Instituto Federal de Educação, Ciência e Tecnologia da Paraíba
Programa de Pós-Graduação em Engenharia Elétrica

Dissertação de Mestrado

Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet

Mikaelle Oliveira Santos

João Pessoa – PB

Abril – 2015

Instituto Federal de Educação, Ciência e Tecnologia da Paraíba
Programa de Pós-Graduação em Engenharia Elétrica

Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet

Mikaelle Oliveira Santos

Dissertação de Mestrado apresentada à Coordenação do Programa de Pós Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba como requisito necessário para obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Processamento de Sinais.

Suzete Élida Nóbrega Correia, D.Sc.
Orientadora

Silvana Luciene do Nascimento Cunha Costa, D.Sc.
Co-Orientadora

João Pessoa, Paraíba, Brasil
10 de Abril de 2015

Dados Internacionais de Catalogação na Publicação – CIP
Biblioteca Nilo Peçanha – IFPB, *campus* João Pessoa

S237a Santos, Mikaelle Oliveira.
Análise acústica de desvios vocais infantis utilizando a transformada Wavelet / Mikaelle Oliveira Santos. – 2015.
79 f. : il.
Dissertação (Mestrado em Engenharia Elétrica) – Instituto Federal de Educação, Ciência e Tecnologia da Paraiba – IFPB / Coordenação de Pós-Graduação em Engenharia Elétrica, 2015.
Orientadora: Suzete Élida Nóbrega Correia, D.Sc.
Co-Orientadora: Silvana Luciene do Nascimento Cunha Costa, D.Sc.
1. Engenharia elétrica. 2. Processamento Digital de Sinais. 3. Desordens Vocais. 4. Transformada Wavelet. I. Título.

CDU 621.391

Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet

Mikaelle Oliveira Santos

Suzete Élida Nóbrega Correia, D.Sc.
Orientadora
(IFPB)

Silvana Luciene do Nascimento Cunha Costa, D.Sc.
Co-Orientadora
(IFPB)

Francisco Madeiro Bernardino Junior, D.Sc.
Membro da Banca
(UNICAP)

Leonardo Wanderley Lopes, Dr.
Membro da Banca
(UFPB)

Luiz Guedes Caldeira, D.Sc.
Membro da Banca
(IFPB)

João Pessoa – PB, Abril de 2015

Lista de Siglas e Abreviaturas

Ac – Medida de acurácia

AMDF – *Average Magnitude Difference Function*

ANN - Redes Neurais Artificiais

BBA – Algoritmo Best Basis

BBT – *Best Basis Tree*

Db – Wavelet de Dabechies

EAV – Escala Analógico-Visual

EN – Energia Normalizada

Esp – Especificidade

F0 – Frequência Fundamental

F1 – Primeiro Formante

F2 – Segundo Formante

F3 – Terceiro Formante

FN – Falso Negativo

FP – Falso Positivo

GG1 – Grau Geral 1 (grupo de sinais de vozes com grau geral normal)

GG2 – Grau Geral 2 (grupo de sinais de vozes com grau geral leve)

GG3 – Grau Geral 3 (grupo de sinais de vozes com grau geral moderado)

GG2 e GG3 – Grau Geral 2 e Grau Geral 3 (grupo de sinais de vozes alteradas)

H – Entropia de Shannon

LDA – Função Discriminante Linear

LS-SVM – *Least Square Support Vector Machines*

MFCC – Coeficientes Cepstrais de Frequência Mel

QDA – Função Discriminante Quadrática

RUG – Grupo de sinais de vozes com a disfonia Rugosidade

SDL – Grupo de sinais de vozes Saudáveis

Sen – Sensibilidade

SOP – Grupo de sinais de vozes com a disfonia Soprosidade

STFT – *Short Time Fourier Transform*

SVM – Máquina de Vetor de Suporte

TWD – Transforma Wavelet Discreta

VN – Verdadeiro Negativo

VP – Verdadeiro Positivo

WPT – Transformada Wavelet Packet

Lista de Figuras

Lista de Tabelas

Sumário

Aos Meus pais Inaldete e Adilson e Meu esposo Ítalo Arthur.

Agradecimentos

- ★ A Deus, Senhor da vida, por tudo que eu pude vivenciar até hoje, pelas pessoas que conheci, e por tudo que ainda está por vir;
- ★ Aos meus pais, Inaldete e Adilson, por todo amor, educação, carinho e paciência para comigo. Ao meu esposo, Ítalo Arthur, pelo incentivo, companheirismo e paciência com meus momentos de ausência e aos meus irmãos, Kleiton e Kleilton, por todo apoio;
- ★ À Professora Suzete Correia, minha Orientadora, um agradecimento carinhoso, por todos os momentos de paciência, dedicando parte do seu tempo, desde os últimos anos, para compartilhar comigo seus valiosos conhecimentos, não apenas na área acadêmica, mas também dando conselhos e ensinando valores humanos;
- ★ À Professora Silvana Costa, um agradecimento especial, por sempre ter acreditado em mim, aceitando tal papel nesta pesquisa, pelos ensinamentos e orientações em sala de aula, e por todas as conversas e conselhos dados;
- ★ Ao Professor Leonardo Lopes, membro da Banca, por ter disponibilizado, em nome do Departamento de Fonoaudiologia da Universidade Federal da Paraíba, o banco de dados com as vozes infantis. Além disso, por ter aceitado fazer parte desta Banca, bem como por compartilhar os seus valiosos conhecimentos ao longo desta pesquisa;
- ★ Aos Professores Francisco Madeiro e Luis Caldeira, membros da Banca, por aceitar avaliar este trabalho, de forma a compartilhar os seus valiosos conhecimentos e acrescentar mais valor a esta pesquisa;
- ★ Aos colegas do Mestrado, pela torcida, pelo conhecimento compartilhado, pelas conversas e palavras de motivação. Aos amigos pioneiros do Mestrado em Engenharia Elétrica do IFPB, tais como Sérgio, Vinícius, Leidiane, com os quais pude aprender muito, e em especial à Taciana, que me acolheu em sua casa, nos momentos em que precisei de abrigo por morar em uma outra cidade;
- ★ Ao Professor Jefferson Costa e Silva, Coordenador do Programa de Pós Graduação em Engenharia Elétrica (PPGEE) do IFPB, e a todos os Professores do Colegiado do Programa;
- ★ Ao Professor Carlos Danilo Miranda Regis, pelo incentivo para estar aqui hoje, o meu muito obrigada.

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”

(Arthur Schopenhauer)

Resumo

Distúrbios da voz podem atingir diferentes faixas etárias, afetando a qualidade vocal, prejudicando a comunicação por meio da voz. Técnicas de processamento digital de sinais de voz podem ser empregadas para auxiliar outros métodos de avaliação de distúrbios da voz, tais como análise otorrinolaringológica e análise perceptivo-auditiva. Crianças com distúrbios de voz podem apresentar efeitos negativos no seu desenvolvimento social, educacional e físico. A investigação e o diagnóstico precoce do desvio vocal infantil permite maior eficácia no tratamento. Entretanto, a avaliação de desordens vocais em crianças apresenta alguns desafios relacionados às dificuldades de cooperação das mesmas durante os exames tradicionais. Nesta pesquisa, as medidas de energia e entropia dos coeficientes de detalhe da transformada wavelet são empregadas na avaliação da qualidade vocal em crianças. Dois estudos de caso são abordados nesta pesquisa: 1) Análise acústica do grau da intensidade do desvio vocal; e 2) Análise acústica da qualidade vocal predominante (rugosidade e soprosidade). As medidas de energia e entropia dos coeficientes de detalhe da transformada wavelet são utilizadas de maneira individual e combinada a fim de se obter uma maior eficácia na classificação dos sinais. Para o primeiro estudo de caso, utilizando-se de um vetor híbrido de medidas combinadas, foram obtidas acurácias acima de 95% e, para o segundo, utilizando-se também do vetor de medidas combinadas, as medidas de acurácia foram superiores a 90%. Os sinais das vozes desviadas apresentaram elevação em suas frequências dos formantes, comparados às vozes sem desvio. Os resultados obtidos nesta pesquisa indicam que o uso das medidas de energia e entropia dos coeficientes de detalhe da transformada wavelet mostra-se como uma técnica promissora, que pode ser considerada para ser empregada como uma ferramenta para análise acústica da qualidade vocal em crianças.

Palavras-Chave: Processamento Digital de Sinais de Voz, Desordens Vocais, Energia, Entropia, Transformada Wavelet.

Abstract

Voice disorders may target different age groups, affecting voice quality, impairing communication through voice. Digital processing techniques for speech signals can be used to assist other evaluation methods of voice disorders, such as analysis ENT and perceptual analysis. Children with voice disorders may present negative effects on their social, educational and physical development. The research and early diagnosis of a child dysphonia allows greater treatment efficacy. However, the evaluation of vocal disorders in children presents some challenges related to the their difficulties to cooperate in traditional tests. In this research, energy and entropy measures of the wavelet transform detail coefficients are employed to evaluate children?s dysphonia. Two studies of case are covered in this research: 1) Acoustic analysis of the degree of intensity of vocal deviation; and 2) Acoustic analysis of the predominant voice quality (hoarseness and breathiness). Energy and entropy measures of wavelet transform detail coefficients are used individually and combined in order to obtain greater accuracy. For the first case of study, using a hybrid vector of combined measures, accuracies above 95% were obtained and in the second case, also using the combined vector of measures, the accuracy values were greater than 90%. Signs of disordered voices showed an increase in their frequency of formants compared to the voices without deviation . The results obtained in this study indicate that the use of energy and entropy measures of the wavelet detail coefficients is shown as a promising technique , which can be considered to be used as a tool for acoustic analysis of voice quality in children.

Key-Words: Digital Processing of Speech Signals, Voice Disorders, Energy, Entropy, Wavelet Transform.

Capítulo 1

Introdução

1.1 – Motivação

O homem utiliza diversos meios de comunicação para desenvolver a sua capacidade intelectual e o seu meio social. A fala é a principal ferramenta para o convívio entre as pessoas, pois com ela é possível expressar os sentimentos e ideias, além de possibilitar a troca de informações.

O sistema vocal, apesar de pequeno, possui uma capacidade de produção complexa e potente. Sua representação máxima está focada nas pregas vocais. O trato vocal atua como um filtro e suas frequências de ressonância designam-se por formantes. As vogais são reconhecidas pelos seus formantes, que são produzidos em nível glótico e modificados pelos ajustes específicos do trato vocal [1].

Os distúbios ou desvios da voz, podem afetar diferentes grupos etários. Muitas desses desvios o ser humano traz consigo desde o seu nascimento, sendo diagnosticadas ainda na infância, por meio da identificação de dificuldades respiratórias ou choro anormal ou de forma tardia, por meio de manifestações sutis que ocorrem ao longo do crescimento [2] [3].

