	34
Quadro 4 – Arquivos para treinamento e validação.	34
Quadro 5 – Características da RNA aplicadas	41
Quadro 6 – Melhor topologia encontrada (rugosidade – uma camada)	43
Quadro 7 – Resultado dos treinamentos (rugosidade – uma camada)	44
Quadro 8 – Melhor topologia encontrada (rugosidade – duas camadas).....	44
Quadro 9 – Resultado dos treinamentos (rugosidade – duas camadas).....	44
Quadro 10 – Melhor topologia encontrada (rugosidade – três camadas).....	45
Quadro 11 – Resultado dos treinamentos (rugosidade – três camadas)	45
Quadro 12 – Melhor topologia encontrada (soprosidade – uma camada).....	46
Quadro 13 – Resultado dos treinamentos (soprosidade – uma camada)	46
Quadro 14 – Melhor topologia encontrada (soprosidade – duas camadas).....	46
Quadro 15 – Resultado dos treinamentos (soprosidade – duas camadas).....	47
Quadro 16 – Melhor topologia encontrada (soprosidade – três camadas).....	47
Quadro 17 – Resultado dos treinamentos (soprosidade – três camadas).....	47
Quadro 18 – Melhor topologia encontrada (tensão – uma camada)	48
Quadro 19 – Resultado dos treinamentos (tensão – uma camada)	48
Quadro 20 – Melhor topologia encontrada (tensão – duas camadas).....	48
Quadro 21 – Resultado dos treinamentos (tensão – duas camadas).....	49
Quadro 22 – Melhor topologia encontrada (tensão – três camadas).....	49
Quadro 23 – Resultado dos treinamentos (tensão – três camadas)	49
Quadro 24 – Melhores topologias para cada parâmetro estudado.....	50
Quadro 25 – RNA especialista em rugosidade para áudio rugoso.....	51
Quadro 26 – RNA especialista em soprosidade para um áudio rugoso	52
Quadro 27 – RNA especialista em tensão para áudio rugoso.....	53
Quadro 28 – RNA especialista em soprosidade para áudio soproso.	54
Quadro 29 – RNA especialista em rugosidade para áudio soproso.	55
Quadro 30 – RNA especialista em tensão para áudio soproso.....	56
Quadro 31 – RNA especialista em tensão para áudio tenso.	57
Quadro 32 – RNA especialista em soprosidade para áudio tenso.	58
Quadro 33 – RNA especialista em rugosidade para áudio tenso.	59
Quadro 34 – Estimação dos parâmetros para um áudio rugoso.	60
Quadro 35 – Estimação dos parâmetros para um áudio soproso.	60
Quadro 36 – Estimação dos parâmetros para um áudio tenso.	60

SUMÁRIO

1	INTRODUÇÃO	12
1.1	CONTEXTUALIZAÇÃO	12
1.2	MOTIVAÇÃO	14
1.3	OBJETIVOS	14
1.3.1	Objetivo geral	14
1.3.2	Objetivos específicos	15
2	A ANÁLISE PERCEPTIVO-AUDITIVA E ACÚSTICA	16
2.1	FISIOLOGIA DA VOZ	16
2.2	A AVALIAÇÃO PERCEPTIVO-AUDITIVA	18
2.2.1	Escala perceptivo-auditiva CAPE-V	19
2.2.2	Avaliadores da escala perceptivo-auditiva	22
2.3	A AVALIAÇÃO ACÚSTICA	23
2.4	IMPLICAÇÕES CLÍNICAS	25
3	TRANSFORMADA WAVELET E REDES NEURAS ARTIFICIAIS	26
3.1	TRANSFORMADA WAVELET DISCRETA	26
3.2	TRANSFORMADA <i>WAVELET PACKET</i>	27
3.3	REDES NEURAS ARTIFICIAIS (RNAs)	28
3.3.1	Perceptron multicamadas	31
4	MATERIAIS E MÉTODOS	33
4.1	BANCO DE DADOS	33
4.2	PRÉ-PROCESSAMENTO DO BANCO DE DADOS	34
4.3	NORMALIZAÇÃO DAS AMOSTRAS DE ÁUDIOS	37
4.4	JANELAMENTO DAS AMOSTRAS DE ÁUDIOS	37
4.5	EXTRAÇÃO DE CARACTERÍSTICAS UTILIZANDO WPT	39
4.6	CARACTERÍSTICAS DAS RNAs E PROCESSO DE TREINAMENTO	41
5	RESULTADOS	43
5.1	MELHORES TOPOLOGIAS	43
5.2	MELHORES REDES NEURAS ARTIFICIAIS	50
5.3	VALIDAÇÃO DA GENERALIZAÇÃO DAS RNAs	50
5.3.1	Análise de áudio rugoso	51
5.3.2	Análise de áudio soproso	53
5.3.3	Análise de áudio tenso	57
5.3.4	Discussão das análises realizadas	60
6	CONSIDERAÇÕES FINAIS	61
	REFERÊNCIAS	63

1 INTRODUÇÃO

Neste capítulo serão descritos os fundamentos para o estudo da voz, desde os aspectos fisiológicos até o processamento do sinal.

1.1 CONTEXTUALIZAÇÃO

Sob a visão psicossocial e cultural, a voz traz consigo a necessidade da comunicação humana e carrega a identidade de cada um. Pode-se dizer que a voz possui um caráter único para cada indivíduo, sendo ela uma das expressões mais fortes da personalidade humana, podendo ser tão pessoal quanto as impressões digitais (CUERVO, 2010; SANTOS, 2015).

A voz humana é produzida pela interação de um conjunto de órgãos que formam o aparelho fonador: pulmões, brônquios, traqueia, músculos, ligamentos e ossos. O sistema de produção vocal possui uma capacidade de produção complexa e potente. Por meio do fluxo de ar expelido pelos pulmões que passa pelas pregas vocais, fenômeno que comporta grandes variações que necessita de uma complexa e interdependente atividade de todos os músculos e integridade dos tecidos do aparelho fonador. Sua representação está centralizada nas pregas vocais, de onde origina-se o sinal de voz (BEHLAU, 2001; CUERVO, 2010).

Os distúrbios vocais e seus agravantes podem comprometer a qualidade de vida dos indivíduos, resultando em restrições no campo emocional, social e psicológico. Alguns hábitos são prejudiciais, tais como: tabagismo, etilismo, hidratação insuficiente, estresse e automedicação; além das condições desfavoráveis nas quais as pessoas estão inseridas, como, ambientes climatizados de forma incorreta, ausência de recursos de reprodução vocal em situações necessárias, entre outras que muitas vezes passam despercebidos e podem ser responsáveis pelo surgimento de vários distúrbios vocais. Caso os distúrbios sejam detectados em estágios mais avançados, são necessários procedimentos mais complexos para a reabilitação do paciente (HOGIKYAN; SETHURAMAN, 1999; ARAÚJO, 2014).

O desvio vocal é definido como um distúrbio da comunicação oral, em que a voz não é capaz de transmitir a mensagem, tanto verbal como emocional, representando uma dificuldade ou alteração na emissão, condicionado a sua produção. Como tal, é compreensível que seja difícil usar um único método que avalie, de forma abrangente e precisa, a qualidade vocal ou a sua deterioração (FREITAS, 2010).

A avaliação fonoaudiológica da voz é um instrumento eficaz para descrever a qualidade vocal do indivíduo. O conceito de qualidade vocal está diretamente relacionado com preceitos fisiológicos, perceptivos e acústicos. Os métodos de avaliações mais utilizados na rotina da clínica vocal são as análises acústica e a perceptivo-auditiva da voz. A análise acústica baseia-se na linearidade do modelo fonte-filtro para a produção de vogais. Esta é uma das mais importantes limitações deste método de avaliação, pelo fato de assumir a independência da fonte e do filtro. Por sua vez, a análise perceptivo-auditiva é uma avaliação subjetiva fundamentada na impressão auditiva que o avaliador tem sobre a emissão vocal do indivíduo, permitindo a inferência de dados anatomofisiológicos importantes, fornecendo informações sobre os aspectos psicossociais da voz (ZITTA, 2005; FREITAS, 2010; GAMA; COUTO; MARTINS, 2015).

Como alternativa aos métodos convencionais de análise e visando contornar as dificuldades da análise acústica, esse estudo tem como objetivo estimar o grau dos parâmetros subjetivos de rugosidade, soprosidade e tensão do sinal de voz com o intuito de auxiliar os profissionais na avaliação diagnóstica perceptivo-auditiva. Por meio da utilização de redes neurais artificiais (RNAs) e aplicação da transformada *wavelet packet* (WPT) como extrator de características será implementado uma ferramenta computacional capaz de estimar tais parâmetros subjetivos, tendo como base de dados, áudios de pacientes que sustentam a vogal /e/.

As RNAs são modelos computacionais inspirados no sistema nervoso de seres vivos. Possuem a capacidade de aquisição e manutenção do conhecimento (fundamentado em informações) e podem ser definidas como um conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligados por um grande número de interconexões (sinapses artificiais), sendo as mesmas representadas por vetores/matrizes de pesos sinápticos. Assim, podem ser

empregadas em diversos problemas relacionados às engenharias e ciências (DA SILVA; SPATTI; FLAUZINO, 2010).

A transformada *wavelet packet* fornece uma análise do sinal em diferentes resoluções, possibilitando nesse estudo a observação da energia associada às faixas de frequência dos seus diferentes níveis, podendo assim apontar uma desordem vocal (SANTOS, 2015).

1.2 MOTIVAÇÃO

A motivação para a realização deste trabalho é apontar a necessidade de integração entre a engenharia e a fonoaudiologia, especialidade médica que compreende o estudo da fonação e da audição, de seus distúrbios e das suas formas de tratamento, proporcionando uma ferramenta computacional no auxílio de procedimentos mais instantâneos, na avaliação perceptivo-auditiva.

Por último, prevê-se a possibilidade de criar uma ferramenta computacional capaz de classificar e quantificar os parâmetros subjetivos que serão analisados.

1.3 OBJETIVOS

Nessa seção serão destacados os objetivos referentes a proposta do trabalho.

1.3.1 Objetivo geral

Esta proposta de trabalho traz como objetivo desenvolver e aplicar RNAs capazes de classificar e estimar o grau dos parâmetros subjetivos: rugosidade, soprosidade e tensão de sinais vocais, a fim de auxiliar a avaliação perceptivo-auditiva realizada por profissionais da voz.

1.3.2 Objetivos específicos

- Efetivar o pré-processamento do sinal de voz contendo a vogal sustentada /e/;
- Abordar métodos de extração de característica utilizando a transformada *wavelet packet*;
- Construir o conjunto de entrada das redes neurais a partir da característica extraída;
- Encontrar a melhor topologia que se enquadre nas limitações do trabalho, para o problema por meio de treinamentos automatizados das RNAs especialistas;
- Validar o código para amostras desconhecidas na avaliação da generalização da rede;
- Analisar estatisticamente a acurácia do estimador com dados do classificador.

2 A ANÁLISE PERCEPTIVO-AUDITIVA E ACÚSTICA

Neste capítulo será evidenciado o método utilizado para avaliação perceptivo-auditiva e acústica.

2.1 FISILOGIA DA VOZ

A análise da voz implica em uma complexa combinação de dados, que permite a construção de um quadro conceitual, possibilitando explicações do “como” e “porquê” de um distúrbio vocal apresentado. O uso do termo voz disfônica engloba uma série de sintomas, tais como: desvios na qualidade vocal, esforço na emissão do som, fonoastenia (fadiga), perda de potência/projeção vocal (diminuição do volume), variações descontroladas da frequência fundamental, quebras da intensidade vocal, baixa resistência vocal e sensações desagradáveis na emissão. Contudo, as perturbações vocais podem ser reconhecidas quando: a altura tonal e a sensação de intensidade e/ou a qualidade vocal são desagradáveis ou inadequadas para a idade e sexo do indivíduo, acusando desconforto ou dor ao falar, ou se causa alguma preocupação ou limitação ao mesmo (FREITAS, 2010).

Deste modo, Behlau (2001) discorre sobre a dificuldade em definir as distinções entre uma voz normal e disfônica, caracterizando as disfunções como:

- **Orgânicas:** Independem do uso da voz e podem ser causadas por uma série variada de processos: tumores benignos e malignos; traumatismos por uso de arma branca ou de fogo; doenças neurológicas, entre outras.
- **Organofuncionais:** Relacionadas ao funcionamento da voz com lesões; representam uma etapa posterior na evolução de uma disfonia funcional. Assim sendo, caracteriza-se pelo seu diagnóstico tardio e/ou atraso na busca da solução do problema pelo próprio paciente. As lesões organofuncionais mais comuns englobam nódulos, pólipos e

edemas de Reinke e são diagnosticadas por meio de exame laringoscópico.

A videolaringoestroboscopia, exame clínico representado pela Figura 1, permite a visualização do comportamento vibratório das pregas vocais através da utilização de uma microcâmera, e a eletromiografia é um método de registro dos potenciais elétricos gerados nas fibras musculares em ação. Essas técnicas visuais resultam em uma avaliação qualitativa, de resultados difíceis de serem quantificados, e necessitam do conhecimento e da experiência do avaliador (SANTOS, 2015).

Figura 1 – Videolaringoestroboscopia.



Fonte: Laringoestroboscopia Laríngea. InstitutoOtosur.

A produção da voz normal necessita de apurado controle neuromuscular e integridade das estruturas laríngeas, e a maior parte dos casos de disfonias organofuncionais (decorrentes de mau uso ou abuso vocal) podem ser solucionadas apenas com a fonoterapia. Muitas vezes, o quadro orgânico permanece após a reabilitação, mas ocorrem compensações funcionais, tais como: aumento do fluxo aéreo respiratório durante a fonação, maior eficiência da adução glótica e melhor

projeção vocal, satisfazendo as necessidades de comunicação do paciente. (PINHO; TSUJI; BOHADANA, 2006).

Técnicas de processamento digital de sinais estão sendo desenvolvidas para avaliar a qualidade vocal, bem como avaliar quantitativamente a intensidade do desvio vocal (rugosidade, sopro, tensão e instabilidade) por meio da análise acústica, com o objetivo de auxiliar o diagnóstico de forma menos invasivas e com menor custo, comparadas àquelas baseadas em exames videolaringoscópicos. Desta forma, para minimizar as variações decorrentes da subjetividade, foram criadas e validadas escalas de avaliação, entre as quais estão as escalas numérica (EN) e visual analógica (EVA) (BEHLAU et al, 2015).

Os métodos não invasivos de avaliação vocal podem ser divididos em dois grupos básicos: avaliação acústica e avaliação perceptivo-auditiva. Ambos métodos são complementares à avaliação visual da laringe, realizada pelo médico otorrinolaringologista para um diagnóstico mais completo sobre a situação do paciente (SANTOS, 2015).

2.2 A AVALIAÇÃO PERCEPTIVO-AUDITIVA

A análise perceptivo-auditiva é uma avaliação não invasiva amplamente usada na clínica fonoaudiológica, na qual o profissional julga uma amostra vocal produzida pelo falante. Geralmente, as provas pedidas implicam a sustentação de vogais, a fala encadeada, entre outras, e ficam registradas em áudio e/ou vídeo (preferencialmente) para posterior análise e suporte demonstrativo (ao paciente) e comparativo com avaliações futuras (FREITAS, 2010).

Por outro lado, a avaliação possui caráter essencialmente subjetivo, portanto, sujeita a erros e variações de avaliador para avaliador. Este tipo de avaliação é influenciado por diversos fatores, como a experiência e habilidade prévia do avaliador, amostra do material, dimensão qualitativa perceptual e quantitativa da escala. (Yu P et al, 2002).

2.2.1 Escala perceptivo-auditiva CAPE-V

A avaliação perceptivo-auditiva utiliza um protocolo de avaliação, denominado CAPE-V (Consensus Auditory Perceptual Evaluation of Voice), que quantifica seis parâmetros subjetivos vocais (Severidade Global, Rugosidade, Soprosidade, Tensão, Pitch e Loudness) por meio de uma escala visual analógica de 100mm. O protocolo também possibilita a avaliação de dois parâmetros adicionais, como ressonância e tremor, e possui dois descritores complementares (consistente ou intermitente), para cada parâmetro avaliado, demonstrado pela Figura 2 (FREITAS, 2010).

Baseia-se em avaliações de vogais sustentadas, leitura de frases pré-definidas e conversação espontânea. As instruções para o seu uso e consideração podem ser descritas da seguinte forma:

- **Grau Geral:** impressão global integrada de desvio de vocal.
- **DI:** Grau referente a um desvio discreto.
- **MO:** Grau referente a um desvio moderado.
- **AC:** Grau referente a um desvio acentuado.
- **Rugosidade:** Acréscimo de tremulações na voz. Exemplo característico de rugosidade ocorre nos nódulos vocais, hiperemias e edemas. Essa situação gera ruídos aditivos de baixa frequência, nos diagramas espectrográficos a presença de ruído mascara os harmônicos inferiores (ZITTA, 2005).
- **Soprosidade:** corresponde à presença de ruído de fundo audível, devido ao escape de ar no fechamento das pregas vocais (ZITTA, 2005).
- **Tensão:** está associada ao aumento do esforço vocal para vibração das pregas vocais e está geralmente relacionado ao aumento da atividade da musculatura extrínseca da laringe. Esta característica é caracterizada por

ruído nas frequências altas do espectro e harmônicos agudos marcados (ZITTA, 2005).

- **Pitch:** correlato perceptivo-auditivo da frequência fundamental. A taxa de escala do indivíduo desvia da normalidade de acordo com o sexo e idade da pessoa em questão. A direção de desvio (alta ou baixa) deve ser indicado no espaço em branco proporcionado acima da escala (FREITAS, 2010).
- **Loudness:** correlato perceptivo da intensidade sonora. Esta escala indica se o indivíduo desvia do normal de acordo com o sexo e idade. A direção de desvio deve ser indicada no espaço em branco proporcionado acima da escala (FREITAS, 2010).

Para esta escala, o clínico indica o grau de desvio percebido em cada parâmetro, utilizando uma marca de tique. O profissional pode colocar marcas em qualquer local ao longo da linha. Pontuações são fundamentadas em observações diretas do clínico e sua habilidade audível para com a vocalização do paciente durante a avaliação. O avaliador deve também circundar a letra que melhor descreve a consistência do parâmetro considerado. Um julgamento de "consistente" indica que o atributo está continuamente presente em todas as tarefas. Uma ponderação de "intermitente" indica que o atributo ocorreu de forma inconsistente dentro ou entre tarefas.

Ao concluir todas as classificações, o profissional deve medir com uma régua a distância em milímetros (mm) entre sua marca de tique e a extremidade esquerda da escala. A pontuação em mm deve ser escrita no espaço em branco, assim relacionando os resultados em uma proporção ao comprimento total de 100 mm da linha. Os resultados podem ser relatados através da indicação da distância em mm para descrever o grau de desvio, por exemplo "73/100".

Figura 2 – Escala CAPE-V.

**PROTOCOLO - CONSENSO DA AVALIAÇÃO PERCEPTIVO AUDITIVA DA VOZ
(CAPE-V) – ASHA 2003, SID3**

Nome _____ Data: _____

Os parâmetros da qualidade vocal deverão ser preenchidos conforme as seguintes tarefas:

- 1) Vogal sustentada com 3 a 5 segundos
- 2) Produção das seguintes sentenças:

a) Érica tomou suco de pêra e amora.	d) Agora é hora de acabar.
b) Sonia sabe sambar sozinha.	e) Minha mãe namorou um anjo
c) Olha lá o avião azul.	f) Papai trouxe pipoca quente.
- 3) Fala espontânea, com os seguintes conteúdos: "Fale-me sobre o seu problema de voz" ou "Diga-me como está a sua voz".

Legenda: C = consistente I = Intermitente

GRAU GERAL	DI	MO	AC	SCORE C I /100
RUGOSIDADE				C I /100
SOPROSIDADE				C I /100
TENSÃO				C I /100
<i>PITCH</i>	indique a natureza de desvio de <i>pitch</i> _____			C I /100
<i>LOUDNESS</i>	indique a natureza de desvio de <i>loudness</i> _____			C I /100
_____				C I /100
_____				C I /100

Comentários sobre ressonância: NORMAL OUTRA (descreva): _____

Características adicionais (por exemplo: diplofonia, som basal, falsete, astenia, afonia, instabilidade de frequência, tremor, qualidade molhada ou outras observações relevantes)

Clínico: _____

2.2.2 Avaliadores da escala perceptivo-auditiva

A constituição de grupos de avaliadores para a análise perceptivo-auditiva visa reunir um conjunto próximo de opiniões, que reduzam a subjetividade e minimizem alguns erros, que Freitas (2010) descreve como:

- **Erros de medida por fatores associados aos avaliadores:** estes fatores incluem as referências internas do avaliador relativamente à qualidade vocal a ser percebida, aos seus hábitos perceptivos e erros/distorções próprias e a sensibilidade geral face à qualidade de vozes a escutar. O treino e a exposição a uma vasta gama de áudios de alterações vocais ajudam a modelar estes fatores. Nesta categoria estão ainda incluídos o cansaço do avaliador, os lapsos de atenção e os erros de transcrição, assumidos como erros aleatórios.
- **Erros de medida por fatores associados à tarefa de avaliação:** estes dizem respeito às características da própria avaliação como, por exemplo, uma má interpretação dos parâmetros ou a falta de proximidade com a realidade perceptiva. Inclui-se ainda o contexto perceptivo, pois pode alterar a referência interna do avaliador. Por exemplo, ao escutar uma voz classificada como moderada após várias amostras de vozes ligeiramente desviadas, então o avaliador poderá tender a agravar a sua percepção da voz moderada e a alterar a sua referência.
- **Erros de medida por fatores associados à interação entre o avaliador e a tarefa de avaliação:** a sensibilidade do avaliador pode interagir com a resolução da escala usada, no sentido de que a sua percepção face a um determinado parâmetro possa ser mais fácil, por ter contato com mais casos com essa característica evidente. Contudo, o avaliador pode não estar tão sensível a outros parâmetros.

2.3 A AVALIAÇÃO ACÚSTICA

A análise acústica vocal tem como objetivo quantificar e caracterizar um sinal sonoro. Algumas técnicas de processamento digital de sinais têm sido usadas para analisar desordens vocais provocadas por patologias na laringe. A análise vocal das características temporais e espectrais dos sinais de fala pode ser utilizada como técnica auxiliar à laringoscopia, para pré-diagnósticos de patologias (TEIXEIRA; FERREIRA; CARNEIRO, 2011).

Quando usada no âmbito do estudo da voz, permite determinar e quantificar a qualidade vocal do indivíduo por meio dos diferentes parâmetros acústicos que compõem o sinal – periodicidade, amplitude, duração e composição espectral. Dessa forma propiciam medidas objetivas que são extraídas automaticamente, mediante um processamento computacional e que permitem caracterizar os atributos físicos da voz nos domínios do tempo, da frequência e da intensidade. A análise é capaz de fornecer o formato da onda sonora, permitindo avaliar determinadas características, tais como a frequência fundamental (F_0), definida como o número de vibrações por segundo produzidas pelas pregas vocais, as medidas de variação da frequência fundamental, ciclo a ciclo, o *jitter* e ainda medidas da variação da amplitude, ciclo a ciclo, o *shimmer*. Estes constituem os principais parâmetros acústicos utilizados atualmente na detecção de patologias. (FREITAS, 2010; TEIXEIRA; FERREIRA; CARNEIRO, 2011).

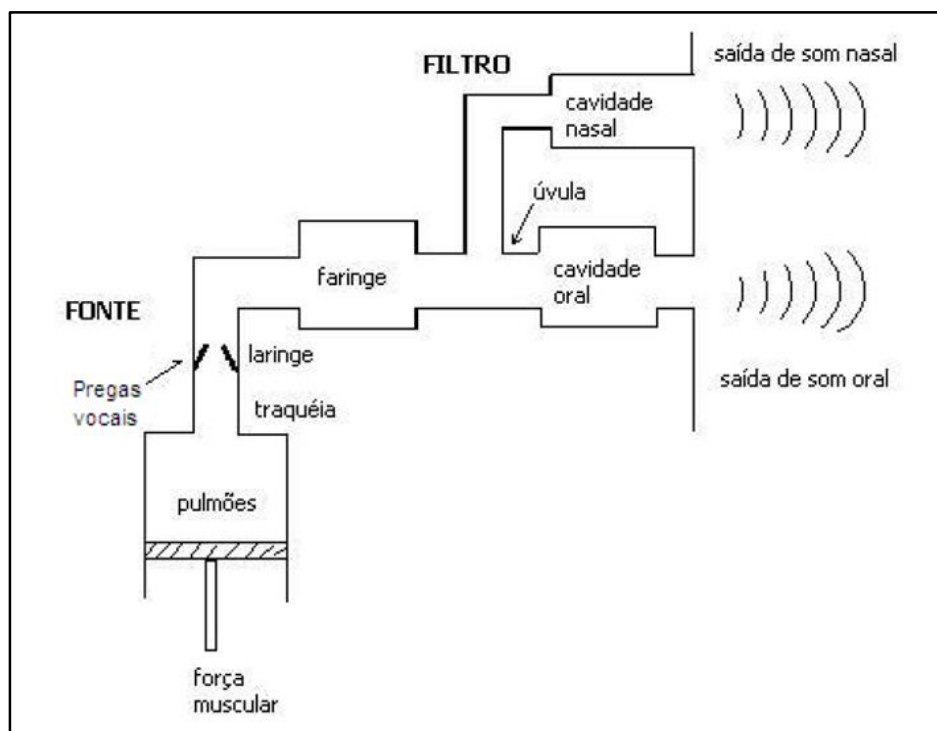
A F_0 da voz de um indivíduo é o reflexo das características biodinâmicas das pregas vocais e de sua integração com a pressão subglótica, sendo afetada pelo sexo e pela idade. No entanto, estes valores não são estacionários, uma vez que, dependem também de fatores como o estado de espírito da pessoa, o período do dia em que se enquadram (de manhã, à tarde e à noite), os hábitos de vida (álcool e tabaco), o uso profissional da voz (falada e cantada) e a disфония (GAMA; ALVES; CERCEAU, 2008; TEIXEIRA; FERREIRA; CARNEIRO, 2011).

As medidas objetivas da qualidade vocal são também suscetíveis de variáveis de confusão, que incluem: as condições de captação do sinal, as características específicas do sistema de *hardware* e *software*, os protocolos de recolhimento e

análise dos dados, a variabilidade individual em termos acústicos e aerodinâmicos e o grau de severidade e tipo de alteração vocal (FREITAS, 2010).

Além disso, a análise acústica não fornece medidas diretas da fonte glótica, uma vez que o sinal de fala registrado é o sinal de saída, que é o somatório do sinal glótico mais os efeitos dos filtros. O sinal glótico (som produzido pelas pregas vocais) sofre modificações ao longo do trato vocal (ação de filtro). Há um somatório das ondas sonoras provenientes da fonte glótica com ondas refletidas ao longo do trato vocal, sendo a resultante final, ou sinal de saída, o sinal irradiado pelos lábios como pode ser observada na Figura 3. Por este motivo, os instrumentos de análise realizam diagnósticos indiretos, a partir de procedimentos matemáticos que permitem, por exemplo, eliminar do sinal vocal de saída os efeitos da atividade supraglótica e apresentar medidas relacionadas à atividade glótica (SANTOS, 2015).

Figura 3 – Diagrama de blocos da produção da voz humana.



Fonte: Santos, (2015).

Contudo, a avaliação acústica oferece medições válidas, mas ainda insuficientes para os clínicos, que a encaram como um resultado complementar de diagnóstico que permite o registo e oferece a possibilidade de maior detalhe analítico

do processo de geração do sinal sonoro, o qual corresponde a eventos das porções glóticas e supraglóticas do aparelho fonador (FREITAS, 2010).

2.4 IMPLICAÇÕES CLÍNICAS

A aplicação de diversas formas de análise depende da correta digitalização do sinal acústico. Vários autores oferecem sugestões ou normas para obter um adequado áudio de vozes e conseguir uma correta análise do ponto de vista visual (espectrografia) e quantitativa (extração de medidas temporais, de periodicidade, amplitude, perturbação, ruído ou outras).

O sinal acústico decorrente do vozeamento de uma vogal sustentada é captado por um microfone, digitalizado e, posteriormente, analisado por meio de um procedimento computacional para medição de fatores de perturbação, associados à sua forma de onda.

Uma pesquisa que comparou a EVA com a EN na avaliação da gravidade da rinite alérgica apontou que a EVA representa melhor a gravidade do problema, uma vez que, por meio dela, o paciente apresenta melhor sua percepção da alteração. Além disso, os médicos e pacientes que participaram da pesquisa relataram que a utilização da EVA é mais rápida e fácil que a da EN (BEHLAU et al, 2015).

Na correlação entre os dados da análise auditiva e da análise acústica, as variáveis foram diversas, estando apresentadas a seguir: tipo de voz (auditiva) X GNE - Glottal to Noise Excitation ratio é uma medida acústica para calcular o ruído em uma série de pulsos produzidos pela oscilação das pregas vocais (acústica); tipo de voz (auditiva) X subharmônicos (acústica); tipo de voz (auditiva) X ruído entre harmônicos (acústica); tipo de voz (auditiva) X definição dos harmônicos na espectrografia (acústica); tipo de voz (auditiva) X Jitter (acústica); voz trêmula (auditiva) X Jitter (acústica) e pitch (auditiva) X frequência fundamental (acústica) – por sexo (ORTIZ; CARRILLO, 2008).