O sistema de produção vocal infantil possui uma complexidade estrutural menor que a adulta, pois nesta fase, diversos órgãos como a laringe ainda estão em formação [3]. No entanto, o sinal vocal infantil é mais complexo e instável. As bases anatômicas e fisiológicas da laringe infantil são relativamente pouco conhecidas se comparadas às bases da laringe adulta. O tamanho e o formato do trato vocal são fatores determinantes das características do som a ser emitido e dependem diretamente da idade e sexo do falante [4].

Esses distúrbios, em crianças, podem ser causadas por diversos fatores, tais como: patologias (de origem orgânica, neurológica ou genética), abuso vocal (gritos, cantos excessivos, fala excessiva, entre outros comportamentos inerentes à faixa etária) e fatores psicogênicos, tais como distúrbios emocionais, problemas familiares e traumas físicos [2].

1.2 – Justificativa

O diagnóstico da qualidade vocal inicialmente é feito pelo otorrinolaringologista, que inclui a anamnese, seguido de exames físicos e visual da laríngе, a exemplo da videolaringoscopia

direta, videoestroboscopia e eletromiografia, exames esses de caráter invasivo, que podem trazer desconforto ao paciente [5].

A videolaringoscopia direta é um exame realizado pelo médico com o objetivo de visualizar a laringe utilizando uma microcâmera. A videoestroboscopia permite a visualização do comportamento vibratório das pregas vocais, 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 [6] [7].

Técnicas de processamento digital de sinais tem sido desenvolvidas para avaliar a qualidade vocal, bem como avaliar quantitativamente a intensidade do desvio vocal (rugosidade, soprosidade, tensão e instabilidade) através da análise acústica. Essas, são técnicas automáticas de auxílio diagnóstico, menos invasivas e de baixo custo, comparadas àquelas baseadas em exames videolaringoscópicos [5]. A terapia vocal, realizada pelos fonoaudiólogos, inclui a audição da voz do paciente e análise acústica da voz.

Crianças com distúrbios de voz podem apresentar efeitos negativos no seu desenvolvimento social, educacional e físico [8]. A investigação e o diagnóstico precoce do desvio vocal infantil permite maior eficácia no tratamento. Entretanto, a avaliação de desordens vocais em crianças apresenta alguns desafios relacionados às dificuldades de cooperação das mesmas durante os exames tradicionais.

Clínicos e pesquisadores têm buscado novas medidas discriminativas, de caráter não invasivo, que sejam capazes de imprimir uma boa avaliação da qualidade vocal, bem como o seu diagnóstico e monitoramento do tratamento. A literatura ainda não traz um consenso a cerca das medidas de maior acurácia na avaliação dessas desvios vocais. Por isso, se fazem necessários estudos que possam revelar o poder de discriminação das medidas acústicas de maneira isolada e/ou combinadas para serem empregadas na discriminação entre vozes saudáveis/alteradas. Uma alteração das frequências dos formantes da voz, por exemplo, podem indicar algum tipo de desvio vocal.

A extração de características do sinal de voz que representem bem o desvio vocal que se pretende investigar é de fundamental importância para uma classificação mais acurada do tipo e do grau de intensidade do desvio vocal, para acompanhamento do processo de terapia fonoaudiológica.

Uma classificação eficiente pode auxiliar o terapeuta a avaliar o quanto a terapia está sendo efetiva, de forma objetiva. Para tanto, é necessário que a técnica proposta tenha confiabilidade e apresente as informações das mudanças ocorridas no sinal antes e após a terapia vocal, necessárias para um diagnóstico mais preciso.

Apesar de haver muitos trabalhos relacionados à identificação de distúrbios da voz, não há uma confirmação precisa de um método que seja capaz de encontrar os parâmetros mais adequados para modelagem de uma patologia em particular. Muitas dessas pesquisas são

focadas na discriminação entre laringes saudáveis e patológicas de adultos, sem discriminar entre tipos de desvio vocal e seus graus de intensidade em crianças [7] [9] [10] [11].

A discriminação de distúrbios da voz ainda é objeto de investigação mais precisa por parte dos pesquisadores. Portanto, o estudo de técnicas de análise acústica é uma área bastante promissora, uma vez que a interdisciplinaridade dos procedimentos pode proporcionar a investigação com mais precisão de um distúrbio da voz [12].

A transformada wavelet fornece uma análise dos sinais em diferentes resoluções, de forma que, em cada uma delas, diferentes aspectos dos sinais podem ser observados. Características obtidas a partir da análise wavelet têm sido empregadas na avaliação de desordens vocais em adultos [13] [14] [15], causadas por patologias laríngeas. Para a população pediátrica, no entanto, ainda há poucos trabalhos relacionados [16].

Apesar de ser uma técnica relativamente recente, a transformada wavelet, tem apresentado resultados significativos na discriminação entre vozes normais e patológicas, [15], [17], [18], [19], [20], [21], [22]. A energia e a entropia do sinal associada às faixas de frequência dos diferentes níveis de resolução das wavelets podem apontar uma desordem vocal. [23] [24].

No tocante à aplicação de técnicas de processamento digital de sinais voz no monitoramento da qualidade vocal, não foi encontrada, na literatura, nenhuma pesquisa que relacione as medidas de Energia e Entropia com a avaliação do grau de desvio fonatório em crianças e a classificação da qualidade vocal predominante.

A alta prevalência de desvios vocais na infância exige uma atenção especial na avaliação e diagnóstico de vozes infantis, sugerindo o desenvolvimento de medidas objetivas que proporcionem a compreensão da intensidade do desvio vocal e sua manifestação em diferentes períodos entre os 3 e 9 anos de idade [25].

1.3 – Objetivos

1.3.1 – Objetivo Geral

Avaliar o desempenho da Energia Normalizada e da Entropia dos coeficientes de detalhe da Transformada Wavelet em nove níveis de resolução, na avaliação da intensidade do desvio vocal e da qualidade vocal predominante em crianças.

1.3.2 – Objetivos Específicos

- Empregar técnicas de classificação de padrões tal como análise discriminante, para discriminar entre os graus de intensidade do desvio vocal em vozes infantis e a qualidade vocal predominante;

- Avaliar o potencial discriminativo das medidas de Energia Normalizada e Entropia dos coeficientes de detalhe da Transformada Wavelet entre os graus de intensidade do desvio fonatório dos sinais de vozes infantis e entre tipos de qualidades vocais;
- Avaliar diversas bases wavelets para determinar a família que melhor se adequa ao problema em questão;
- Identificar uma medida ou um conjunto de medidas combinadas que melhor caracterizem os distúrbios de voz considerados.

1.4 – Organização do Trabalho

Este documento está organizado da seguinte forma: o Capítulo 2 trata da análise acústica dos sinais de vozes, descrevendo o mecanismo de produção da fala baseado no modelo linear e ressalta os formantes como modelo de análise acústica para classificação entre tipos de desvios vocais. No Capítulo 3 é apresentada a ferramenta matemática utilizada no desenvolvimento desta pesquisa, a Transformada Wavelet. No Capítulo 4 é apresentada a metodologia empregada na pesquisa. No Capítulo 5, estão apresentados os resultados obtidos e sua discussão e, no Capítulo 6, encontram-se as considerações finais e as sugestões para trabalhos futuros.

Capítulo 2

Análise Acústica dos Sinais de Voz

Do ponto de vista fisiológico, a voz humana é o resultado da interação de órgãos de diferentes sistemas do corpo humano [26], um conjunto de estruturas do trato vocal, cujas partes mais intimamente associadas à produção do som são os pulmões, a traqueia, a laringe, a faringe as cavidades nasais e a cavidade oral [6].

O trato vocal possui uma capacidade de produção complexa e potente. Sua representação máxima está focada nas pregas vocais. A voz é utilizada tanto para comunicação, quanto para expressar emoções, pensamentos e sentimentos, para satisfazer suas necessidades, além de representar a identidade de cada indivíduo, sendo considerada tão pessoal quanto à impressão digital.

Neste capítulo, são apresentados diversos aspectos da voz, tais como: o processo de produção, os conceitos de voz normal e voz desviada, avaliação perceptivo-auditiva da qualidade vocal, com as escalas mais utilizadas, análise acústica dos sinais de voz e as medidas acústicas do sinal de voz mais comumente utilizadas.

2.1 – O Processo de Produção da Voz

A fonação é uma função neurofisiológica inata, mas a voz forma-se ao longo da vida, baseada nas características anatomofuncionais do indivíduo, bem como nos aspectos emocionais de sua história pessoal. Assim sendo, a voz é o resultado da fonação acrescida de ressonância [26].

A Figura 2.1 ilustra a anatomia do aparelho fonador. Os pulmões, brônquios e traqueia produzem o “ar”, matéria prima da produção vocal; a laringe (onde se encontram as pregas vocais) produz a energia da fala e, a faringe, fossas nasais e boca são responsáveis pela ressonância.

Os sons sonoros ocorrem quando o fluxo de ar sai dos pulmões e atinge a traqueia até alcançar a laringe, produzindo uma vibração nas pregas vocais. Diferente dos sons sonoros, os sons surdos não provocam vibrações, pois quando o fluxo de ar atinge a traqueia as pregas vocais estão relaxadas.

Na produção de sons orais, o véu palatino está levantado e o fluxo de ar é irradiado pela boca e na produção de sons nasais o véu palatino está abaixado e a cavidade oral fechada (lábios, dentes, palato), assim, o fluxo de ar é radiado pelas narinas [27].

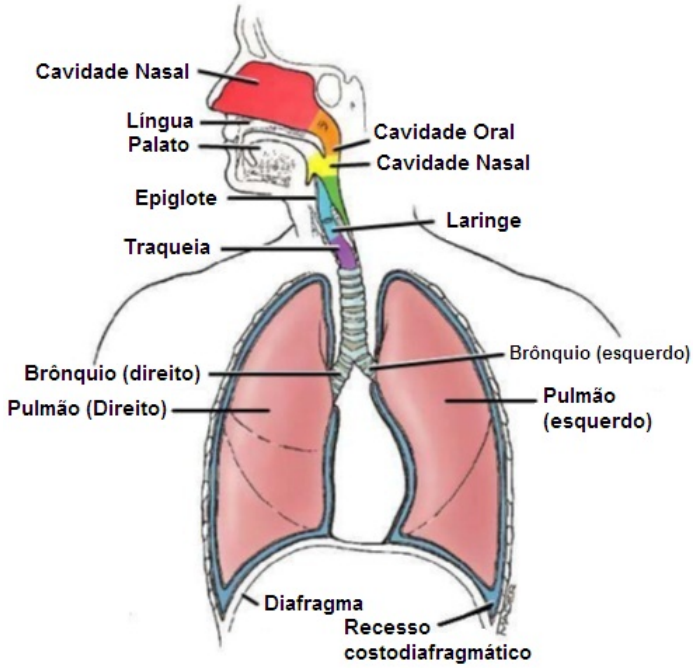


Figura 2.1 – Anatomia do aparelho fonador.
Fonte: fttp://www.medicalexcel.com (adaptação).

A laringe é um órgão tubular, um arcabouço esquelético membranoso, situada no plano mediano e anterior superficial do pescoço. Comunica-se inferiormente com a traqueia e superiormente com a faringe [28]. As funções básicas da laringe, em ordem de importância são proteção ds vias aéreas inferiores, respiração e fonação.