3 TRANSFORMADA WAVELET E REDES NEURAIIS ARTIFICIAIS

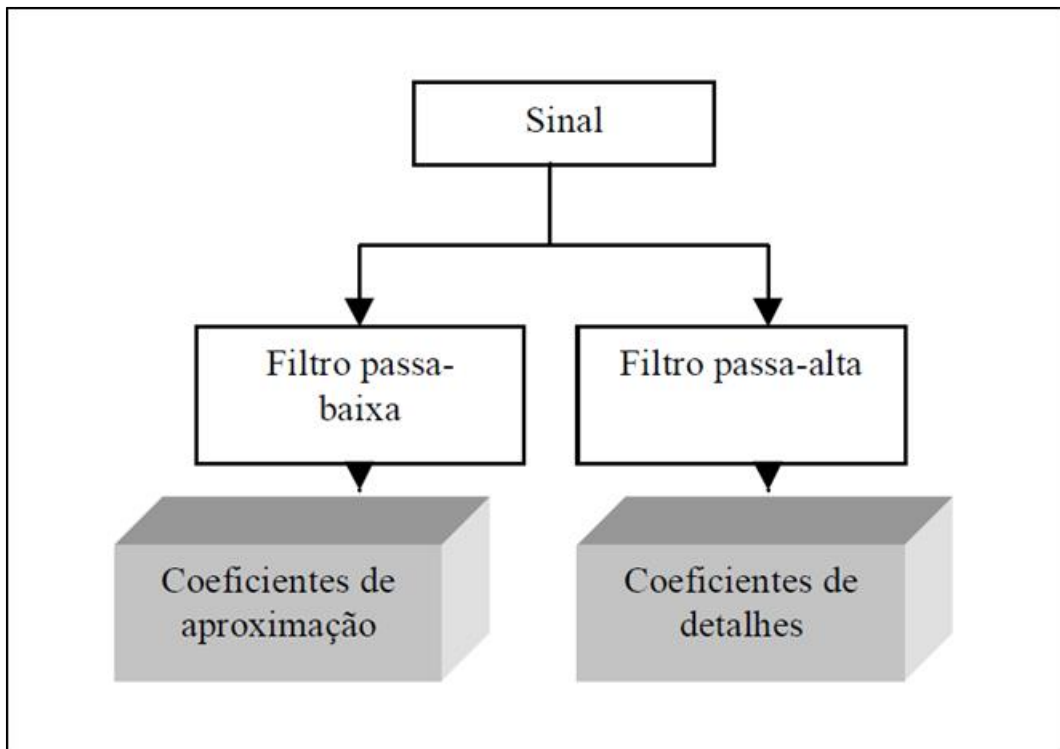
Neste capítulo são descritas as etapas para realização de um sistema computacional, que seja capaz de classificar o parâmetro subjetivo predominante de uma amostra de áudio vocal e estimar o seu grau através de ferramentas inteligentes (RNAs) combinadas com a WPT.

3.1 TRANSFORMADA WAVELET DISCRETA

Diversos sinais encontrados na natureza possuem características não estacionárias, ou seja, variam com o tempo, como os sinais de voz. Tais sinais para serem melhor observados exigem uma aproximação mais flexível, em que o tamanho da janela seja variável, determinando mais precisamente informações sobre tempo ou frequência de um determinado sinal. Uma maneira eficiente de aplicar a Transformada Wavelet Discreta (TDW) é por meio de filtros, apresentando alta capacidade de concentrar características importantes em seus coeficientes, sendo ela uma ferramenta que permite decompor um sinal em diferentes componentes de frequências, permitindo, assim, estudar cada componente separadamente em seu nível correspondente (SANTOS, 2015).

Sendo sua implementação equivalente a bancos de filtros não uniformes com uma melhor resolução em frequência baixas, em sua análise de multirresolução, o sinal é dividido em coeficientes de aproximação e detalhes, representadas pela Figura 4, mas apenas os coeficientes de aproximação são novamente divididos em coeficientes de aproximação e detalhes, gerando o segundo nível de decomposição, podendo ser decompostos por quantas vezes achar-se necessário (CROVATO, 2004).

Figura 4 – Decomposição wavelet discreta multiresolução.



Fonte: Parraga, (2002).

3.2 TRANSFORMADA WAVELET PACKET

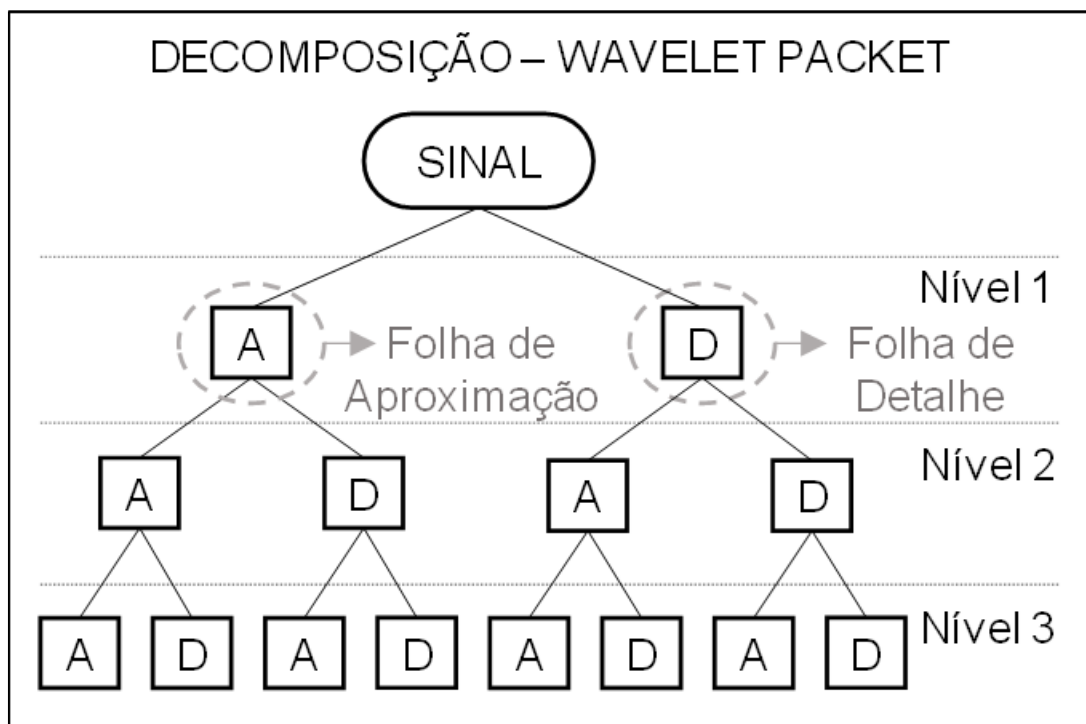
A transformada *wavelet packet* (TWP) é uma generalização do conceito da DWT, na qual a resolução tempo-frequência pode ser escolhida de acordo com o sinal, embora isto seja feito dentro dos limites do princípio de incerteza de Heisenberg¹ (CROVATO, 2004).

Neste trabalho é proposta a aplicação da TWP como extrator de características para a estimação automática do grau de parâmetros subjetivos vocais. As funções que a compõe possuem a importante propriedade de decomposição do plano tempo-frequência. A decomposição de sinais nestas bases forma uma estrutura de árvore binária. Os coeficientes de detalhe e aproximação e que também são conhecidos de “folhas” da *wavelet packet*.

¹ O princípio da incerteza de Werner Heisenberg, estabelece que é impossível conhecer simultaneamente a posição e a energia de uma partícula tal como o elétron. O princípio da incerteza pode ser assim interpretado: quanto mais de perto tentamos olhar uma partícula diminuta, tanto mais difusa se torna a visão da mesma.

A Figura 5 representa o diagrama referente as decomposições em árvore binária resultantes da aplicação da WPT, em que a cada nível de decomposição, os coeficientes são divididos em coeficientes de aproximação e detalhes. Desta forma, tem-se o sinal representado em bandas de frequências com diferentes resoluções (CROVATO, 2004).

Figura 5 – Decomposição da transformada wavelet packet em árvore binária.



Fonte: Adaptado de Fermino et al (2016).

Conforme visto na Figura 5, o sinal é decomposto tanto em alta como em baixa frequência, e a cada novo nível de decomposição, $2^{nível}$ bandas de decomposição são acrescentadas.

3.3 REDES NEURAIIS ARTIFICIAIS (RNAs)

A estrutura das RNAs originou-se das tentativas matemáticas de representar como a informação é processada em sistemas nervosos biológicos e do próprio cérebro humano, e vem sendo utilizada até hoje para cobrir diversos modelos diferentes especializados em soluções específicas de problemas. (BISHOP, 2007), (DA SILVA; SPATTI; FLAUZINO, 2010).

As RNAs simulam o aproveitamento dos aspectos do raciocínio humano para resolução de problemas e são recomendadas para extrair conhecimento de determinada área ou problema, se adaptando para atingir a melhor função dentro do contexto inserido, mediante treinamento, conseguindo então atingir o processo de generalização para estimar uma solução ou resultado.

Os neurônios artificiais utilizados nos modelos de RNAs fornecem saídas tipicamente contínuas e realizam funções simples, como coletar os sinais existentes em suas entradas, agregá-los de acordo com sua função operacional e produzir uma resposta, levando em consideração sua função de ativação inerente (DA SILVA; SPATTI; FLAUZINO, 2010).

Por meio da Figura 6 pode-se observar as principais estruturas do neurônio artificial, que são:

- **Sinais de entrada: $\{x_1, x_2, \dots, x_n\}$**
São sinais ou medidas advindas do meio externo e que apresentam os valores assumidos pelas variáveis de uma aplicação específica. Os sinais de entrada são usualmente normalizados visando incrementar a eficiência computacional dos algoritmos de aprendizagem.
- **Pesos sinápticos $\{w_1, w_2, \dots, w_n\}$**
São os valores que servirão para ponderar cada uma das variáveis de entrada da rede, permitindo-se quantificar as suas relevâncias em relação à funcionalidade do respectivo neurônio.
- **Combinador linear $\{\Sigma\}$**
Sua função é agregar todos os sinais de entrada que foram ponderados pelos respectivos pesos sinápticos a fim de produzir um valor de potencial de ativação.

- **Limiar de ativação $\{\theta\}$**

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

- **Potencial de ativação $\{u\}$**

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

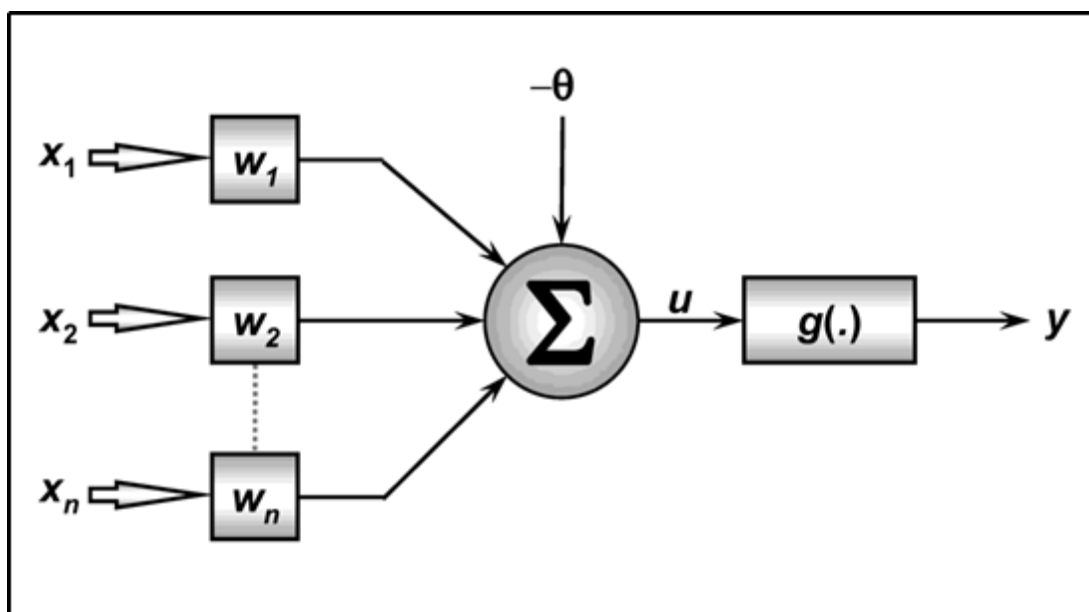
- **Função de ativação $\{g\}$**

Seu objetivo é limitar a saída do neurônio dentro de um intervalo de valores razoáveis.

- **Sinal de saída $\{Y\}$**

Consiste do valor produzido pelo neurônio em relação a um determinado conjunto de sinais de entrada, podendo ser também utilizado por outros neurônios que estão sequencialmente interligados.

Figura 6 – Representação do neurônio artificial.



Fonte: Spatti (2011).

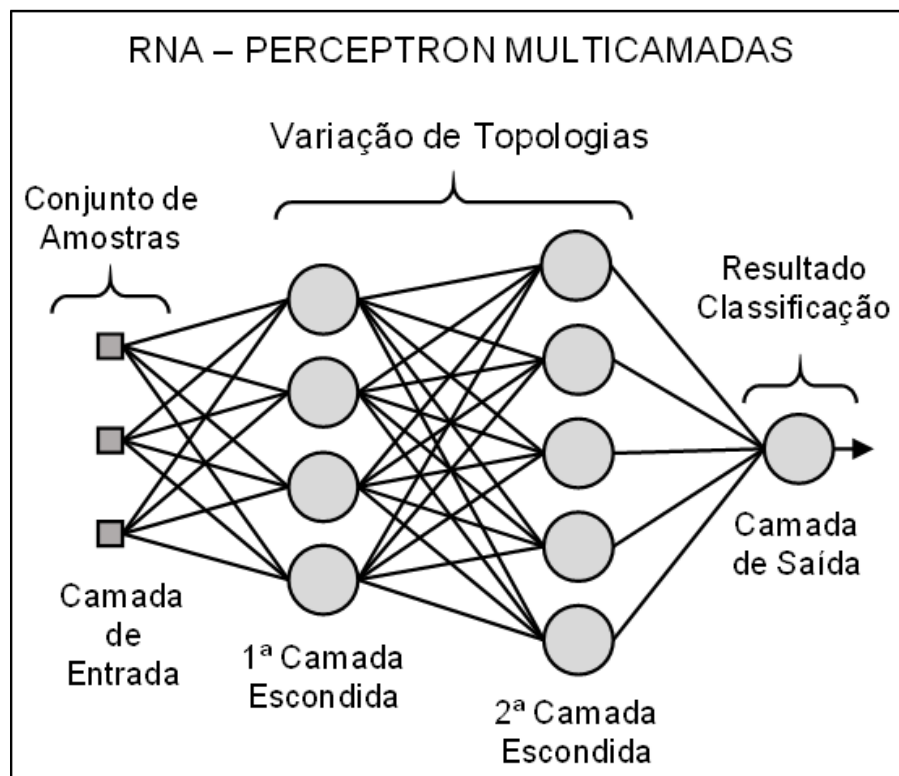
O funcionamento das RNAs mais frequentes é definido pelo conceito de *feedforward* (em tradução livre, “alimentação para frente”), que representa o processamento dentro da estrutura, seguindo um fluxo da entrada para saída, sem laços ou realimentações.

3.3.1 Perceptron multicamadas

As redes neurais artificiais de arquitetura Perceptron multicamadas (PMC) são caracterizadas pela presença de pelo menos uma camada intermediária (escondida) de neurônios, situada entre a camada de entrada e a respectiva camada neural de saída. Conseqüentemente, as redes PMC possuem no mínimo duas camadas de neurônios, os quais estarão distribuídos entre as camadas intermediárias e a camada de saída. (DA SILVA; SPATTI; FLAUZINO, 2010)

Tal descrição pode-se ser melhor compreendida por meio da Figura 7

Figura 7 – Exemplo de rede feedforward de camada múltiplas.



Fonte: adaptado de Fermino et al (2016).

No Quadro 1 é possível visualizar as principais funções de ativação, bem como seu comportamento em relação a entrada e saída e a simbologia que representa seu princípio de operação:

Sendo,

u : valor recebido pela função de ativação

Y : valor de saída

Quadro 1 - Principais funções de ativação.

NOME	RELAÇÃO ENTRADA (u) E SAÍDA (Y)	SÍMBOLO
Degrau	$u < 0 \quad Y = 0$ $u \geq 0 \quad Y = 1$	
Degrau bipolar	$u < 0 \quad Y = -1$ $u \geq 0 \quad Y = +1$	
Linear	$Y = u$	
Linear simétrica	$u < 0 \quad Y = 0$ $0 \leq u \leq 1 \quad Y = u$ $u > 1 \quad Y = 1$	
Rampa simétrica	$u < -1 \quad Y = -1$ $-1 \leq u \leq 1 \quad Y = u$ $u > 1 \quad Y = 1$	
Logística	$Y = \frac{1}{1 + e^{-u}}$	
Tangente hiperbólica	$Y = \frac{e^u - e^{-u}}{e^u + e^{-u}}$	
Linear positiva	$u < 0 \quad Y = 0$ $0 \leq u \quad Y = u$	
Competição	$Y = 1$ neurônio com maior u $Y = 0$ outros neurônios	-----

Fonte: Adaptado de Hagan (2014).

Para esse estudo, foi utilizada a função de ativação tangente hiperbólica para suas camadas escondidas e a linear da camada neural de saída da PMC.

4 MATERIAIS E MÉTODOS

A partir de uma amostragem de áudios, procedimentos foram realizados para a composição de um classificador de predominância/estimador de parâmetros subjetivos (Rugosidade, Soprosidade e Tensão) da voz humana, sendo a ferramenta utilizada o *software* MATLAB®R2016a. Foram utilizadas também as funções implementadas nos pacotes instalados em conjunto com o MATLAB®, por exemplo, para aplicação de transformadas e extração de parâmetros por meio da *toolbox* de processamento de sinais e treinamento da RNA presente na *toolbox* de redes neurais.

4.1 BANCO DE DADOS

Os arquivos de áudios com resultados da avaliação perceptivo-auditiva foram concedidos pelo Centro de Estudos da Voz (CEV). Objetivando compor um banco de dados consistente para treinamento de fonoaudiólogos especialistas em voz, os pesquisadores do CEV selecionaram e analisaram prontuários médicos de pacientes que apresentavam sinais de voz com vozes predominantemente rugosas, vozes predominantemente soprosas e vozes predominantemente tensas. Em outras palavras, cada sinal de voz devia ter apenas um parâmetro perceptivo predominante.

Dos pacientes selecionados, efetuaram-se gravações acústicas da vogal sustentada /e/ em níveis confortáveis para os mesmos. As gravações foram coletadas em uma sala acusticamente tratada e com o acompanhamento de uma fonoaudióloga com experiência na realização de testes vocais. Para a captura dos sinais de voz 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, todos os áudios foram gravados em arquivos com formato “.wav”.

4.2 PRÉ-PROCESSAMENTO DO BANCO DE DADOS

As amostras de áudios foram separadas por gênero e predominância de parâmetros subjetivos, sendo cada amostra composta pela sustentação da vogal /e/ de indivíduos distintos, representado pelo Quadro 2:

Quadro 2 – Amostragem do banco de dados

Parâmetros / gêneros	Masculinos	Femininos
Rugosidade	5	5
Soprosidade	5	5
Tensão	5	5

Fonte: Autoria própria.

Os áudios foram agrupados em três pastas distintas, descartando assim a dependência com gêneros, como é melhor observado no Quadro 3:

Quadro 3 – Relação dos arquivos independentes do gênero.

Pastas / amostras	Áudios
Rugosa	10
Soprosa	10
Tensa	10

Fonte: Autoria própria.

Feito isso, um áudio de cada pasta foi retirado com o intuito de testar a capacidade de generalização posteriormente com amostras desconhecidas, utilizando-os na etapa que será descrita na seção dos resultados. Logo, há a disposição das amostras apresentadas pelo Quadro 4 para o procedimento de treinamento e validação:

Quadro 4 – Arquivos para treinamento e validação.

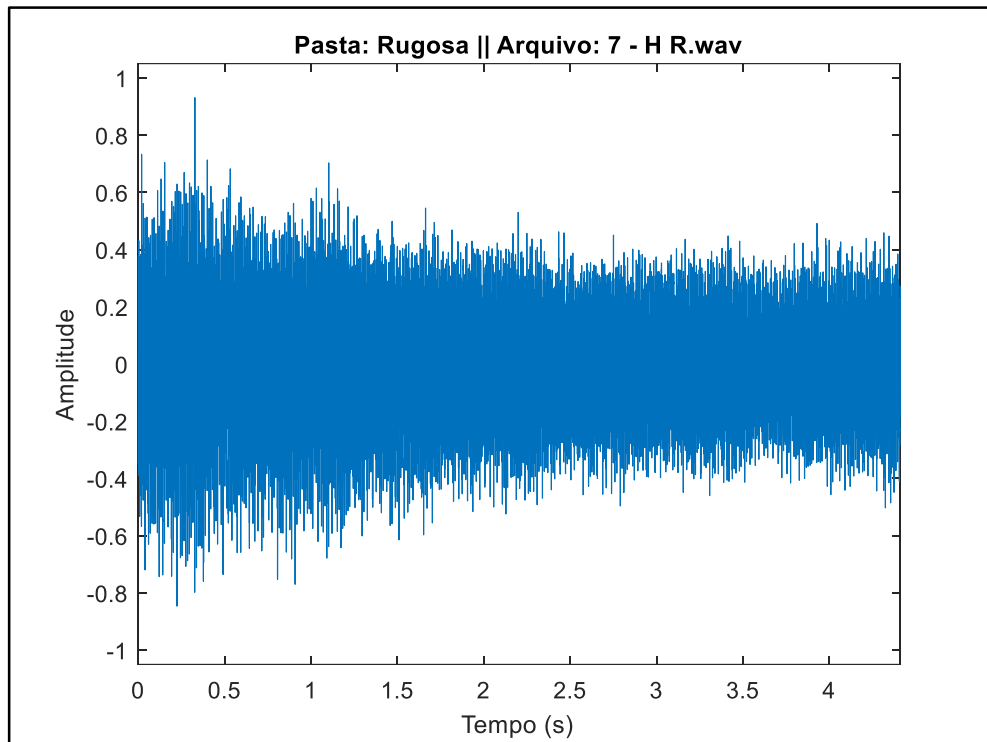
Pastas/ amostras	Áudios
Rugosa	9
Soprosa	9
Tensa	9

Fonte: Autoria própria.

A partir do software MATLAB®R2016a, efetua-se o carregamento de forma automática das três pastas, ocorrendo uma varredura por seus diretórios, para que

assim, tenha a quantidade exata de arquivos a serem processados. Utilizando-se, então, a função *audioread*, obtém-se o sinal vocal representados em formas de gráficos. Concebidos os arquivos 7 – HR.wav, 93-HT.wav e o 20- HS.wav², é possível gerar seus gráficos, que são representados nas Figuras 8, 9 e 10, respectivamente.

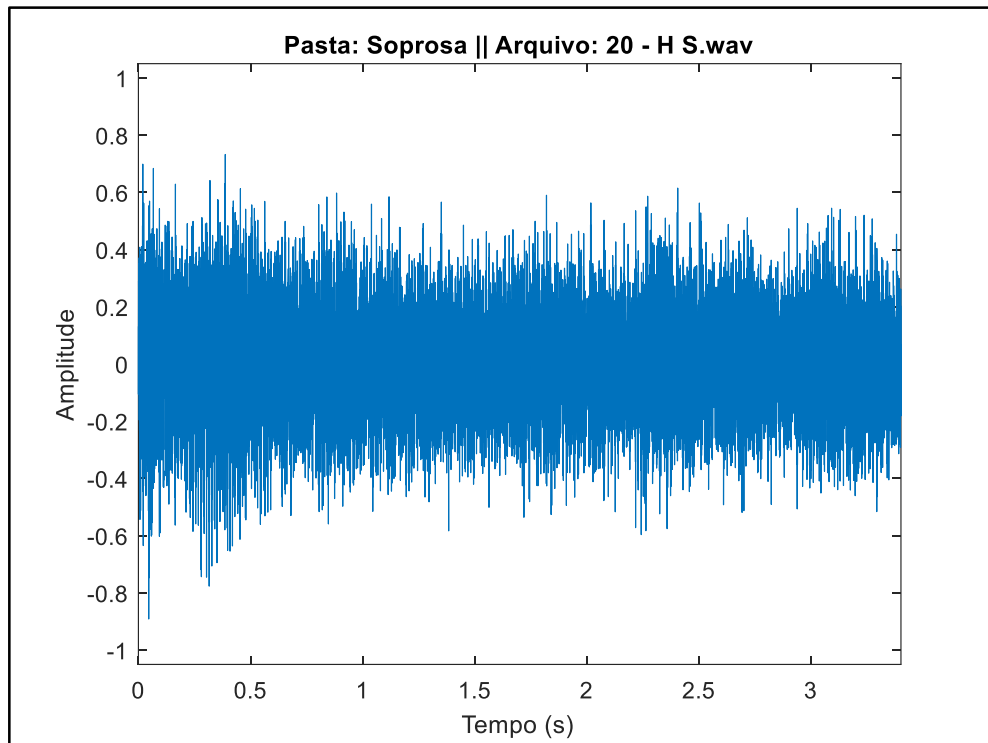
Figura 8 – Sinal vocal predominantemente rugoso – H R.wav.



Fonte: Autoria própria.

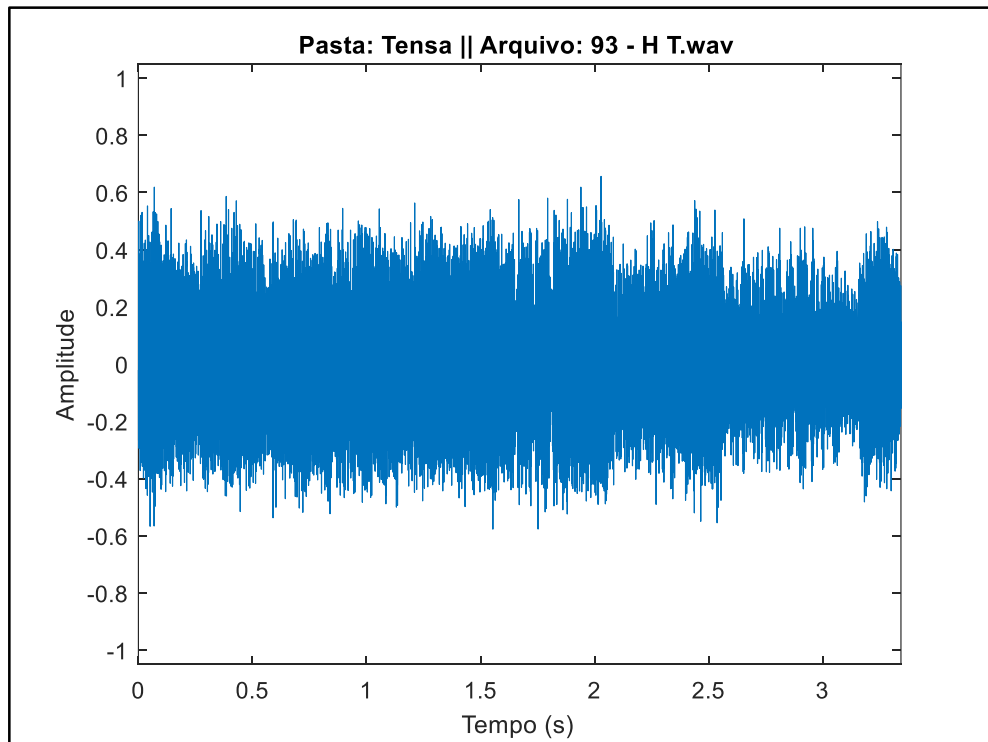
² Como critério de organização, os arquivos recebem identificações compostas pelo gênero (H – Homem, M – Mulher), pelo tipo de parâmetro predominante (R – Rugosidade, S – Soprosidade, T – Tensão) e a numeração apresentada como forma de organização do grupo de pesquisas que cedeu o banco de dados (7, 20, 90, ...).

Figura 9 – Sinal vocal predominantemente soproso– H S.wav.



Fonte: Autoria própria.

Figura 10 – Sinal vocal predominantemente tenso – H T.wav.



Fonte: Autoria própria.

4.3 NORMALIZAÇÃO DAS AMOSTRAS DE ÁUDIOS

A amostragem tem o propósito de exemplificar os cuidados necessários com a preparação dos arquivos de áudio utilizados no banco de dados. É importante que seja garantido uma uniformidade na amostragem dos sinais, mesmo na presença de diferentes frequências de amostragem, como encontradas nas amostras do banco de dados, sendo elas de 22kHz ou 44kHz.

É necessário, então, saber qual é a mínima taxa de amostragem apresentada nos áudios para que, com isso, seja possível reamostrar esses sinais e posteriormente construir um vetor no domínio do tempo.

Desta forma, será efetuado um novo carregamento dos arquivos de áudios reamostrados a partir da mínima taxa de amostragem de 22kHz encontrada por meio da função *resample* do *software*

4.4 JANELAMENTO DAS AMOSTRAS DE ÁUDIOS

Dando sequência a etapa anterior, é necessário o janelamento do sinal em amostras menores, com o intuito de detalhar as informações e incrementar a quantidade de segmentos apresentado as RNAs, possibilitando condições melhores de treinamento. Sabe-se que a WPT realiza uma decomposição por dois a cada nível. Então, faz-se interessante a escolha de segmentos de base 2. Portanto, o tamanho da janela escolhida tem extensão de 512 pontos.