As pregas vocais são duas dobras de músculos, ligamentos e mucosas que se estendem horizontalmente na laringe. Na Figura 2.2, são ilustrados os processos de abdução (afastamento) e adução (fechamento) das pregas vocais que ocorrem durante a fonação. Uma desordem nesse movimento, pode acarretar o surgimento de alguns tipos de desordens vocais [29].

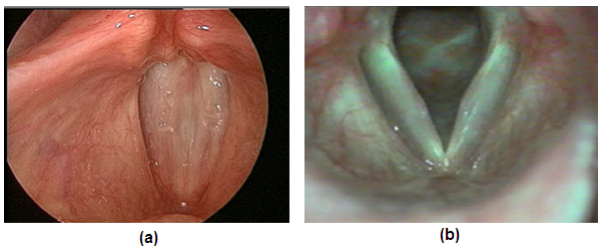


Figura 2.2 – Pregas vocais em: (a) adução e (b) abdução - visão endoscópica. Fonte: [30].

As bases anatômicas e fisiológicas da laringe infantil são relativamente pouco conhecidas se comparadas às da laringe adulta. No entanto, sabe-se que a laringe infantil não corresponde a uma miniatura da laringe do adulto, uma vez que as diferenças entre elas não se restringem apenas ao tamanho (Figuras 2.3 e 2.4). O tamanho e o formato do trato vocal são

fatores determinantes das características do som a ser emitido e dependem diretamente da idade e sexo [4].

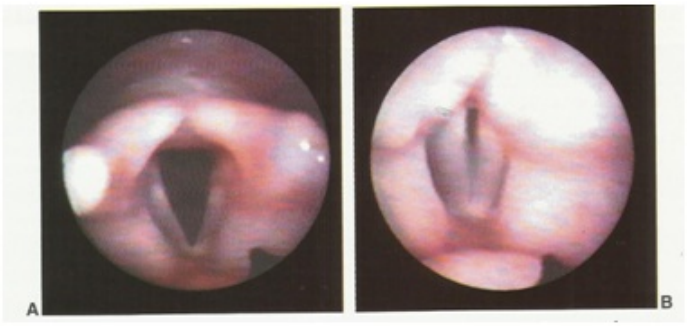


Figura 2.3 – Imagens da laringe infantil, obtidas por nasolaringoscopia. A. Durante a respiração. B. Durante a fonação. Fonte: [26].

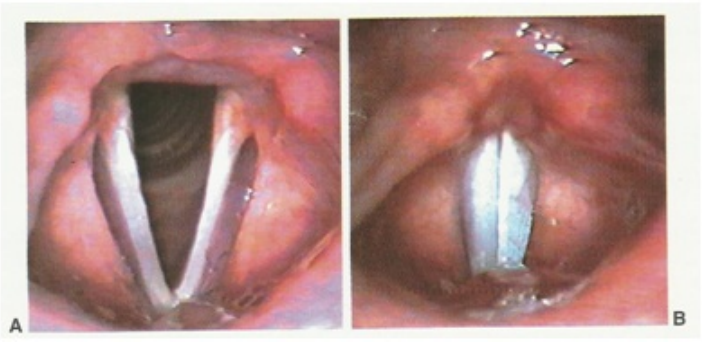


Figura 2.4 – Imagens da laringe adulta, obtidas por telelaringoscopia. A. Durante a respiração. B. Durante a fonação. Fonte: [26].

No início da vida, a laringe se apresenta muito alta e seguindo ao desenvolvimento orgânico, ela inicia sua descida em relação à posição no pescoço, o que continua por toda a vida, permanecendo na mesma posição entre os 15 e 20 anos e segue descendo discretamente durante a terceira idade. A consequência direta a esse fato é o alongamento do tubo de ressonância que pode amplificar melhor as frequências graves [31].

Na seção a seguir, serão apresentados os aspectos relativos à diferença entre voz normal e voz desviada, bem como as implicações das mesmas em crianças, suas causas e os desvios ou distúrbios da voz estudos neste trabalho.

2.2 – Voz Normal x Voz desviada

A literatura não apresenta consenso quanto aos conceitos de voz normal e voz desviada. Não existe uma definição aceitável de voz normal e não há padrões nem limites definidos [32].

Desordens vocais podem afetar diferentes grupos etários. Muitas desses desvios vocais podem ser diagnosticadas ainda na infância, por meio da identificação de dificuldades

respiratórias ou choro anormal ou, ainda, de forma tardia, por meio de manifestações sutis que ocorrem ao longo do crescimento [2] [3]. Em crianças, estima-se que a taxa de prevalência de desvios vocais está entre 6% a 23%, aproximadamente [33] [8].

Behlau & Pontes [31] conceituam desvio vocal como um distúrbio da comunicação oral, no qual a voz não consegue cumprir o seu papel básico de transmissão da mensagem verbal e emocional de um indivíduo.

Nesse contexto, desvio vocal ou distúrbio da voz, é considerado um sintoma presente em vários e diferentes distúrbios da voz, ora se apresentando como sintoma secundário, ora como principal. O desvio da voz tanto pode apresentar-se como o sintoma mais importante de uma desordem ou doença, quanto como um sintoma discreto inserido num quadro de outras doenças a exemplo do mal de Parkinson.

A alta prevalência de desvios vocais na infância exige uma atenção especial na avaliação e diagnóstico de vozes infantis, com o desenvolvimento de medidas objetivas que proporcionem a compreensão da intensidade do desvio vocal e sua manifestação em diferentes períodos entre os 3 e 9 anos de idade [25]. A análise acústica pode ser empregada como um método de apoio ao diagnóstico e tratamento de desvios vocais, de forma rápida e confortável.

Dois desses desvios, por estarem atreladas a diversos tipos de patologias e acometerem grande parte do público infantil, foram escolhidas para serem estudas neste trabalho. São elas: rugosidade e soprosidade.

Na seção que se segue, serão apresentadas as escalas que medem a qualidade vocal através da análise perceptivo-auditiva, além de mostrar como esses e outros tipos de distúrbio da voz são classificadas.

2.3 – Avaliação Perceptivo-Auditiva da Qualidade Vocal

A avaliação da voz é uma das componentes principais do diagnóstico vocal e precede a intervenção terapêutica. Normalmente é realizada de acordo com um protocolo contendo duas componentes: a avaliação de acordo com parâmetros perceptivos, também designada de avaliação perceptiva, e a análise de acordo com parâmetros objetivos, também designada de avaliação acústica [34].

No primeiro caso, o especialista (fonoaudiólogo), observa as características sonoras da voz, de acordo as referências perceptivas, adquiridas pelo especialista durante a sua formação ou exercício profissional, de vozes categorizadas como normais. Existem procedimentos de avaliação padronizados que permitem quantificar a intensidade das perturbações percebidas.

A avaliação perceptivo-auditiva pode ser de caráter exclusivamente impressionístico (voz rouca, soprosa, áspera, etc.), e envolver escalas e índices para uma determinação menos subjetiva e mais confiável do desvio encontrado.

Segundo Pontes *et al.* [35] existem diferenças espectrográficas marcantes entre as vozes roucas e ásperas das vozes saudáveis. Os harmônicos estão presentes em grande quantidade

e com melhor definição nas vozes saudáveis, com uma média de alcance nas vozes femininas de 4.868,6 Hz e nas masculinas de 4.242,6 Hz. Já nas vozes ásperas estas faixas alcançaram a média de 2.145,6 Hz no sexo feminino e no masculino de 2.104,6 Hz, representando praticamente a metade da média dos normais; nos roucos a média superior foi de 1.311,6 Hz para os casos de vozes femininas e de 983,3 Hz para as masculinas, representando mais de um quarto do resultado das vozes normais.

De acordo com Martens *et al.* [36], 70 vozes de pacientes com diversas patologias foram avaliados e, dentre outros resultados, percebeu-se que a presença de ruído na faixa de 1500 a 4500 Hz está correlacionada a soprosidade. Os autores em um estudo sobre a correlação feita entre parâmetros acústicos, perceptivo-auditivos, aerodinâmicos e anatômicos, avaliando 87 vozes de pacientes disfônicos [37], foram encontradas relações significantes entre ruídos em altas frequências no espectro e impressão perceptivo-auditiva de soprosidade na voz.

A literatura traz uma série de escalas para avaliação auditiva da voz, com emprego de diferentes tarefas para a avaliação perceptivo-auditiva da qualidade vocal. Dentre as diferentes escalas abordadas pela literatura para utilização na clínica vocal, serão abordadas duas delas: a escala GRBAS [38], e a escala visual analógica [39].

Escala GRBAS

A escala GRBAS, (G = avaliação do grau global do desvio vocal (grade); R = rugosidade (*roughness*); B = soprosidade (*breathiness*); A = astenia (*asteny*); S = tensão (*strain*) [38], usada internacionalmente, é um método simples de avaliação do grau global do desvio vocal pela identificação da contribuição de quatro fatores independentes: rugosidade, soprosidade, astenia e tensão, considerados os mais importantes na definição de uma voz disfônica. Ressaltando que apenas os fatores astenia e tensão são excludentes entre si [26]. Os fatores indicados, são definidos como [40]:

- Rugosidade: irregularidade de vibração das pregas vocais. Engloba o conceito de rouquidão, crepitação, bitonalidade e também aspereza. Assim, a voz é percepcionada com ruídos presentes em baixa frequência, com característica rugosa e ruidosa. Este parâmetro verifica-se em casos de: fenda glótica, presença isolada de uma alteração orgânica ou fenda de qualquer dimensão com alterações da mucosa das pregas vocais (exemplo: nódulos, pólipos ou edemas).

- Soprosidade: presença de ruído de fundo, audível, que corresponde fisiologicamente à fenda glótica (abertura entre as pregas vocais).

- Astenia: relacionada com o mecanismo de hipofunção das pregas vocais e reduzida energia de emissão do som. Exemplo: *miastenia gravis* ou outras perturbações neurológicas do controle vocal.

- Tensão: associada a esforço vocal por aumento da adução glótica (hiperfunção), geralmente inerente ao aumento da atividade da musculatura extrínseca da laringe, com elevação desta. Exemplo: disfonia espasmódica e síndromes de abuso vocal com consequente alteração da mucosa (i.e. nódulos ou pólipos).

Os parâmetros avaliados são classificados em uma escala de 4 pontos: 0 = normal ou ausência de desvios; 1 = ligeiro desvio ou discretas modificações; 2 = desvio moderado ou alterações evidentes; 3 = desvio severo/grave ou com variações extremas. São também contemplados valores intermédiarios. Esta é uma escala de triagem vocal que se aplica sobre a fonte glótica durante a produção de vogais sustentadas (/a/ ou /ε/) ou fala encadeada [40].

Os resultados são anotados com os níveis de avaliação subscritos ao lado das iniciais dos fatores. Assim sendo, exemplificando, um indivíduo com desvio vocal em grau global moderado, caracterizada por rugosidade moderada, soprosidade discreta, sem astenia e sem tensão, seria classificada como $G_2R_2B_1A_0S_0$.