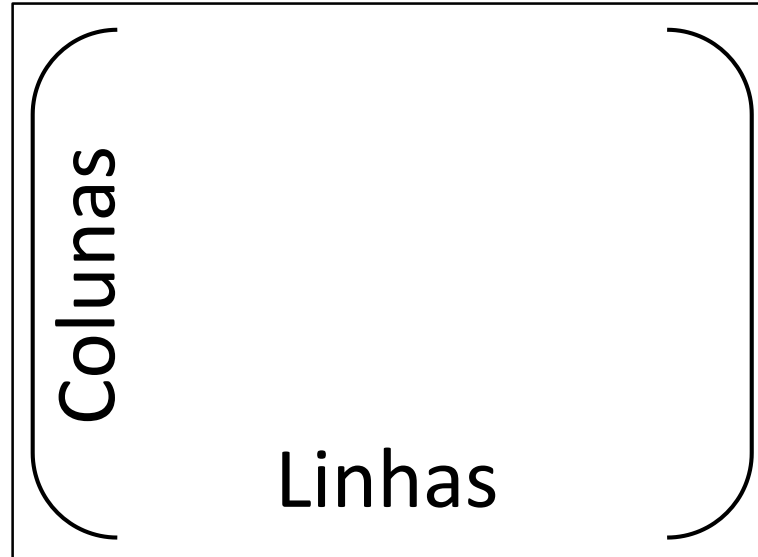
Posteriormente, é possível determinar o tamanho da matriz do banco de dados que será utilizado para realizar o cálculo da WPT por meio da Equação 1, 2 e 3, bem como, representada visualmente pela Figura 11:

$$\text{Quantidade de janelas} = \frac{\text{comp. vetor do sinal [pontos]}}{\text{tamanho da janela escolhida [pontos]}} \quad (1)$$

$$\text{Colunas} = (\text{Tamanho da janela escolhida} + \text{Quantidades de saídas}) \quad (2)$$

$$\text{Linhas} = (\text{Quantidades de janelas} * \text{Quantidades de arquivos}) \quad (3)$$

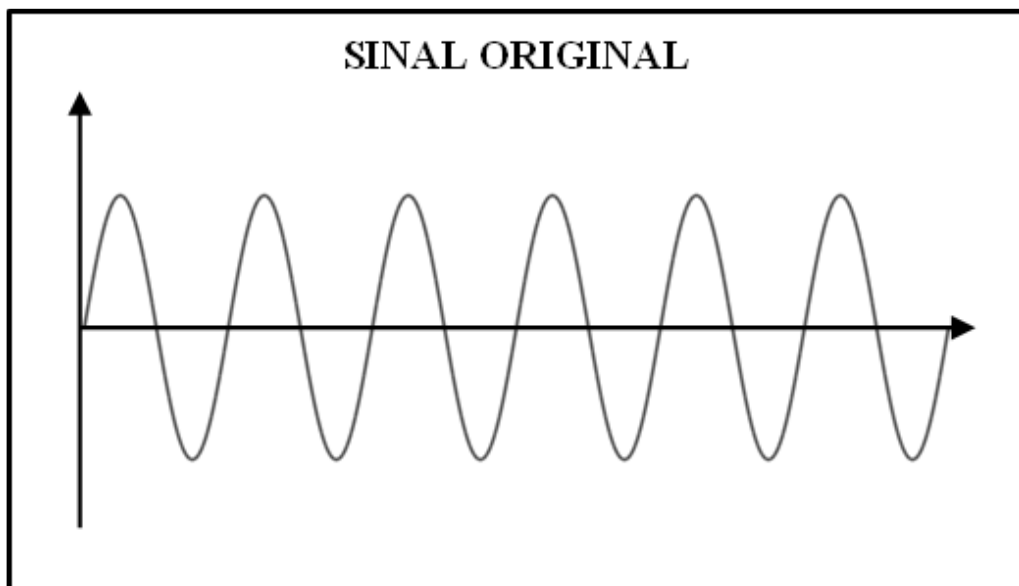
Figura 11 – Exemplificação da matriz de janelamento.



Fonte: Autoria própria.

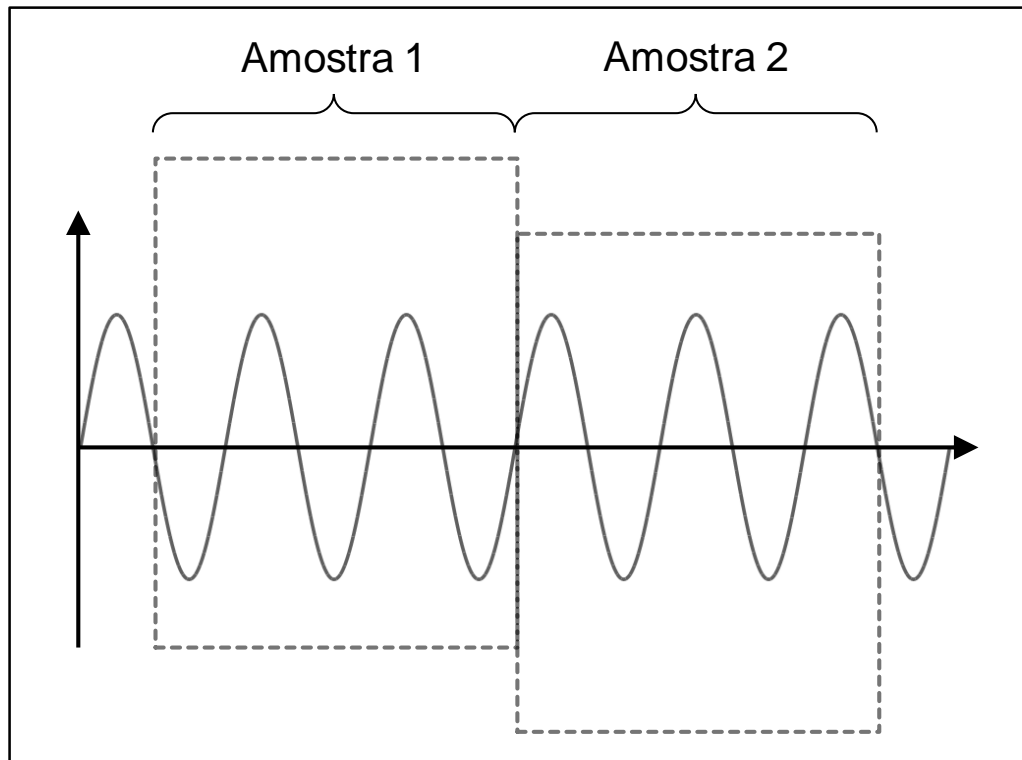
Considera-se o método do janelamento normal, em que o término de uma janela representa início de outra, sendo possível entender de maneira mais clara o procedimento por meio das Figuras 12 e 13.

Figura 12 – Exemplo de sinal original reamostrado e normalizado.



Fonte: Adaptado de Fermino et al (2016).

Figura 13 – Janelamento simples do sinal.



Fonte: Adaptado de Fermino et al (2016).

4.5 EXTRAÇÃO DE CARACTERÍSTICAS UTILIZANDO WPT

Assumindo que a energia do sinal associada às faixas de frequência dos diferentes níveis de resolução pode apontar um desvio vocal, utilizou-se para o desenvolvimento deste trabalho o nível de energia como característica relevante ao estimador proposto. Tal nível, como é definido por em Hsu (1995), no caso de um sinal discreto, é calculado por meio da equação (4):

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

Obtém-se,

$x[n]$: janela a ser analisada;

E : o valor da energia obtido.

Em seguida, há a elaboração de um processo automatizado para extração das características de todas as janelas criadas e composição das amostras que serão apresentadas como entradas para as RNAs.

Para cada segmento do sinal (ou janela) foi aplicada a WPT, utilizando o nível 4 de resolução, gerando 16 folhas. Feito isso, é calculado o nível de energia de cada folha, estruturando-se, então, 16 valores de entrada para as RNAs. Obtém-se também o preenchimento das saídas, atribuindo o valor 1 para o parâmetro predominante presente em dada janela e valor 0 para os outros parâmetros. Desta forma, é composta a matriz dos dados de entradas para o estimador elaborado. Tal composição é representada visualmente na Figura 14.

Figura 14 – Matriz dos dados de entradas das RNAs.

		ENTRADAS																SAÍDAS		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	RUGOSA	SOPROSA	TENSA
1 3861																				

Fonte: Autoria própria.

É importante ressaltar que para melhorar o desempenho da etapa de treinamento das RNAs, faz-se necessário que as entradas das redes estejam normalizadas, adequando as variações das entradas em relação à função de ativação do neurônio.

4.6 CARACTERÍSTICAS DAS RNAS E PROCESSO DE TREINAMENTO

Após a execução dos procedimentos anteriores, o objetivo seguinte é aplicar o método de treinamento das RNAs com uma rotina automatizada, utilizando-se do conjunto de entradas construído. Será aplicada a rede do tipo PMC (como já foi discutido nos capítulos anteriores), treinada com um algoritmo supervisionado.

Foram realizados dez treinamentos na rotina automatizada antes de se alterar a topologia, pois considerando a inicialização aleatória, a prática de vários treinamentos na mesma topologia evita descartar uma configuração promissora devido ao confinamento em um mínimo local da função de aproximação em que a PMC será capaz de prever.

As variações de topologias, bem como, as características das RNAs e das famílias de *wavelets* estão resumidas no Quadro 5:

Quadro 5 – Características da RNA aplicadas

Arquitetura da rede	Perceptron Multicamadas
Tipo de treinamento	Supervisionado – <i>Off-line</i>
Algoritmo de treinamento	Levenberg-Marquardt (<i>backpropagation</i>)
Taxa de aprendizagem	10^{-5}
Critério de parada (Erro)	10^{-2}
Épocas	4000
Proporção Treinamento / Validação	60% / 40%
Número total de amostras	3805
Número de repetições para cada topologia	10
Quantidade de topologias testadas	265
Número de treinamentos	2650
Quantidade mínima de neurônios por camada	2
Quantidade máxima de neurônios por camada	26
Famílias wavelets	Coiflets (1, 2, 3, 4, 5), Daubechies (1, 2, 3, 4, 5), Haarn e Symlets (2, 3, 4, 5)

Fonte: Autoria própria.

Sabendo-se que a rede PMC operando como estimador apresenta sua funcionalidade fundamentada em um aproximador universal de funções, é importante ressaltar que o melhor desempenho alcançado será por meio da separação do estimador em três redes especialistas, pois a capacidade de aproximar a função que descreve o comportamento dos parâmetros subjetivos será consideravelmente inferior quando trabalhadas de forma conjunta.

Ademais, é factível mencionar que as famílias *wavelets* abordadas para esse trabalho foram escolhidas no intuito de considerar possibilidades para que o problema tivesse uma melhor resposta tanto de treinamento quanto de validação, resultando em um erro percentual baixo para cada RNA especialista.

5 RESULTADOS

Este capítulo é dedicado a apresentar os resultados obtidos na aplicação dos métodos propostos no capítulo anterior.

Nas três primeiras subseções são apresentadas as melhores topologias para uma, duas e três camadas escondidas, assim como seus respectivos desempenhos. Em seguida, há uma seção que reúne os melhores treinamentos para a escolha das topologias com maior eficiência.

Por fim, na última subseção será realizado o teste de generalização da rede, aplicando o estimado em amostras de áudio desconhecidas (retiradas do conjunto no início da preparação do banco de dados), permitindo observar o comportamento real da melhor RNA obtida.

5.1 MELHORES TOPOLOGIAS

Nesta seção serão relatados os seis melhores resultados de treinamento/validação das RNAs especialistas em rugosidade, sopro e tensão.

As características das melhores topologias especialistas em rugosidade para uma camada, duas camadas e três camadas escondidas bem como as famílias das wavelets referentes aos melhores resultados são representados nos Quadros 6, 8 e 10. Por sua vez, os resultados dos 10 treinamentos são apresentados nos Quadros 7, 9 e 11.

Quadro 6 – Melhor topologia encontrada (rugosidade – uma camada)

TOPOLOGIA DA REDE	
Camadas	1
Nº de neurônios	20
Família <i>wavelet</i>	Coiflets 5

Fonte: Autoria própria.

Quadro 7 – Resultado dos treinamentos (rugosidade – uma camada)

TESTE	ERRO (%)
1º	6,131767776
2º	5,348988911
3º	5,740378343
4º	4,631441618
5º	5,414220483
6º	5,479452055
7º	7,175472929
8º	6,979778213
9º	5,283757339
10º	7,436399217

Fonte: Autoria própria.

Quadro 8 – Melhor topologia encontrada (rugosidade – duas camadas)

TOPOLOGIA DA REDE	
Camadas	2
Nº de neurônios	13 – 18
Família <i>wavelet</i>	Daudechies 5

Fonte: Autoria própria.

Quadro 9 – Resultado dos treinamentos (rugosidade – duas camadas)

TESTE	ERRO (%)
1º	5,218525766
2º	5,870841487
3º	33,13763862
4º	5,02283105
5º	4,69667319
6º	33,13763862
7º	6,327462492
8º	5,740378343
9º	5,936073059
10º	5,870841487

Fonte: Autoria própria.

Quadro 10 – Melhor topologia encontrada (rugosidade – três camadas)

TOPOLOGIA DA REDE	
Camadas	3
Nº de neurônios	12 – 9 – 5
Família <i>wavelet</i>	Coiflets 5

Fonte: Autoria própria.

Quadro 11 – Resultado dos treinamentos (rugosidade – três camadas)

TESTE	ERRO (%)
1º	4,827136334
2º	6,58838878
3º	32,9419439
4º	5,870841487
5º	4,631441618
6º	4,827136334
7º	4,500978474
8º	7,175472929
9º	5,153294194
10º	4,892367906

Fonte: Autoria própria.

As características da melhor topologia especialista em soproside para uma camada, duas camadas e três camadas escondida, bem como as famílias das *wavelets* referentes ao melhor resultado são representados nos Quadros 12, 14 e 16. Por sua vez, os resultados dos 10 treinamentos são apresentados nos Quadros 13, 15 e 17.

Quadro 12 – Melhor topologia encontrada (soproside – uma camada)

TOPOLOGIA DA REDE	
Camadas	1
Nº de neurônios	16
Família <i>wavelet</i>	Daubechies 3

Fonte: Autoria própria.

Quadro 13 – Resultado dos treinamentos (soproside – uma camada)

TESTE	ERRO (%)
1º	6,58838878
2º	5,675146771
3º	5,153294194
4º	5,348988911
5º	5,348988911
6º	6,523157208
7º	4,69667319
8º	5,936073059
9º	5,088062622
10º	4,435746902

Fonte: Autoria própria.

Quadro 14 – Melhor topologia encontrada (soproside – duas camadas)

TOPOLOGIA DA REDE	
Camadas	2
Nº de neurônios	8 – 6
Família <i>wavelet</i>	Daubechies 5

Fonte: Autoria própria.

Quadro 15 – Resultado dos treinamentos (soprosidade – duas camadas)

TESTE	ERRO (%)
1º	3,13111546
2º	4,370515329
3º	4,631441618
4º	4,370515329
5º	4,370515329
6º	5,02283105
7º	4,305283757
8º	2,804957599
9º	4,69667319
10º	4,174820613

Fonte: Autoria própria.

Quadro 16 – Melhor topologia encontrada (soprosidade – três camadas)

TOPOLOGIA DA REDE	
Camadas	3
Nº de neurônios	9 – 5 – 3
Família <i>wavelet</i>	Daubechies 5

Fonte: Autoria própria.

Quadro 17 – Resultado dos treinamentos (soprosidade – três camadas)

TESTE	ERRO (%)
1º	4,109589041
2º	3,979125897
3º	4,827136334
4º	6,523157208
5º	4,240052185
6º	4,69667319
7º	5,414220483
8º	5,283757339
9º	4,69667319
10º	4,631441618

Fonte: Autoria própria.

As características das melhores topologias especialistas em tensão para uma camada, duas camadas e três camadas escondidas, bem como as famílias das *wavelets* referentes aos melhores resultados são representados nos Quadros 18, 20 e 22. Por sua vez, os resultados dos 10 treinamentos são apresentados nos Quadros 19, 21 e 23.

Quadro 18 – Melhor topologia encontrada (tensão – uma camada)

TOPOLOGIA DA REDE	
Camadas	1
Nº de neurônios	15
Família <i>wavelet</i>	Symlets 4

Fonte: Autoria própria.

Quadro 19 – Resultado dos treinamentos (tensão – uma camada)

TESTE	ERRO (%)
1º	4,109589041
2º	3,718199609
3º	4,174820613
4º	3,979125897
5º	4,566210046
6º	3,913894325
7º	4,761904762
8º	4,109589041
9º	4,631441618
10º	3,261578604

Fonte: Autoria própria.

Quadro 20 – Melhor topologia encontrada (tensão – duas camadas)

TOPOLOGIA DA REDE	
Camadas	2
Nº de neurônios	9 – 6
Família <i>wavelet</i>	Symlets 4

Fonte: Autoria própria.

Quadro 21 – Resultado dos treinamentos (tensão – duas camadas)

TESTE	ERRO (%)
1º	5,348988911
2º	3,979125897
3º	3,45727332
4º	2,674494455
5º	3,000652316
6º	2,609262883
7º	3,45727332
8º	3,979125897
9º	5,02283105
10º	3,848662753

Fonte: Autoria própria.

Quadro 22 – Melhor topologia encontrada (tensão – três camadas)

TOPOLOGIA DA REDE	
Camadas	3
Nº de neurônios	10 – 6 – 2
Família <i>wavelet</i>	Symlets 3

Fonte: Autoria própria.

Quadro 23 – Resultado dos treinamentos (tensão – três camadas)

TESTE	ERRO (%)
1º	4,240052185
2º	4,631441618
3º	4,305283757
4º	3,587736464
5º	32,81148076
6º	4,566210046
7º	5,02283105
8º	3,718199609
9º	3,13111546
10º	4,566210046

Fonte: Autoria própria.

5.2 MELHORES REDES NEURAIS ARTIFICIAIS

Considerando os resultados apresentados na subseção anterior, pode-se afirmar que as redes conseguiram absorver o conhecimento presente na massa de dados, permitindo estimar os valores desejados com erros inferiores à 5%.

Portanto, as melhores topologias dentre as testadas para o estudo proposto estão reunidas no Quadro 24.

Quadro 24 – Melhores topologias para cada parâmetro estudado.

MELHORES TOPOLOGIAS					
Parâmetro	Primeira Camada	Segunda Camada	Terceira Camada	Erro (%)	Famílias wavelets
Rugosidade	12 neurônios	9 neurônios	5 neurônios	4,500978474	Coiflets 5
Soprosidade	8 neurônios	6 neurônios	-----	2,804957599	Daubechies 5
Tensão	9 neurônios	6 neurônios	-----	2,609262883	Symlets 4

Fonte: Autoria própria.

5.3 VALIDAÇÃO DA GENERALIZAÇÃO DAS RNAS

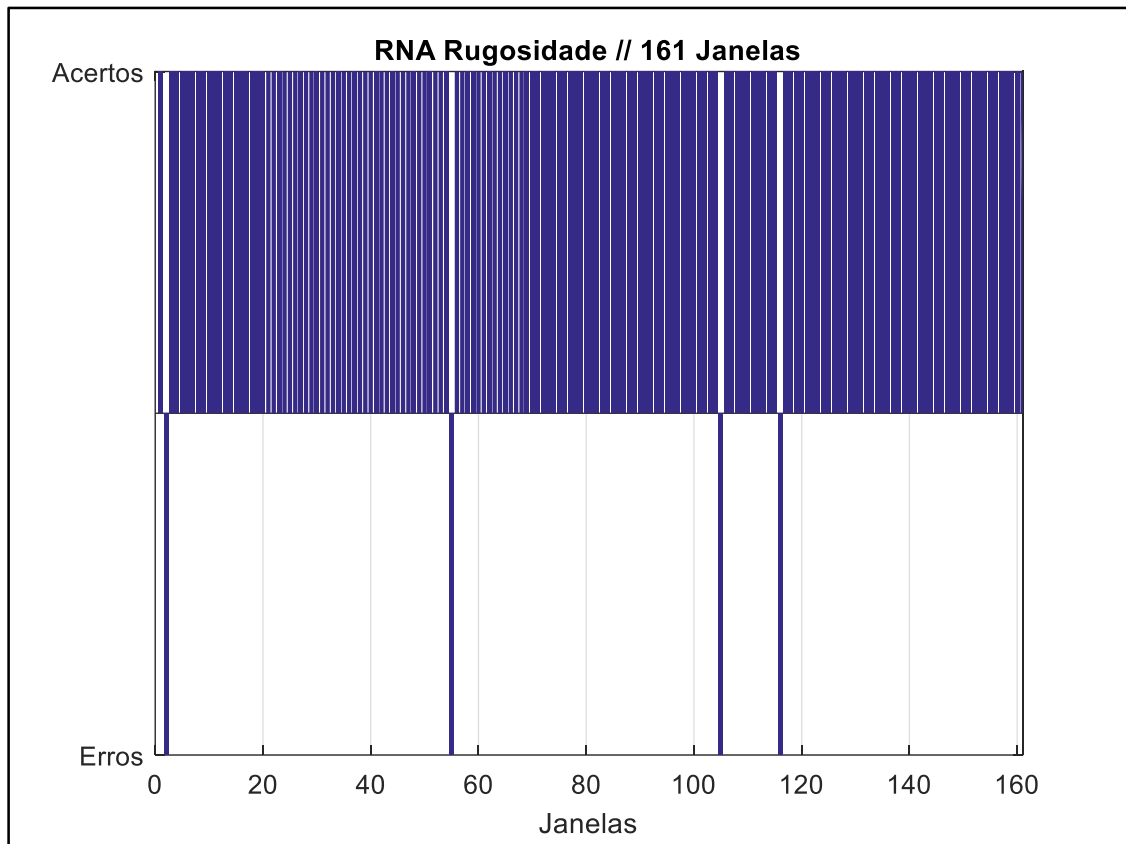
A legitimação da topologia mais adequada para a RNA capaz de solucionar o problema deu-se por meio de análise de amostras desconhecidas. Após obter RNAs com desempenho adequados na validação, utilizaram-se os três áudios que foram retirados da base de dados no início do processo. Tais arquivos passaram pelo mesmo procedimento de preparação da taxa de amostragem, janelamento, extração de características e parâmetros de normalização.

Com o intuito de estimar o grau dos parâmetros subjetivos, o método utilizado será com base no método de classificação. Sendo as amostras geradas a partir do arquivo de áudio, a quantidade de janelas pertencentes ao parâmetro especialista da rede em relação ao total de janelas obtidas indicará o grau de tal parâmetro do indivíduo analisado.

5.3.1 Análise de áudio rugoso

Os resultados da análise do áudio rugoso para cada uma das RNAs especialistas estão representados graficamente nas Figuras 15, 16 e 17, bem como seus resultados nos Quadros 25, 26 e 27:

Figura 15 – RNA especialista em rugosidade para áudio rugoso.



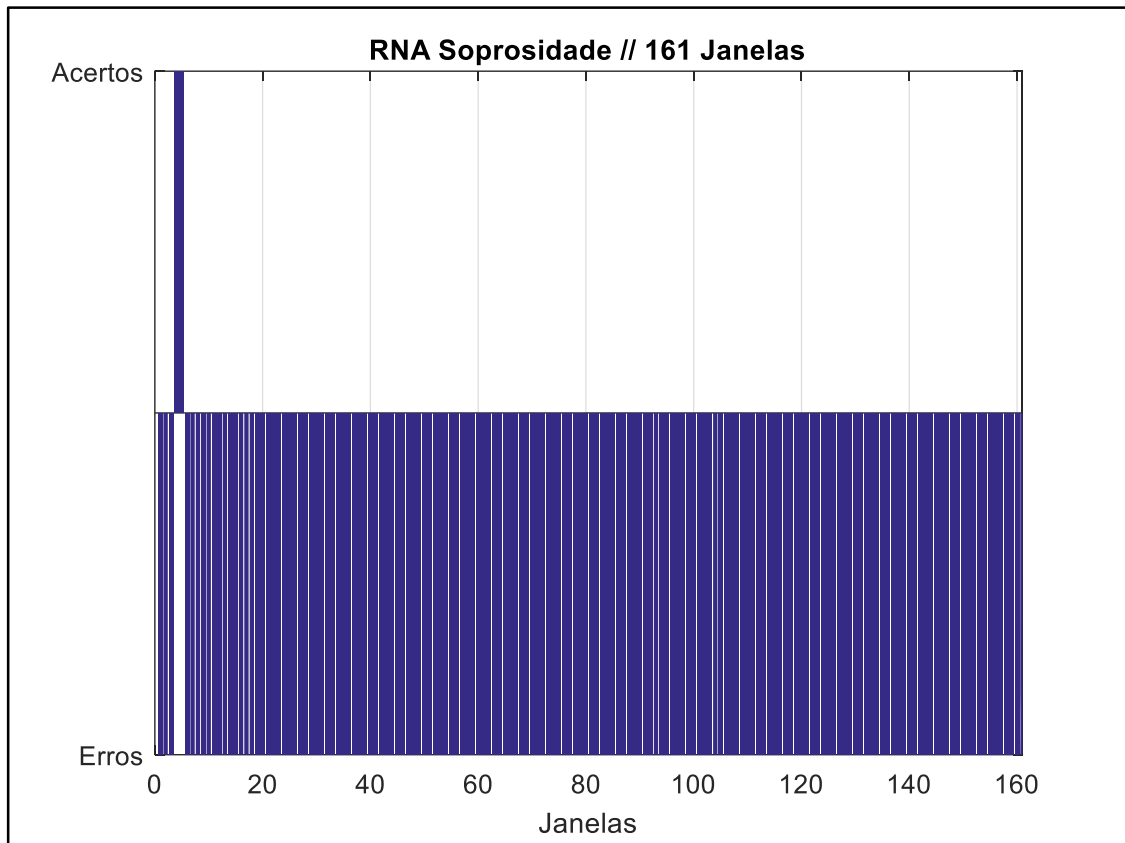
Fonte: Autoria própria.

Quadro 25 – RNA especialista em rugosidade para áudio rugoso.

RESULTADOS PARCIAIS – RUGOSIDADE	
Total de amostras	161
Amostras com rugosidades	157
Amostras não – rugosas	4
Amostras indeterminadas	0
Grau de rugosidade	97,52%

Fonte: Autoria própria.

Figura 16 – RNA especialista em soprosidade para áudio rugoso.



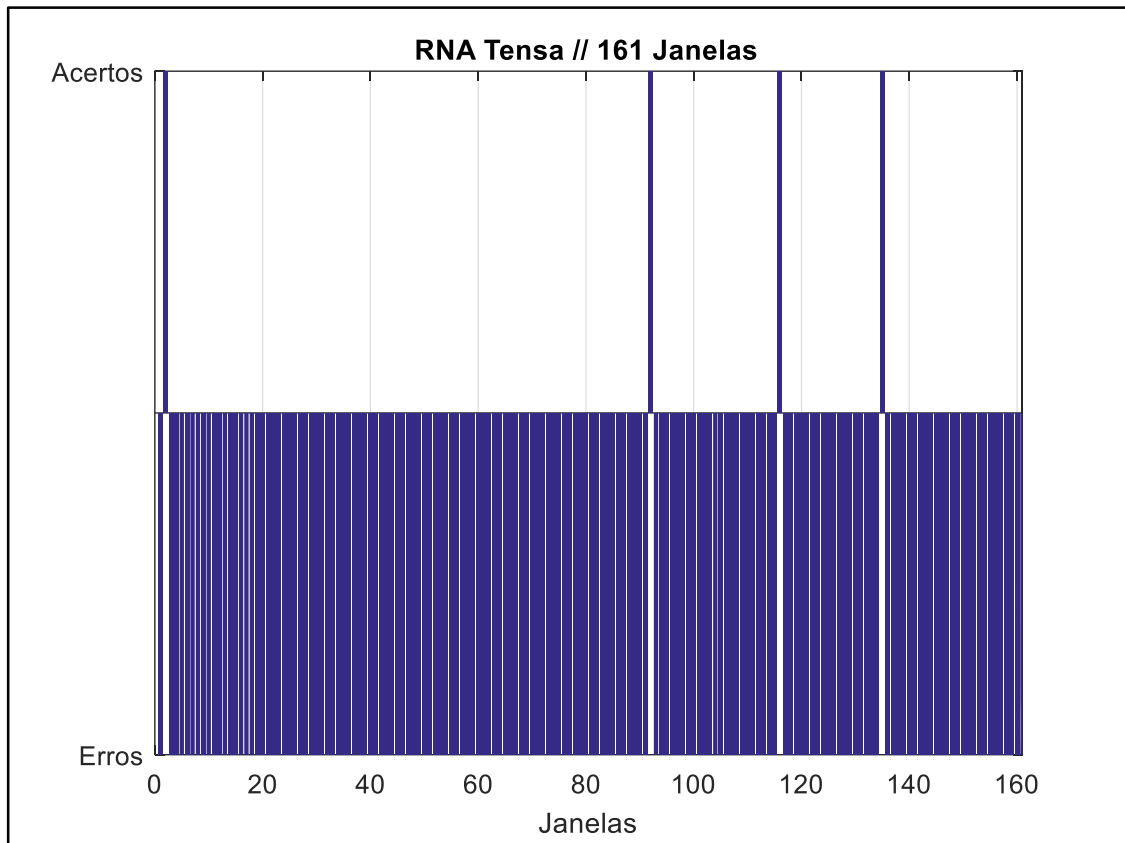
Fonte: Autoria própria.

Quadro 26 – RNA especialista em soprosidade para um áudio rugoso

RESULTADOS PARCIAIS – SOPROSIDADE	
Total de amostras	161
Amostras com soprosidade	2
Amostras não – soprosas	159
Amostras indeterminadas	0
Grau de soprosidade	1,24%

Fonte: Autoria própria.

Figura 17 – RNA especialista em tensão para áudio rugoso.



Fonte: Autoria própria.

Quadro 27 – RNA especialista em tensão para áudio rugoso.