Escala EAV

Outra forma de se estabelecer os graus de intensidade do desvio vocal é através da escala analógico-visual ou EAV. Escalas analógico-visuais (EAV) são amplamente utilizadas na área de saúde, particularmente na enfermagem, para a mensuração de fenômenos subjetivos como dor, ansiedade, náusea, fadiga e dispneia.

Tais escalas correspondem a uma linha de 100mm, vertical ou horizontal, na qual o paciente, ou o avaliador, é orientado a marcar a quantidade de sensação experienciada no momento. Cada milímetro corresponde a um grau de desvio e, portanto, a escala oferece 100 possibilidades de graduação.

A EAV é geralmente ancorada por termos que representam os extremos (ausente e máximo) ou graus intermediários (leve, médio e intenso) dos fenômenos subjetivos [41] [42]. Não existe um limite específico para definir uma voz como normal, mas reconhece-se uma faixa de distribuição de normalidade vocal [26] [32].

Um estudo realizado por Yamasaki [43] reproduziu no Brasil o estudo Finlandês de Simberg [39], para definir o critério de diferenciação entre variações normais da qualidade vocal e alterações vocais por análise perceptivo-auditiva, concluindo que o valor de 35,5 pontos (Tabela 2.1), em uma EAV de 100 pontos (Figura 2.5) seria o critério de diferenciação, sendo que vozes assinaladas acima deste ponto representam falha na triagem vocal e deveriam ser encaminhadas para avaliação médica.

Essas escalas avaliam o sinal de voz de maneira perceptivo-auditiva, tornando-se uma avaliação subjetiva. Essas técnicas visuais resultam em uma avaliação qualitativa, de resultados difíceis de serem quantificados, e necessitam do conhecimento e da experiência do avaliador [6] [7].

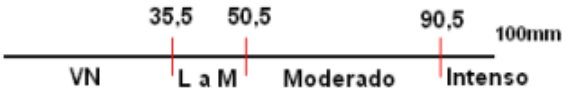


Figura 2.5 – Régua de graduação na escala analógico-visual, com base nos respectivos valores de corte, de acordo com a análise perceptivo-auditiva. [43].

Tabela 2.1 – Faixas de distribuição dos graus de desvio vocal, em pontos.

Grau de Desvio Vocal	Faixa de Desvio
Variabilidade Normal	0 a 35,5
Leve a Moderado	35,6 a 50,5
Moderado a Intenso	50,6 a 90,5
Intenso	90,6

Fonte: [43]

Para auxiliar o diagnóstico médico, técnicas de processamento digital de sinais podem ser desenvolvidas para avaliar a qualidade vocal, bem como avaliar quantitativamente a intensidade do desvio vocal (rugosidade, soprosidade, tensão e instabilidade) através da análise acústica [5].

A seção a seguir apresenta a análise acústica dos sinais de voz, seus objetivos e como ela pode ser utilizada na diferenciação entre vozes normais e disfônicas.

2.4 – Análise Acústica dos Sinais de Voz

A análise acústica de sinais de voz tem como objetivo quantificar e caracterizar um sinal sonoro, possibilitando a integração de dados fornecidos pela avaliação perceptivo-auditiva com o plano fisiológico. Tal método, permite um detalhamento do processo de geração do sinal sonoro, fornecendo uma estimativa indireta dos padrões vibratórios das pregas vocais, bem como dos formatos do trato vocal e das modificações nestes formatos [29].

Quando usada no âmbito do estudo da voz, a análise acústica permite, de forma não invasiva, comparada aos exames laringoscópicos usuais, determinar e quantificar a qualidade vocal do indivíduo através dos diferentes parâmetros acústicos que compõem o sinal: periodicidade, amplitude, duração e composição espectral. Constituindo-se, assim, um método de avaliação objetiva que permite, entre outras utilidades, um diagnóstico precoce de problemas vocais.

Clínicos e pesquisadores tem buscado, constantemente, medidas discriminativas de caráter não invasivo, que sejam capazes de imprimir uma boa avaliação da alteração vocal, bem como o seu diagnóstico e monitoramento do tratamento.

Por meio da análise acústica, os atributos físicos da voz são analisados no domínio do tempo, da frequência e da intensidade, além de outras medidas complexas, que conjugam do cruzamento de tais domínios [1].

Historicamente, o século XX marca o período moderno da análise acústica. As primeiras análises iniciaram-se com o oscilógrafo, em 1920, que produzia gráficos relacionando a amplitude do som e o tempo [26].

Na década de 40, foi desenvolvido o espectrógrafo sonoro, aparelho que teve implicação revolucionária, por permitir um registro tridimensional do sinal sonoro, integrando os aspectos de tempo, frequência e intensidade num único gráfico de dois eixos, chamado de espectrograma [44].

Somente no início dos anos 70, começaram a operar os primeiros processadores digitais de sinais, com definições mais acuradas e mais claras [45], possibilitando o armazenamento digital, bem como, o surgimento de uma série de outras medidas [26].

As medidas obtidas na análise acústica correspondem a medidas físicas definidas. O sinal glótico (sinal da fonte) sofre efeitos ao longo do trato vocal supraglótico até a saída deste para o meio externo (ação de filtro). Há uma somatória das ondas sonoras provenientes da fonte glótica com outras refletidas ao longo do trato vocal, sendo a resultante final (sinal de saída), o sinal irradiado pelos lábios [46] [47] como pode ser observada na Figura 2.6.

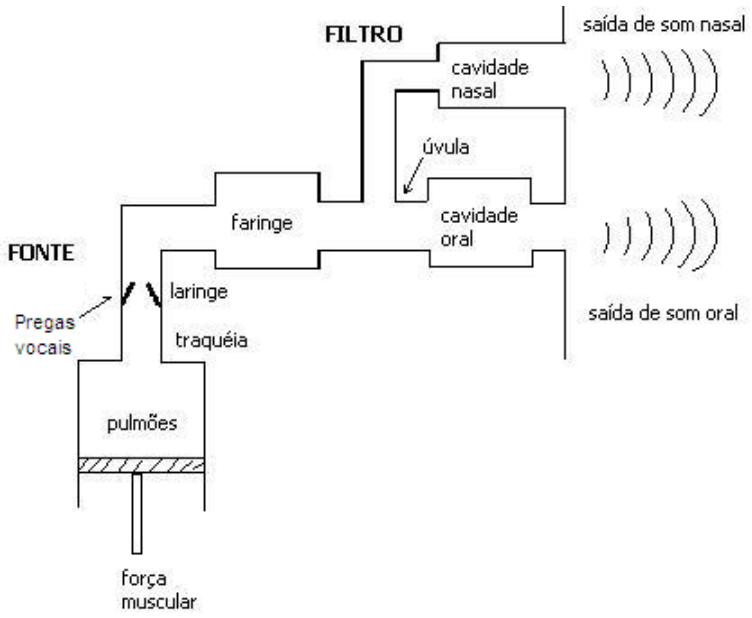


Figura 2.6 – Diagrama de blocos das produção da voz humana. [6].

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 a somatória do sinal glótico mais os efeitos dos filtros. Por este motivo, os instrumentais de análise realizam análises indiretas, 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. As principais medidas da análise acústica vocal são apresentadas na seção a seguir.

2.5 – Medidas Acústicas do Sinal de Voz

Os dados encontrados através da analise acústica são complementares a análise perceptivo-auditiva. Além da percepção do sinal sonoro, a analise acústica permite ao avaliador captar as alterações vocais precoces, sendo também um ótimo recurso para promoção e prevenção da saúde vocal.

Na técnica da análise acústica, são extraídas características do sinal que possam representar bem suas variações, desordens, contendo detalhes do sinal que possam diferenciá-los ou classificá-los de acordo com critérios estabelecidos para os objetivos da análise, tais como: pré-diagnóstico de alterações no funcionamento laríngeo, avaliação da qualidade vocal, redução de ruído, entre outras.

As medidas acústicas geralmente são escolhidas baseadas em análise estatística, verificando o poder discriminatório das mesmas, baseada em análise subjetiva visual dos padrões comportamentais das mesmas, ou empregando técnicas de classificação (redes neurais, máquinas de vetor de suporte, análise discriminante, entre outras).

Frequentemente os desvios vocais mais significativos são caracterizados acusticamente pelos avaliadores e fonoaudiólogos por meio da leitura das representações visuais fornecidas, a exemplo da análise espectrográfica e não apenas pelas medidas numéricas obtidas automaticamente. Tal aspecto destaca a importância da observação e apreciação visual de padrões espectrográficos num primeiro momento, para depois relacioná-los às medidas numéricas obtidas [1].

As principais medias acústicas utilizadas atualmente na detecção de desvios vocais são a frequência fundamental, o *Jitter* e o *Shimmer*. Existem ainda outras características do sinal sonoro capazes de fornecer informações importantes, tais como os formantes, as medidas de ruído, a intensidade, e o tempo máximo de fonação.

Frequência Fundamental (*F0*) - medida mais frequentemente em Hertz, corresponde ao número de vibrações por segundo das pregas vocais, que por sua vez é o equivalente ao primeiro harmônico da emissão [46].

A *F0* reflete a eficiência do sistema fonatório, a biomecânica laríngea e a sua interação com a aerodinâmica, sendo, portanto, um importante parâmetro na avaliação anatômica e funcional da laringe. Esta medida é também usada para distinção entre locutores, uma vez que depende de características físicas do trato vocal tais como comprimento, tensão e massa.

Os valores desta frequência fundamental(*F0*) variam de acordo com a idade, com uma distribuição média de 80 a 250Hz, nos adultos jovens, sendo que nos homens a faixa de frequências varia entre 80 a 150 Hz, nas mulheres de 150 a 250 Hz e em crianças apresentam valores acima de 250 Hz, como pode ser visto na Figura 2.7. [48] [49].

No entanto, estes valores não são estacionários uma vez que, além de variarem com o sexo e a idade, podem depender também, de fatores como o estado de espírito da pessoa, o

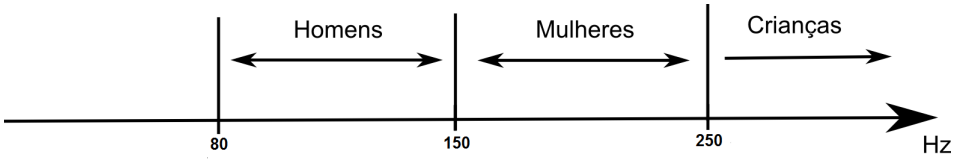


Figura 2.7 – Faixas de normalidade da frequência fundamental para homens, mulheres e crianças.

período do dia em que se enquadram (de manha, à tarde e à noite), os hábitos de vida (alcoolismo e tababagismo), o uso profissional da voz (voz falada e cantada) e os distúrbios da voz.

As medidas da *F0* mais referidas na literatura são a média, a mediana, o desvio padrão, o máximo e o mínimo. A literatura mostra que os indivíduos com patologia apresentam, tendencialmente, uma extensão da *F0* mais restrita e mais baixa. Por essas razões, considera-se que as medidas de variabilidade da *F0* são úteis para a avaliação do grau da patologia vocal.

Vozes com crepitação e roucas tendem a apresentar *F0* grave, enquanto que vozes ásperas geralmente apresentam *F0* aguda. Situações de extrema tensão psicológica podem produzir vozes extremamente agudas.

Existem vários métodos para medição da frequência fundamental [50]. Esta frequência pode ser medida determinando o inverso do intervalo de tempo transcorrido entre dois pulsos glotais sucessivos, ou selecionando a frequência correspondente à primeira harmônica do espectro de frequências.

Outras formas de medição da frequência fundamental são realizadas no domínio do tempo: Método da Função da Média de Diferenças de Amplitudes (AMDF - *Average Magnitude Difference Function*) [50]; Método da função de autocorrelação [50] [51]; Algoritmos que utilizam análise cepstral [52] e Medição a partir do resíduo da análise LPC [53]. A AMDF e a Função de Autocorrelação são mais comumente utilizados.

Jitter - é uma medida ciclo a ciclo e refere-se a pequenas variações involuntárias na frequência fundamental, que permite determinar o grau de estabilidade do sistema fonatório.

O *jitter* altera-se principalmente com a falta de controle de vibração das pregas vocais. Os sinais de vozes de pacientes com patologias vocais apresentam, frequentemente, uma maior porcentagem de *jitter*.

A presença de um pequeno grau de perturbação e irregularidade do sinal vocal é aceitável, uma vez que, fatores de ordem neurológica, emocional e biomecânica, tornam o sinal de voz instável.

A literatura considera como valor típico normal a variação entre 0,5 e os 1,0% para as fonações sustentadas em adultos jovens [38]. O *jitter* altera-se principalmente com a falta de controle da vibração das pregas vocais, como ocorre nas disfonias neurológicas e está correlacionado com a aspereza [26].

Shimmer - é uma medida da irregularidade na amplitude da onda sonora a curto prazo. É muitas vezes referida como a perturbação da amplitude.

O *shimmer*, portanto, mede a variação na intensidade dos ciclos adjacentes de vibração das pregas vocais e altera-se com a redução da resistência glótica e lesões de massa nas pregas vocais, estando correlacionado com a presença de ruído à emissão (rouquidão) e com a soprosidade [26].

2.6 – Formantes

Os pulsos de ar que passam pelas pregas vocais vibram no trato vocal e as ressonâncias aí ocorridas são chamadas de formantes [54]. Os principais correlatos acústicos associados à qualidade vocálica de um segmento são os formantes e a duração.

Os formantes das vogais variam, dependendo das características anatomofuncionais de cada indivíduo e do posicionamento dos órgãos fonoarticulatórios no momento da emissão [55]. O trato vocal infantil é mais curto do que o trato vocal do adulto e, considerado o sexo da criança, observa-se uma diferença nas medidas de comprimento. Tendo como referência o trato vocal adulto masculino, o trato infantil (aos oito anos) apresenta, em média, medidas 25% e 42% menores, para meninos e meninas, respectivamente. Dessa forma, as frequências dos formantes são mais agudas em crianças do que em adultos, e mais agudas em meninas do que em meninos [26].

Os três primeiros formantes de cada vogal são mais representativos no que diz respeito à descrição acústica das vogais [1]. O primeiro formante, denominado F1, depende da abertura da mandíbula, abaixamento da língua, deslocamento vertical da língua e constrição laríngea. O segundo formante, F2, depende do movimento horizontal da língua e elevação posterior da mesma e F3 depende do tamanho da cavidade oral [26].

Uma pesquisa realizada por Behlau *et. al.* [48], com 90 falantes do português brasileiro do Brasil, da cidade de São Paulo, divididos em grupos iguais de ambos os sexos, crianças e adultos jovens, provenientes de três classes socioeconômicas e culturais distintas, apresentam os valores médios dos formantes para homens, mulheres e crianças, saudáveis, cujos resultados encontram-se na Tabela 2.2. Os valores obtidos pela pesquisadora foram extraídos por leitura manual, com o auxílio de uma transparência milimetrada, a partir dos espectrogramas produzidos pelo espectrógrafo de som V.I. 700.

Durante esta pesquisa foi desenvolvido um estudo detalhado dos formantes em vozes infantis com e sem desvio vocal, a fim de investigar o quanto essas frequências podem ser alteradas na presença de algum distúrbio da voz. Para isso, foi utilizada a mesma base de dados utilizada para obter os resultados desta dissertação que está descrita no Capítulo 4. O software Praat foi utilizado para obter as frequências formantes. A análise dos formantes foi dividida em dois estudos de caso: crianças com sinal de voz saudável *x* crianças com desvios vocais (rugosidade e/ou soprosidade) e crianças com qualidade vocal predominante rugosidade *x* crianças com qualidade vocal predominante soprosidade.

Tabela 2.2 – Valores médios em Hertz dos formantes para homens, mulheres e crianças, falantes do português brasileiro da cidade de são Paulo.

Grupos	Formantes	“i”	“ê”	“é ”	“a”	“ô”	“ ó”	“u”
Homens	F1	398	563	699	807	715	558	400
	F2	2.456	2.339	2.045	1.440	1.201	1.122	1.182
	F3	3.320	2.995	2.848	2.524	2.481	2.520	2.452
Mulheres	F1	4,25	6.28	769	956	803	595	462
	F2	2.984	2.712	2.480	1.634	1.317	1.250	1.290
	F3	3.668	3.349	3.153	2.721	2.602	2.668	2.528
Crianças	F1	4,65	698	902	1.086	913	682	505
	F2	3.176	2.825	2.606	1.721	1.371	1.295	1.350
	F3	3.980	3.637	3.243	2.873	2.793	2.823	2.667
Média	F1	4,29	629	790	950	810	612	455
	F2	2.989	2.625	2.337	1.598	1.296	1.226	1.274
	F3	3.656	3.327	3.081	2.706	2.626	2.670	2.549
DP	F1	70,5	101,69	117,3	149,6	126,8	84,3	81,7
	F2	343,0	305,23	315,2	224,3	139,8	171,5	159,6
	F3	371,1	335,26	266,3	302,9	227,3	225,4	221,4

Fonte: [26]

Os resultados obtidos (Tabela 2.3) mostraram que, os valores da frequência fundamental em crianças com a qualidade vocal afetada sofreu alterações em relação as crianças com voz normal. Os formantes F1, F2 e F3, para o grupo de crianças que apresentam algum desvio da qualidade vocal (rugosidade e/ou sorposidade) apresentam valores superiores quando comparado ao grupo de crianças com voz normal, o que evidencia, uma alteração dos formantes do sinal de voz na presença de algum tipo de desvio vocal.

Tabela 2.3 – Valores Médios Para Frequência Fundamental e Formantes em crianças de 3 a 9 anos.

	Voz Normal	**Rugosidade**	**Soprosidade**
Fo	261,098	249,76	237,69
F1	946,907	1.179,617	2.701,647
F2	2.779,737	2.791,850	3.293,284
F3	2.857,796	3.334,040	4.924,548

Quando se compara o grupo de crianças com o desvio soprosidade, com o grupo de crianças com o desvio rugosidade, os valores dos três primeiros formantes, para o grupo com soprosidade apresentam-se mais elevados, mais agudos do que o grupo com rugosidade. Desta forma, pode-se justificar esta elevação nos valores dos formantes, na presença de ar turbulento, presente no desvio vocal soprosidade, que pode estar atrelada a um fechamento glótico insuficiente. No Apêndice A, estão todas as informações referentes ao desenvolvimento desta pesquisa.

2.7 – Considerações Finais do Capítulo

Neste capítulo foram apresentados os aspectos inerentes a produção da voz, trazendo a diferenciação entre o sistema de produção vocal infantil e adulto, principais órgãos responsáveis e como uma má formação nesse sistema pode acarretar o surgimento de desvios vocais.

Foi vista a diferenciação entre voz normal e voz desviada e foram apresentados os distúrbios da voz trabalhados nesta pesquisa, a rugosidade e a soprosidade. No âmbito da avaliação vimos a avaliação perceptivo-auditiva, que necessita de um especialista, e a avaliação acústica, que será utilizada neste trabalho, bem como as principais medidas utlizadas neste tipo de avaliação.

No capítulo seguinte, será apresentado o modelo matemático, para extração de características, utilizado na classificação entre vozes normais e disfônicas, seus graus de severidade e na separação entre rugosidade e soprosidade.

Capítulo 3

Análise Wavelet

A extração de características do sinal de voz, que representem bem o desvio vocal que se pretende investigar, é de fundamental importância para uma classificação mais acurada do tipo e do grau do desvio, para acompanhamento do processo de terapia fonoaudiológica.

Uma classificação eficiente pode auxiliar o terapeuta a avaliar o quanto a terapia está sendo efetiva, de forma objetiva. Para tanto, é necessário que a técnica proposta tenha confiabilidade e apresente as informações das mudanças ocorridas no sinal antes e após a terapia vocal, necessárias para um diagnóstico mais preciso.

Diversos sinais encontrados na natureza possuem características não estacionárias, ou seja, variam com o tempo, tais como os sinais de voz [56]. A Transformada de Fourier é mais adequada para análise de sinais estocásticos estacionários, pois, neste tipo de análise a informação de tempo é perdida e apenas a informação de frequência está presente.

Para que fosse possível analisar o sinal no tempo em pequenas porções, Gabor [57] adaptou a Transformada de Fourier, com uma técnica chamada de janelamento (*windowing*) do sinal. Esta adaptação é conhecida como Transformada de Fourier a Curto Intervalo de Tempo (STFT- *Short Time Fourier Transform*). Nela, o sinal encontra-se em uma função de duas dimensões; tempo e frequência [58]. Contudo, esta informação tem precisão limitada pelo tamanho da janela de análise que, uma vez escolhida, será a mesma para todas as frequências.

Porém, muitos sinais, a exemplo dos sinais de voz, exigem uma aproximação mais flexível, onde o tamanho da janela seja variável, determinando mais precisamente informações sobre tempo ou frequência de um determinado sinal [58].

A transformada wavelet é uma ferramenta matemática, desenvolvida em meados dos anos 80, que surgiu como uma alternativa à Transformada de Fourier para análise tempo-frequência. Uma maneira eficiente de aplicar a Transformada Wavelet Discreta (TWD) é através de filtros, técnica desenvolvida por Mallat [59], que possui propriedades úteis e interessantes para o processamento de sinais, como:

I A possibilidade de usar análise multirresolucional, que permite a análise de sinais em resoluções distintas, de modo que em cada escala aspectos diferentes sejam observados;

II O fato de as wavelets não serem únicas, ou seja, existem na literatura vários tipos dessas funções, que podem ser selecionadas de acordo com a aplicação;

III A representação esparsa dos coeficientes, que é importante para a extração de características, por fornecer apenas um pequeno número de coeficientes não-nulos [60] [61].