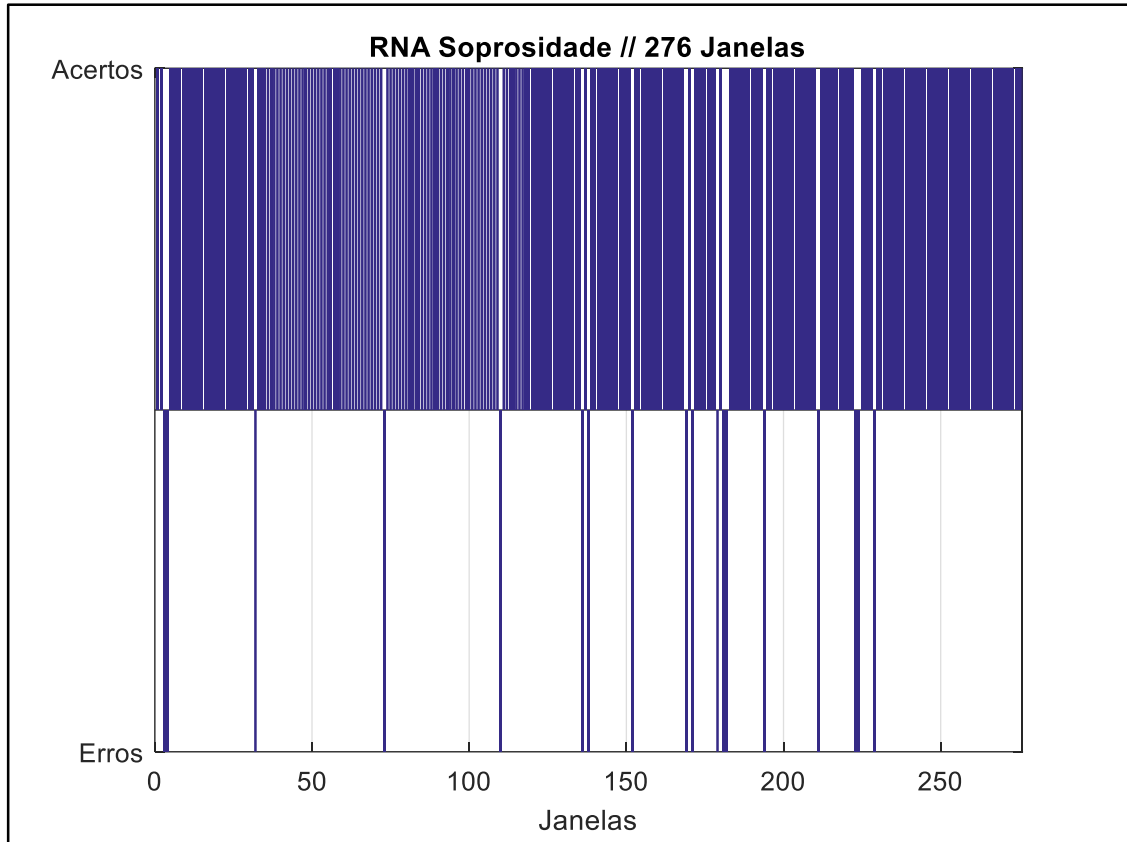
RESULTADOS PARCIAIS – TENSÃO	
Total de amostras	161
Amostras com tensão	4
Amostras não – tensas	157
Amostras indeterminadas	0
Grau de tensão	2,48%

Fonte: Autoria própria.

5.3.2 Análise de áudio soproso

Os resultados da análise do áudio soproso para cada uma das RNAs especialistas estão representados graficamente nas Figuras 18, 19 e 20, bem como seus resultados nos Quadros 28, 29 e 30:

Figura 18 – RNA especialista em soprosidade para áudio soproso.



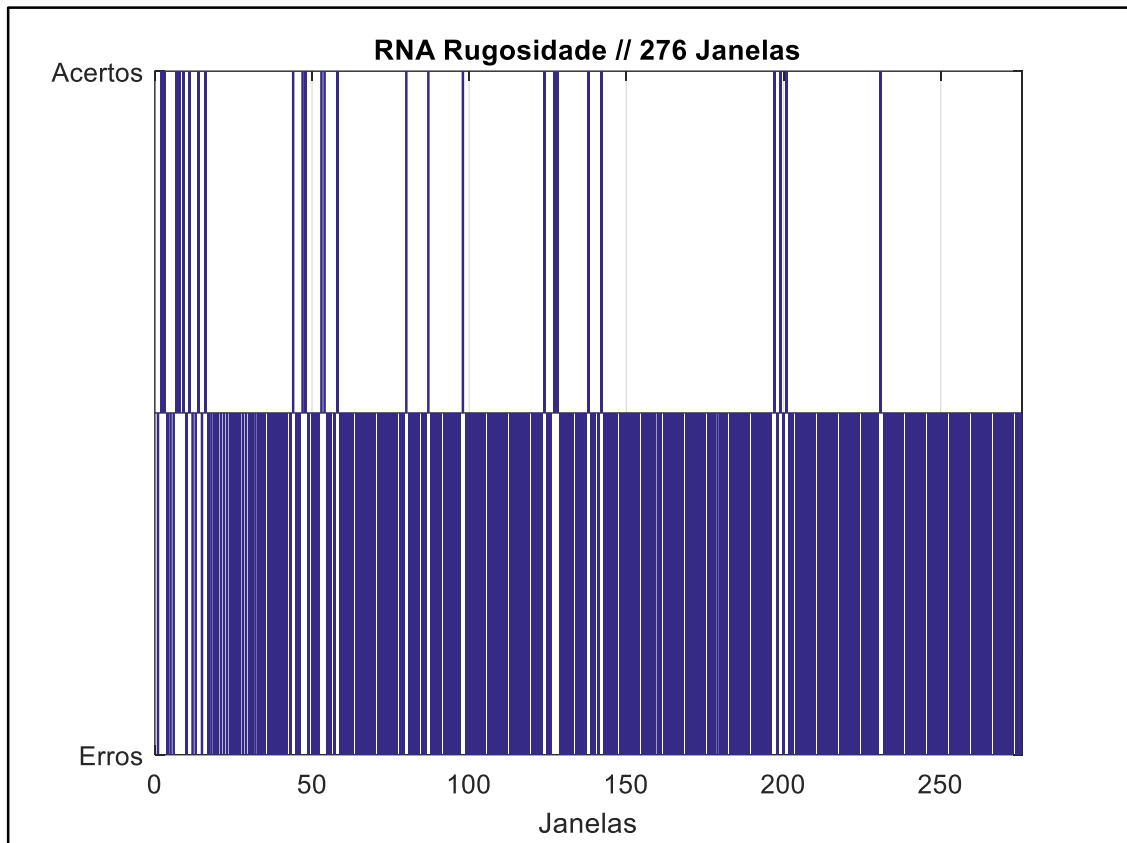
Fonte: Autoria própria.

Quadro 28 – RNA especialista em soprosidade para áudio soproso.

RESULTADOS PARCIAIS – SOPROSIDADE	
Total de amostras	276
Amostras com soprosidade	258
Amostras não – soprosoas	14
Amostras indeterminadas	4
Grau de soprosidade	93,48%

Fonte: Autoria própria.

Figura 19 – RNA especialista em rugosidade para áudio soproso.



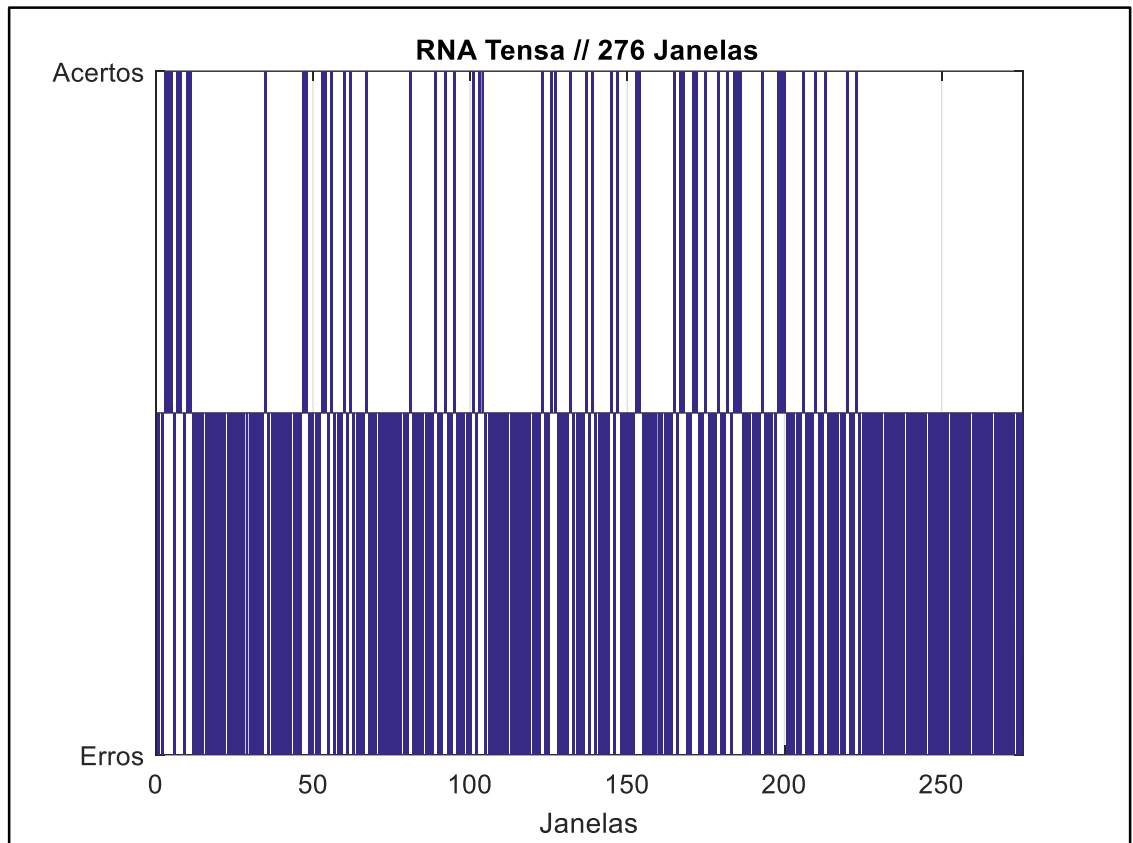
Fonte: Autoria própria.

Quadro 29 – RNA especialista em rugosidade para áudio soproso.

RESULTADOS PARCIAIS – RUGOSIDADE	
Total de amostras	276
Amostras com rugosidade	26
Amostras não – rugosas	250
Amostras indeterminadas	0
Grau de rugosidades	9,42%

Fonte: Autoria própria.

Figura 20 – RNA especialista em tensão para áudio soproso.



Fonte: Autoria própria.

Quadro 30 – RNA especialista em tensão para áudio soproso.

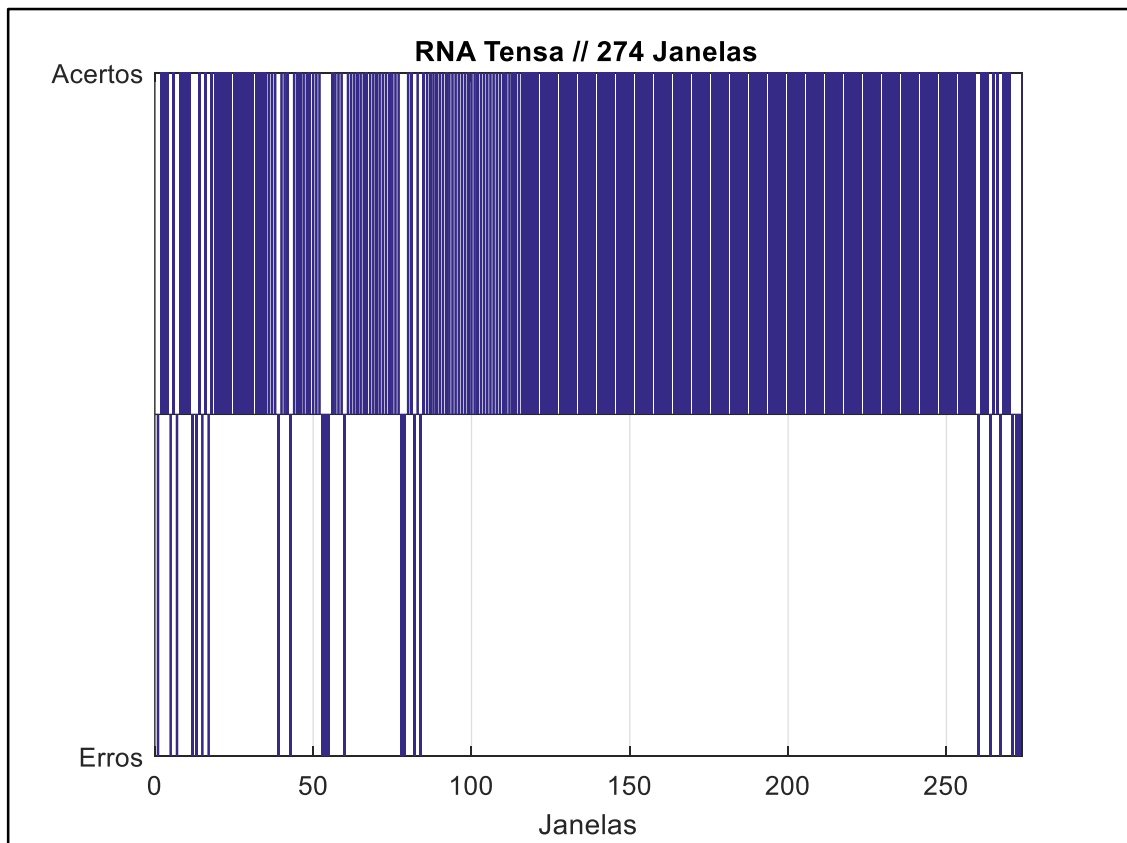
RESULTADOS PARCIAIS – TENSÃO	
Total de amostras	276
Amostras com tensão	53
Amostras não – tensas	214
Amostras indeterminadas	9
Grau de tensão	19,20%

Fonte: Autoria própria.

5.3.3 Análise de áudio tenso

Os resultados da análise do áudio tenso para cada uma das RNAs especialistas estão representados graficamente nas Figuras 21, 22 e 23, bem como seus resultados nos Quadros 31, 32 e 33:

Figura 21 – RNA especialista em tensão para áudio tenso.



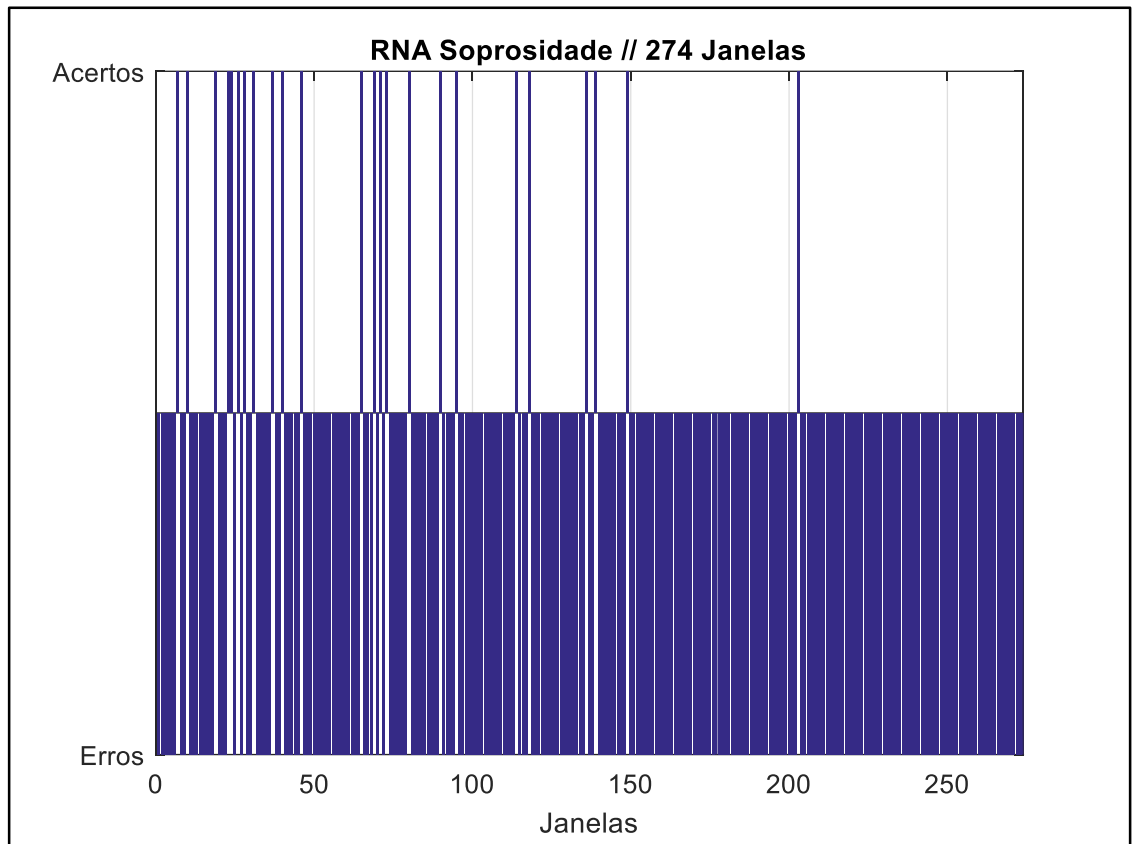
Fonte: Autoria própria.