Uma outra característica da transformada wavelet é sua alta capacidade de concentrar a energia do sinal em um número reduzido de coeficientes, possibilitando a obtenção de uma representação mais compacta [62].

Muitos dos avanços obtidos nos estudos utilizando transformada wavelet foram desenvolvidos devido à cooperação de Ingrid Daubechies e Stephane Mallat. Daubechies [63] desenvolveu uma família de wavelets com base compacta (*compact support*) e Mallat [59] introduziu a transformada wavelet no conceito de decomposição multirresolução de sinais.

A transformada wavelet é uma ferramenta que permite decompor um sinal em diferentes componentes de frequências, permitindo assim, estudar cada componente separadamente em sua escala correspondente. O termo ‘ *wavelet*’ significa ‘pequena onda’ (*small wave* em inglês ou *ondelette* em francês). O termo ‘pequena’ refere-se à condição de que esta função é de tamanho finito (suportada compactamente) [64].

Neste capítulo são introduzidos os conceitos básicos da decomposição wavelet, fornecendo uma base teórica necessária para a aplicação desta teoria nos próximos capítulos desta dissertação. Além disso, são descritas as características extraídas a partir da decomposição wavelet, utilizadas no desenvolvimento deste trabalho.

3.1 – Famílias Wavelets

Existem diferentes tipos de bases ortonormais e não ortogonais, tais como: Haar, Daubechies (dbs), Symlet (syms), Biortogonais (biors), Coiflet, Mexican Hat, B-splines, entre varias outras, utilizadas na construção das funções wavelet [65]. Algumas dessas famílias podem ser visualizadas na Figura 3.1.

A obtenção de melhores resultados em determinadas aplicações tornou-se fundamental para a escolha destas bases. Em processamento digital de sinais, sabe-se que as wavelets de Daubechies possuem características especiais que as tornam mais utilizadas, trazendo resultados de grande importância científica [60].

As wavelets de Daubechies são uma família formada por várias funções, que possuem 45 ordens de filtros de comprimentos diferentes [14]. Tais wavelets são ortogonais e possuem suporte compacto. Segundo [66], as wavelets de Daubechies de ordem 40 são indicadas para análise de desordens vocais.

Neste trabalho, foram analisados o desempenho das 45 wavelets de Daubechies, além das wavelets biortogonais a fim de identificar a que apresentava maior grau de acurácia nas classificações, destacaram-se nesse estudo as wavelets de Daubechies, e dessa formas, esta foi a família escolhida para o desenvolvimento desta pesquisa.

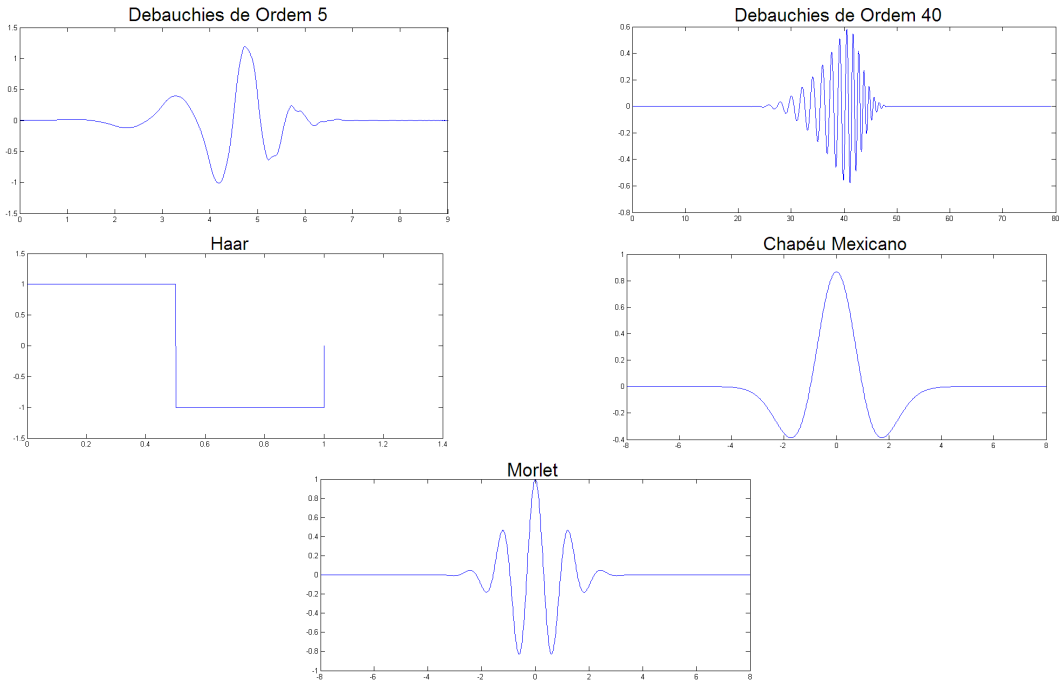


Figura 3.1 – Algumas Famílias Wavelets

3.2 – Decomposição Wavelet

A transformada wavelet consiste na decomposição de um sinal $f(t)$ através de uma família de bases, reais e ortonormais [17]. A função base usada na transfomada wavelet é localizada tanto no tempo como na frequência. Todas as funções wavelet são versões geradas por dilatações e translações de uma função protótipo $\psi(t)$, também conhecida como wavelet "mãe", dada por [63]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a})a, b \in R \tag{3.1}$$

em que os parâmetros $a > 0$ e b são chamados parâmetros de escalonamento e translação respectivamente e $a^{-\frac{1}{2}}$ o fator de normalização que mantém a mesma energia para todas as wavelets independente da escala utilizada.

Quando o fator de escala $a > 1$, a wavelet encontra-se expandida proporcionando a análise em baixas frequências do sinal. Do contrário, quando $a < 1$, as wavelets encontram-se comprimidas e permitem uma análise em altas frequências. Para ser considerada uma wavelet, uma função também tem que atender as seguintes propriedades [63]:

i A área total sob a curva da função é 0, ou seja, $\int_{-\infty}^{+\infty} \psi(t)dt = 0$

ii A energia da função é finita, ou seja, $\int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty$

Essas condições são equivalentes a dizer que $\psi(t)$ é quadrado integrável ou que pertence ao conjunto das funções quadrado integráveis. As propriedades acima sugerem que $\psi(t)$ tende a oscilar acima e abaixo do eixo *t*, e que tem sua energia localizada em uma certa região, já que é finita. Essa característica de energia concentrada em uma região finita é que diferencia a análise usando wavelets da análise de Fourier, já que esta última utiliza as funções periódicas seno e cosseno [17].

A transformada wavelet contínua de um sinal $f(t)$, em que função $f(t) \in L^2\mathbb{R}$, é definida como a correlação entre a função $f(t)$ e a família wavelet $\psi_{a,b}(t)$ para cada *a* e *b* é, dada por [58]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \int f(t)\psi*(\frac{t-b}{a})dt, \tag{3.2}$$

em que o parâmetro de escalonamento a fornece a largura da wavelet, indica a posição e $\psi*(t)$ é o complexo conjugado de $\psi(t)$. Na Figura 3.2 podem ser observadas a wavelet Morlet em diferentes escalas.

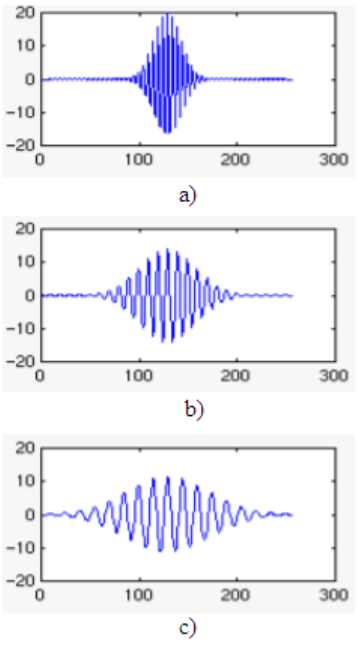


Figura 3.2 – Wavelet Morlet em diferentes escalas. a) wavelet comprimida, b) wavelet mãe e c) wavelet expandida. Fonte: [66].

A transformada wavelet contínua permite uma análise dos sinais de voz por meio de escalogramas, uma representação tempo-frequência do sinal [67] [68]. Na Figura 3.3 podem ser observadas a resolução tempo-frequência para a transformada de Fourier de curto tempo (STFT) e para a transformada wavelet. O módulo ao quadrado da transformada wavelet é definido como escalograma wavelet e mostra como a energia do sinal varia com o tempo e com a frequência. Os padrões obtidos pelo escalograma dependem da família wavelet empregada. Na avaliação de desordem vocais a wavelet Chapéu Mexicano tem sido comumente usada [67]. As Figuras 3.4,

3.5 e 3.6, ilustram os escalogramas de uma voz saudável, uma voz com desvio vocal rugosidade e uma voz com o desvio soprosidade, respectivamente.

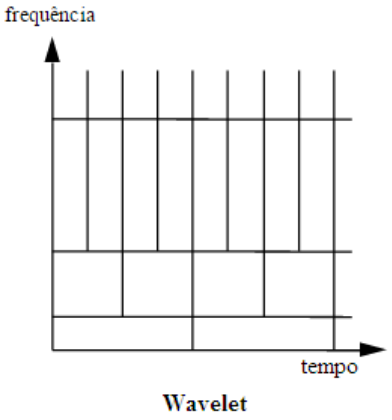


Figura 3.3 – Resolução Tempo-Frequência para transformada wavelet. Fonte: [69] (Adaptação).

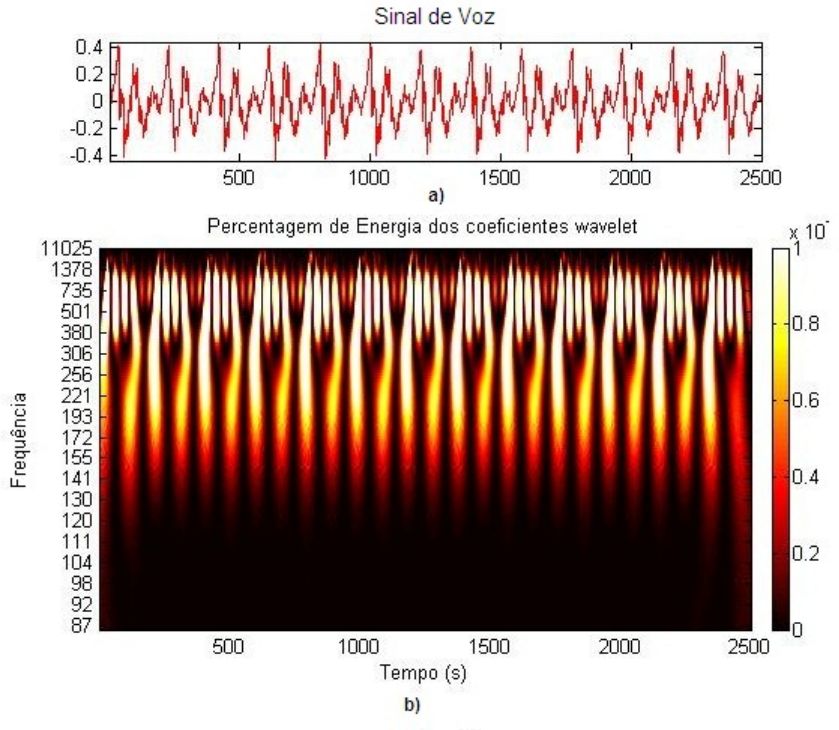


Figura 3.4 – Sinal de Voz (a) e Escalograma (b) de um sinal de voz saudável.

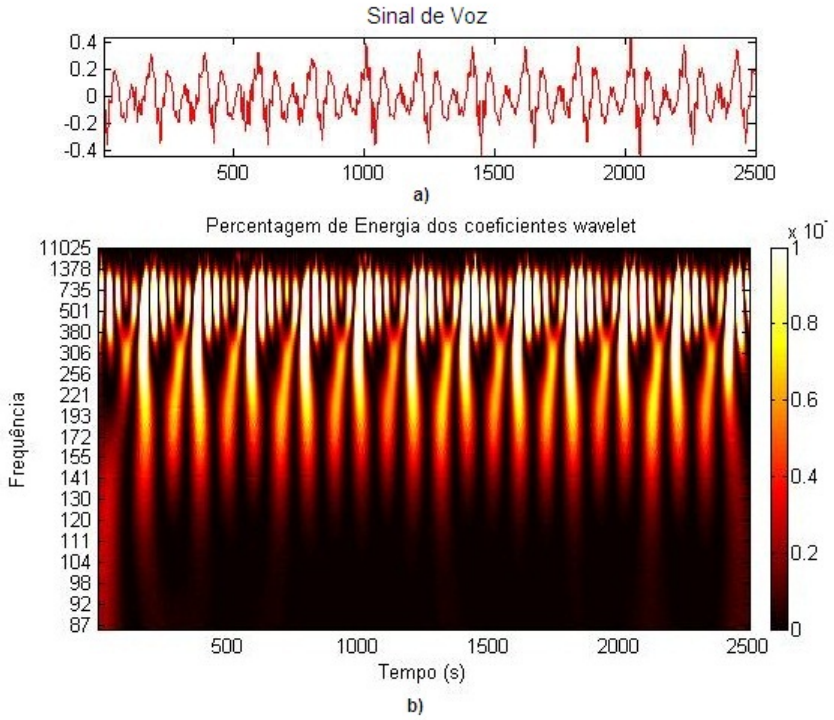


Figura 3.5 – Sinal de Voz (a) e Escalograma (b) de um sinal de voz com desvio vocal rugosidade.

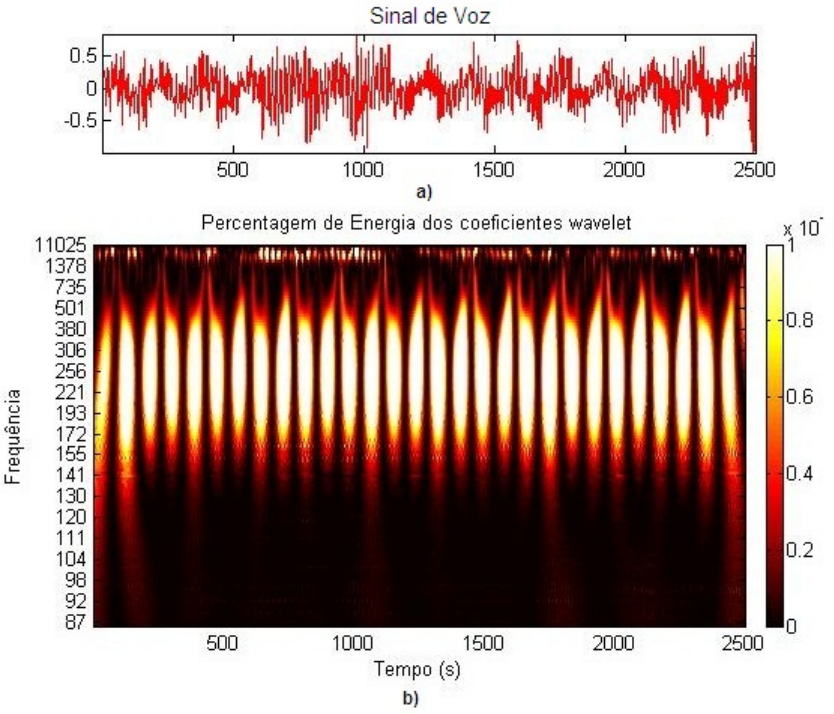


Figura 3.6 – Sinal de Voz (a) e Escalograma (b) de um sinal de voz com desvio vocal soprosidade.

3.3 – Transformada Wavelet Discreta (TWD)

A TWD fornece uma representação não redundante do sinal e seus valores constituem os coeficientes de decomposição *wavelet wavelet*. Os coeficientes *wavelet* fornecem informações

completas de uma forma simples e uma estimativa direta de energias locais em diferentes escalas. Além disso, as informações podem ser organizadas em um esquema hierárquico de subespaços aninhados chamada de análise de multiresolução em $L_2\mathbb{R}$ [70].

A versão discreta da transformada pode ser obtida discretizando as dilatações e as translações. Neste caso, as funções wavelets para a transformada wavelet discreta podem ser representadas pela função wavelet mãe $\psi(t)$ com um conjunto discreto de parâmetros, $a = 2^j$ e $b = k.2^j$, em que j e k são inteiros. O conjunto discreto de wavelets é representado por:

$$\psi_{j,k}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - k). \tag{3.3}$$

Essa família de funções constitui uma base ortonormal do Espaço de Hilbert $L_2\mathbb{R}$ consistindo de sinais de energia finita. Para se construir a wavelet mãe $\psi(t)$, é preciso determinar a função escalonamento $\phi(t)$, que satisfaz a seguinte equação:

$$\phi_{j,k}(t) = \sqrt{2^{-j}}\phi(2^{-j}t - k). \tag{3.4}$$

Uma função contínua $f(t)$ pode ser decomposta na *j*-ésima escala ou resolução, em termos das funções base wavelet e escalonamento por:

$$f(t) = \sum_k (c_j(k)\phi_{j,k}(t) + d_j(k)\psi_{j,k}(t)), \tag{3.5}$$

em que $c_j(k)$ e $d_j(k)$ correspondem aos coeficientes de aproximação e detalhe respectivamente, definidos como:

$$c_j(k) = \sum_m h(m - 2k)c_{j-1}(m) \tag{3.6}$$

$$d_j(k) = \sum_m g(m - 2k)c_{j-1}(m) \tag{3.7}$$

A TWD também pode ser vista como um processo de filtragem do sinal, usando um filtro passa-baixas $h(n)$ e um filtro passa-altas $g(n)$. Então, o primeiro nível de decomposição TWD de um sinal divide em duas faixas, uma versão passa-baixas e uma versão passa-altas do sinal.

As Equações 3.6 e 3.7 representam operações de filtragem por meio das respostas ao impulso de filtros de análise passa-baixas $h(n)$ e passa-altas $g(n)$. Para cada nível de resolução j, o algoritmo da transformada wavelet discreta, proposto por Mallat [60], decompõe o sinal em dois conjuntos de coeficientes: versão passa-baixas que fornece a representação aproximada do sinal (aproximação $c_j(k)$), enquanto a passa-altas indica os detalhes ou variações de altas frequências (detalhe $d_j(k)$). As informações extraídas em uma dada resolução são mantidas nos níveis de resolução superiores. Então, a decomposição wavelet resulta em uma árvore cuja estrutura é dita recursiva [71]. O fator $2k$, no índice dos filtros, representa a decimação por um fator 2 como pode ser visto na Figura 3.7.

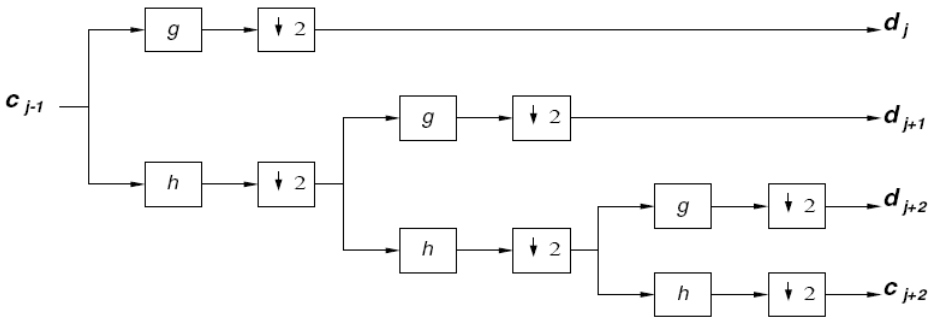


Figura 3.7 – Decomposição de sinal em três níveis, utilizando TWD. [15].

3.4 – Características Wavelets

Algumas características podem ser extraídas a partir dos coeficientes obtidos pela decomposição wavelet de um determinado sinal. Nesta pesquisa, são utilizadas a energia normalizada e a entropia dos coeficientes de detalhes da transformada wavelet, em nove níveis de resolução, utlizando a família wavelet de Daubechies de ordens 5 e 40 para os casos de classificação empregados.

A energia do sinal associada às faixas de frequência dos diferentes níveis de resolução pode apontar um desvio vocal. Medidas de entropia vem sendo empregadas para avaliar desordens vocais provocadas por patologias laríngeas, por medirem o grau de desordem de um sinal [72] [23].

3.4.1 – Energia Wavelet

Utilizando a energia normalizada dos coeficientes de detalhe como característica, pode-se identificar o quanto a energia do sinal de voz encontra-se distribuída ao longo da frequência [73].

Em geral, para sons sonoros, sinais de vozes saudáveis apresentam uma periodicidade no tempo, enquanto sinais com desvios vocais apresentam um comportamento irregular tanto das características temporais como espectrais. Comumente, a qualidade da voz é alterada na presença de desvios vocais por meio de parâmetros como aspereza, rouquidão e soprosidade.

A aspereza ocorre devido a rigidez da mucosa, que causa uma irregularidade vibratória com ruídos nas altas frequências. A rouquidão é proveniente da irregularidade de vibração das pregas vocais, que geram ruídos nas baixas frequências. A soprosidade indica a presença de ruído de fundo, audível, que corresponde fisiologicamente à fenda glótica [34].

O conceito do uso da energia como características em diferentes bandas obtida usando Transformada de Fourier de Tempo Curto (STFT) pode ser extendido para a Transformada Wavelet Discreta (TWD). Então, dado um processo estocástico $x(t)$, seu sinal associado é

assumido ser dado pelos valores amostrados *X=x(n),n=1,...,M*. Os coeficientes wavelet obtidos da decomposição wavelet são dados por:

$$d_j(k) = (2^{\frac{j}{2}}\phi(2^j t - k)) \tag{3.8}$$

com $j = 1, 2, ..., N$ e $N = \log_2 M$. O número de coeficientes de cada nível de resolução é $N_j = 2^j M$. Nota-se que esta correlação dá informações sobre o sinal na escala 2^j e no tempo $j2^j k$. O conjunto de coeficientes wavelet para o nível j, $d_j(k)$, é também um processo estocástico, onde *k* representa a variável de tempo discreto. Ele fornece uma estimativa direta das energias locais em diferentes escalas [74].