Quadro 31 – RNA especialista em tensão para áudio tenso.

RESULTADOS PARCIAIS – TENSÃO	
Total de amostras	274
Amostras com tensão	251
Amostras não – tensas	16
Amostras indeterminadas	7
Grau de tensão	91,61%

Fonte: Autoria própria.

Figura 22 – RNA especialista em soproidade para áudio tenso.



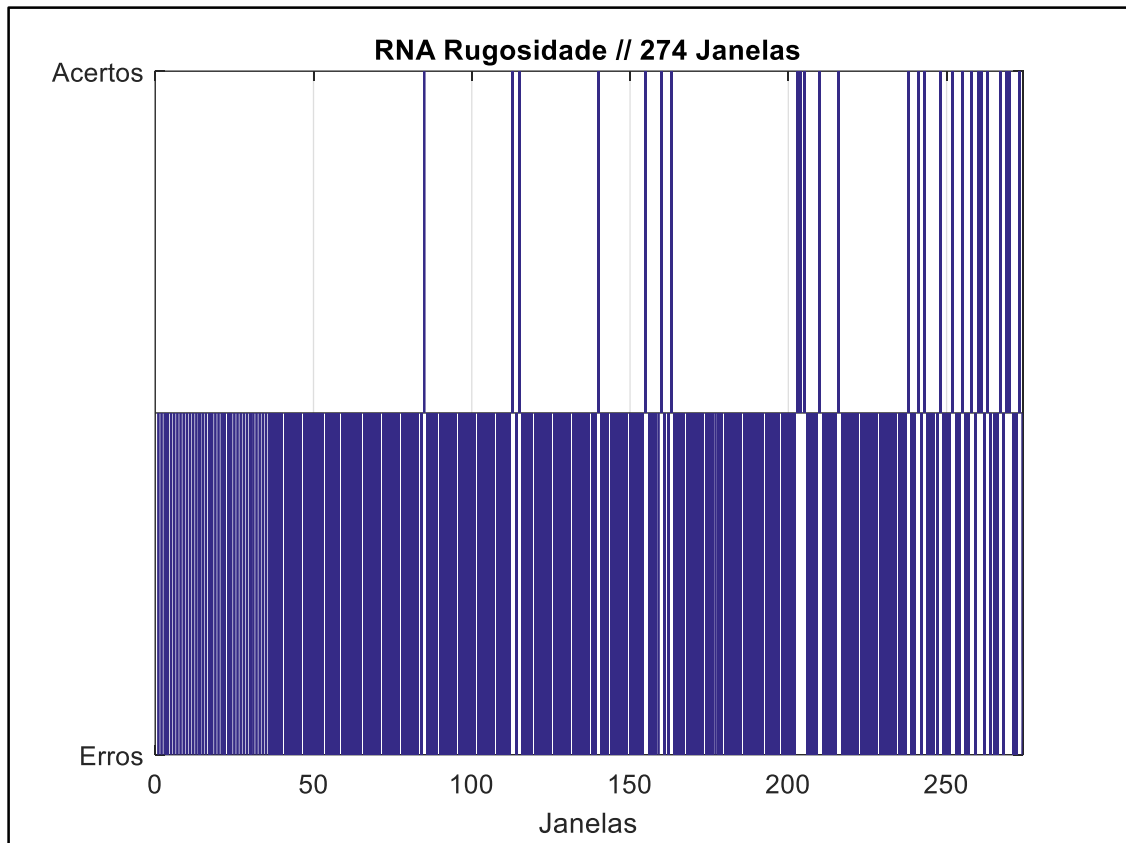
Fonte: Autoria própria.

Quadro 32 – RNA especialista em soproidade para áudio tenso.

RESULTADOS PARCIAIS – SOPROSIDADE	
Total de amostras	274
Amostras com soproidade	24
Amostras não – soprosas	249
Amostras indeterminadas	1
Grau de soproidade	8,76%

Fonte: Autoria própria.

Figura 23 – RNA especialista em rugosidade para áudio tenso.



Fonte: Autoria própria.

Quadro 33 – RNA especialista em rugosidade para áudio tenso.

RESULTADOS PARCIAIS – RUGOSIDADE	
Total de amostras	274
Amostras com rugosidade	26
Amostras não – rugosas	245
Amostras indeterminadas	3
Grau de rugosidade	9,49%

Fonte: Autoria própria.

5.3.4 Discussão das análises realizadas

Reunindo-se os dados das análises anteriores, tem-se os seguintes graus estimados apresentados nos Quadros 34, 35 e 36:

Quadro 34 – Estimação dos parâmetros para um áudio rugoso.

RESULTADOS – ESTIMAÇÃO DO GRAU (ÁUDIO RUGOSO)	
Grau de rugosidade (predominante)	97,52 %
Grau de soproidade	1,24 %
Grau de tensão	2,48 %

Fonte: Autoria própria.

Quadro 35 – Estimação dos parâmetros para um áudio soproso.

RESULTADOS – ESTIMAÇÃO DO GRAU (ÁUDIO SOPROSO)	
Grau de rugosidade	9,42 %
Grau de soproidade (predominante)	93,48 %
Grau de tensão	19,20 %

Fonte: Autoria própria.

Quadro 36 – Estimação dos parâmetros para um áudio tenso.

RESULTADOS – ESTIMAÇÃO DO GRAU (ÁUDIO TENSO)	
Grau de rugosidade	9,49 %
Grau de soproidade	8,76 %
Grau de tensão (predominante)	91,61 %

Fonte: Autoria própria.

Observando-se os valores encontrados na estimação, é possível notar que a variação nos indicadores de redes não especialistas ao grau predominante no arquivo de áudio é superior as taxas de erro encontradas na fase de treinamento, mostrando que, mesmo operando como classificador, os erros encontrados apresentam características de outros parâmetros. Ou seja, ao optar-se por três redes separadas especialistas, foi possível realizar uma tentativa de estimação não somente do parâmetro predominante, mas também do grau dos outros parâmetros presentes no áudio.

6 CONSIDERAÇÕES FINAIS

A integração entre a engenharia e a fonoaudiologia se fez necessária para realização desse trabalho, possibilitando a criação de uma ferramenta computacional capaz de classificar e quantificar os parâmetros subjetivos de distúrbios vocais.

Este trabalho apresentou um método de aplicabilidade de estimação dos parâmetros subjetivos de rugosidade, soprosidade e tensão, que seja robusto e prático a ponto de estimar o seu grau em relação a amostra de sinais apresentadas por meio da configuração correta das RNAs, de forma a auxiliar os profissionais da voz em seus diagnósticos quanto ao grau de desvio vocal desses parâmetros na avaliação perceptivo-auditiva.

Tal estudo foi fundamentado em trabalhos anteriores, buscando, por meio da classificação de parâmetros, obter o grau subjetivo em relação à condição vocal do indivíduo analisado, utilizando-se das medidas de energia dos coeficientes da WPT no quarto nível de resolução.

Com o desafio da ferramenta computacional e ausência dos resultados quantitativos em sua base de dados, realizou-se um mecanismo de comparação para áudios desconhecidos e as RNAs especialistas, utilizando como base que uma RNA especialista em tal parâmetro, ao deparar-se com uma amostra de áudio qualquer, assume que sua especialidade deve conter o número de janelas do sinal adquirido como desconhecido num total de 100% do parâmetro em questão, portanto, sua especialidade apenas relaciona os acertos do classificador, que, por sua vez, compara a quantidade de janelas que foram classificadas como acertos com sua totalidade, gerando, assim, um grau em porcentagem desse parâmetro.

Visto que a avaliação perceptivo-auditiva é realizada por meio de profissionais da voz que possuem aptidão audível acusada para detectar os parâmetros subjetivos, por meio da escala CAPE-V os parâmetros não são complementares até 100% de distúrbios deparados, podendo ser 80% rugoso, 60% soproso e 20% tenso. Portanto o método de estimação elaborado toma como base apenas os acertos referentes ao parâmetro e não um complemento dos parâmetros para se encontrar os demais.

Como principais contribuições deste trabalho, destacam-se: a elaboração de um sistema computacional eficaz, que seja capaz de estimar graus de parâmetros subjetivos pertencentes a sinais vocais implementados na escala visual analógica CAPE-V, que além de auxiliar o avaliador em seu diagnóstico, seja também capaz de nortear o auto treinamento audível dos clínicos desse método de avaliação.

A partir do estudo realizado, é possível motivar trabalhos futuros para o aprimoramento de busca de novas topologia para as RNAs, em um método diferente de processamento da amostra de áudio, de extrações de outras características dos coeficientes das wavelets, utilizando-se o mesmo banco de dados e tendo como referencias os resultados já consolidados.

Deve-se ressaltar que o estudo elaborado não descarta a importância e aplicação dos métodos convencionais de diagnóstico. É proposto apenas um método para complementar as análises acústicas e auxiliar os profissionais das áreas médicas.

REFERÊNCIAS

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

BEHLAU, Mara, et al. **Análise perceptivo-auditiva de vozes rugosas e soprosas: correspondência entre a escala visual analógica e a escala numérica.** São Paulo, São Paulo, 2015.

BEHLAU, Mara. **A voz. O livro do especialista. Vol. I.** Rio de Janeiro, RJ: Revinter, 2001.

BISHOP, Christopher M. **Pattern Recognition and Machine Learning.** New York, US: Springer, 2007.

CERCEU, Janaína da Silva Berto; ALVES, Cláudia Fernanda Tolentino; GAMA, Ana Cristina Côrtes. **Análise acústica da voz de mulheres idosas.** CEFAC, São Paulo, 2008.

Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ASHA Special Interest Division 3, Voice and Voice Disorders. Printed in the US, 2002 -2006. Disponível em:
<<http://www.asha.org/uploadedFiles/members/divs/D3CAPEVprocedures.pdf>>.
Acesso em: 24 out. 2016.

CROVATO, César David. P. **Classificação de sinais de voz utilizando a transformada wavelet packet e redes neurais artificiais.** 143 f. Dissertação de mestrado. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

CUERVO, Luciane. **Introdução à Fisiologia da Voz.** Departamento de Música – UFRGS. Práticas Vocais para a Educação Musical. Porto Alegre. 2010. Material produzido para fins didáticos.

DA SILVA, Ivan N.; SPATTI, Danilo. H.; FLAUZINO, Rogério. A. **Redes Neurais Artificiais para engenharia e ciências aplicadas.** São Paulo, SP: Artliber, 2010.

DAJER, Maria E. **Análise de sinais de voz por padrões visuais de dinâmica vocal.** 154 f. Tese. Universidade de São Paulo, São Carlos, 2010.

FERMINO, Murillo A. et al **Classificação de Distúrbios Vocais Aplicando Redes Neurais Artificiais e Transformada Wavelet Packet**. 12th IEEE/IAS International Conference on Industry Applications – INDUSCON. Curitiba – PR. Brazil, 2016.

FREITAS, Susana V. **Correlação entre avaliação acústica e perceptual na caracterização de vozes patológicas**. 132 f. Relatório do Estado-da-Arte. Faculdade de Engenharia Universidade do Porto, Porto, 2010.

GAMA, Ana Cristina C.; COUTO, Thalita Evaristo; MARTINS, Priscila Campos. **Avaliação perceptivo-auditiva do grau de desvio vocal: correlação entre escala analógica e escala numérica**. Minas Gerais, Belo Horizonte, 2015.

HOGIKYAN, Norman D.; SETHURAMAN, Girish. **Validation of an instrument to measure voice-related quality of life (V-RQOL)**. Journal of voice. v. 13, n. 4, p. 557–569, 1999.

HSU, Hwei P. **Signals and systems**. Schaum outline series. New York, US: McGraw-Hill, 1995.

ORTIZ Karin Zazo; CARRILLO, Luciane. **Comparação entre análises auditivas e acústicas nas disartrias**. Revista Soc Bras Fonoaudiol, 2008.

PARRAGA, Adriane. **Aplicação da transformada wavelet packet na análise e classificação de sinais de vozes patológicas**. 163 f. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

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

SANTOS, Mikaelle Oliveira. **Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet**. 64 f. Dissertação. Instituto Federal de Educação, Ciência e Tecnologia da Paraíba. João Pessoa. 2015.

SILVA, Adelaide Hercília Pescatori. **Para a descrição fonético-acústica das líquidas no português brasileiro: dados de um informante paulistano**. Instituto de Estudos da Linguagem, Unicamp, Campinas, São Paulo, 1996

TEIXEIRA, J. P.; FERREIRA, D. B.; CARNEIRO, S. M. **Análise acústica vocal – Determinação do Jitter e Shimmer para diagnóstico de patologias da fala**. Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Bragança, Portugal, 2011.

TITZE, I. R. **Princípios da Produção Vocal**. Salt Lake City, UT: National Center for Voice & Speech, 2013.

UMAPATHY, Karthikeyan, et al. **Discrimination of pathological voices using a time-frequency approach**. Biomedical Engineering, IEEE Transactions on 52.3: p. 421-430, 2005.

Yu, P.; Revis, J.; Wuyts, F.L.; Zanaret, M.; Giovanni, A. **Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analogue scale**. Folia Phoniatr Logop. 2002.

ZHANG, Yudong, et al. **Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEP-SVM)**. Entropy 17.4: p. 1795-1813, 2015.

ZITTA, Simone Maganhotto. **Análise perceptivo-auditiva e acústica em mulheres com nódulos vocais**. 131 f. Dissertação de Pós – Graduação. CEFET, Paraná, Curitiba, 2005.

Laringoestroboscopia Laríngea. InstitutoOtosur. Disponível em: <<http://otorrinohuelva.es/unidad-de-voz/>>. Acesso em 01 jun. 2017.