Assim, para os coeficientes wavelet dados por $d_j(k)$, a energia em cada nível de decomposição *j = 1, 2, ..., N* será a energia dos detalhes do sinal dada por

$$E_j = \sum_k |d_j(k)|^2 \tag{3.9}$$

E a energia em cada amostra de tempo *k* é

$$E(k) = \sum_{j-1}^{N} |d_j(k)|^2 \tag{3.10}$$

Consequentemente, a energia total do sinal pode ser obtida através da Equação 3.11:

$$E_{TOTAL} = \sum_{j-1}^{N}\sum_k |d_j(k)|^2 = \sum_{j-1}^{N} E_j \tag{3.11}$$

A energia normalizada EN_j dos coeficientes de detalhe em cada resolução j, é obtida através da Equação 3.12:

$$EN_j = \frac{\sum_k |d_j(k)|^2}{\sum_k |c_j(k)|^2 + |d_j(k)|^2} \tag{3.12}$$

3.4.2 – Entropia Wavelet

Outra característica a ser extraída dos coeficientes da decomposição wavelet é a entropia. A entropia de Shannon [75] é um critério útil para analisar e comparar a distribuição de probabilidade, já que fornece uma medida da informação para qualquer distribuição de probabilidade.

A entropia wavelet aparece como uma medida do grau de ordem ou desordem do sinal, fornecendo informações úteis sobre o processo dinâmico subjacente associado ao sinal.

Uma vez que a entropia avalia a quantidade de informação produzida por um processo, a mesma é influenciada pelas irregularidades e aleatoriedade dos sistemas fisiológicos, a exemplo do sistema de produção vocal [13] [73], podendo ser usada como medida na avaliação de desordens vocais.

A entropia de Shannon (*H*) dos coeficientes de detalhe em cada resolução j, é obtida através da Equação 3.13 [66].

$$H_j = -\sum p_j(k) \log p_j(k), \tag{3.13}$$

em que $p_j(k) \frac{|d_j(k)|^2}{\sum_k |d_j(k)|^2}$

3.5 – Revisão Bibliográfica

Apesar de ser uma técnica relativamente recente, a transformada wavelet, tem apresentado resultados significativos na discriminação entre vozes normais e patológicas, [15], [17], [18], [19], [20], [21], [22].

Diversos métodos tem sido propostos na literatura com a tarefa de classificar desordens vocais empregando análise acústica. No entanto, observa-se que determinado método ou característica pode apresentar um bom desempenho para classificar um determinado tipo de desordem ou patologia, mas não ser útil para outro tipo.

Desta forma, a busca por características e métodos mais precisos e eficientes para uma análise acústica com níveis de precisão mais confiáveis ainda é fruto de diversas pesquisas. Nesta seção, será apresentada uma revisão bibliográfica dos trabalhos que também utilizam a transformada wavelet no processamento digital de sinais para análise de desordens vocais com fins de diagnóstico.

Correia *et al.* [15], empregam a energia normalizada dos coeficientes de detalhes obtidos através da transformada wavelet discreta para distinguir sinais de vozes saudáveis dos afetados por edema de Reinke e nódulos nas pregas vocais. A wavelet de Daubechies de ordem 35 é usada para decompor os sinais em oito níveis de resolução. As características extraídas são avaliadas individualmente e de forma combinada, com o intuito de determinar as faixas de frequência que fornecem a melhor discriminação entre as vozes saudáveis e patológicas. Para a classificação é empregada a análise discriminante quadrática. Os resultados atestam que o quarto nível de resolução fornece as melhores taxas de reconhecimento. Uma acurácia de 97% foi obtida na classificação dos sinais de vozes em saudáveis e afetados por nódulos vocais.

Carvalho [17], em seu trabalho de dissertação, traz um extrator de características para diferenciação entre vozes saudáveis e patológicas utilizando a transformada wavelet discreta. O conjunto de dados utilizando em seu trabalho consiste de 60 amostras de sinais de vozes divididas em quatro classes de amostras, uma de indivíduos saudáveis e outras de três de indivíduos acometidos de nódulo vocal, edema de Reinke e disfonia neurológica. A vogal utilizada para gravação das vozes foi a vogal /a/ sustentada e os resultados obtidos mostram que a abordagem proposta, baseada na modificação da decomposição da Transformada Wavelet que é variante à mudança de variância, é uma técnica adequada para discriminação saudável/patológica, com resultados similares ou superiores a técnica clássica de decomposição.

Rodrigues [18], em sua tese, cria uma nova família de filtros digitais específica para o processo de classificação de dados, particularmente aplicada ao pré-diagnóstico de patologias na laringe, baseada na família wavelet de Daubechies. A base de dados utilizada em seu trabalho pertence ao banco de vozes previamente laudado pelo Departamento de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (FMRP-USP). São apresentados os resultados obtidos com base na técnica proposta, verificando-se uma taxa de acerto na classificação de vozes normais de 100% e uma taxa de acerto de 95,52% para vozes patológicas.

Almeida [19], em sua dissertação, propõe o desenvolvimento de um sistema de classificação de vozes para auxiliar no pré-diagnóstico de patologias na laringe, bem como no acompanhamento de tratamentos farmacológicos e pós-cirúrgicos. Os extratores de características foram obtidos através dos coeficientes de Predição Linear (LPC), Coeficientes Cepstrais de Frequência Mel (MFCC) e os coeficientes obtidos através da Transformada Wavelet Packet (WPT). Com o objetivo de maximizar a margem de separação entre as classes envolvidas, foi utilizada na classificação Máquina de Vetor de Suporte (SVM). O hiperplano gerado foi determinado pelos vetores de suporte, que são subconjuntos de pontos dessas classes. De acordo com o banco de dados utilizado no trabalho, os resultados apresentaram um bom desempenho, com taxa de acerto de 98,46% para classificação de vozes normais e patológicas em geral, e 98,75% na classificação de patologias entre si: edemas e nódulos.

Souza [20], em sua dissertação, propõe um modelo não invasivo para o pré-diagnóstico de patologias vocais, baseado em um algoritmo que combina duas máquinas de Vetores de Suporte, treinadas com o uso de um procedimento de aprendizado semi-supervisionado, alimentadas por um conjunto de parâmetros obtidos com o uso da Transformada Wavelet Discreta do sinal de voz do locutor. A base de dados utilizada possui 50 vozes com características normais e outras 50 pertencentes a indivíduos com algumas patologias na laringe, tais como nódulo nas pregas vocais, edema de Reinke, entre outras, em diversos níveis. Todos os indivíduos foram previamente examinados por profissionais da área médica, para confirmar seu estado saudável ou patológico. Os testes realizados com uma base de dados de vozes normais e afetadas por diversas patologias demonstram a eficácia da técnica proposta, que pode, inclusive, ser implementada em tempo-real.

Fonseca [21], em sua tese, utiliza as vantagens da Transformada Wavelet Discreta (TWD), além dos coeficientes de predição linear (LPC) e do algoritmo de inteligência artificial, *Least Squares Support Vector Machines* (LS-SVM), para aplicações em análise de sinais de voz e classificação de vozes patológicas. Os parâmetros de medida para a análise e classificação das vozes patológicas com edema de Reinke e nódulo foram extraídos das componentes da TWD. O banco de dados com as vozes patológicas foi obtido do Departamento de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço do Hospital das Clinicas da Faculdade de Medicina de Ribeirão Preto (FMRP-USP). Utilizando o algoritmo de reconhecimento de padrões, LS-SVM, mostrou-se que a combinação dos componentes da TWD de Daubechies com o filtro LPC inverso levou a um

classificador de bom desempenho alcançando mais de 90% de acerto na classificação das vozes patológicas.

Crovato [22], em sua dissertação, apresenta um sistema de classificação de voz disfônica utilizando a transformada wavelet packet (WPT) e o algoritmo *best basis* (BBA) como redutor de dimensionalidade e seis Redes Neurais Artificiais (ANN) atuando como um conjunto de sistemas denominados especialistas. O banco de vozes utilizado está separado em seis grupos de acordo com as similaridades patológicas (onde o 6º grupo é o dos pacientes com voz normal). O conjunto de seis ANN foi treinado, com cada rede especializando-se em um determinado grupo. A base de decomposição utilizada na WPT foi a Symlet 5 e a função custo utilizada na Best Basis Tree (BBT) gerada com o BBA, foi a entropia de Shannon. Cada ANN é alimentada pelos valores de entropia dos nós da BBT. O sistema apresentou uma taxa de sucesso de 87,5%, 95,31%, 87,5%, 100%, 96,87% e 89,06% para os grupos 1 ao 6 respectivamente, utilizando o método de Validação Cruzada Múltipla (MCV). O poder de generalização foi medido utilizando o método de MCV com a variação *Leave-One-Out* (LOO), obtendo erros em média de 38,52%, apontando a necessidade de aumentar o banco de vozes disponível.

3.6 – Considerações Finais do Capítulo

Neste Capítulo foi apresentado uma abordagem geral da transformada wavelet (TW), sua importância para o processamento digital de sinais e as características extraídas a partir da decomposição wavelet, utilizadas no desenvolvimento deste trabalho.

Dessa forma, a Transformada Wavelet Discreta (TWD) pode ser utilizada para extrair características dos sinais de vozes, permitindo classificar as amostras de voz em saudáveis ou desviadas e ainda classificá-las quanto ao grau de intensidade do desvio vocal, bem como pode ser aplicada na separação entre a qualidade vocal predominante, como será apresentado no capítulo 5.

Foram apresentados também os trabalhos mais recentes que utilizam a transformada Wavelet no processamento digital de sinais de voz, mostrando que essa transformada apresenta resultados significativos para nesta aplicação.

No capítulo seguinte, será apresentada a metodologia empregada nesta pesquisa, bem como os materiais utilizados no desenvolvimento da mesma.

Capítulo 4

Material e Métodos

Neste trabalho, para avaliação da qualidade vocal em crianças, foram considerados dois estudos de caso: 1) Análise acústica da intensidade do desvio vocal; e 2) Análise acústica da qualidade vocal predominante (rugosidade e soprosidade).

Este capítulo apresenta a base de dados utilizada nesta pesquisa, bem como a metodologia empregada.

4.1 – Base de dados

A base de dados, utilizada neste trabalho, foi fornecida pelo Laboratório de Voz e Deglutição, atual Laboratório Integrado de Estudos da Voz (LIEV), do Departamento de Fonoaudiologia da Universidade Federal da Paraíba [30], tendo sido aprovada a sua aquisição pelo comitê de ética em pesquisa daquela Instituição sob o protocolo número 775/10 e cedida para uso nesta pesquisa. A mesma contém 93 sinais de vozes da vogal sustentada /ε/ de ambos os sexos, com idade variando entre 3 e 10 anos, sendo 48 meninas e 45 meninos, todos integrantes de uma escola vinculada a uma instituição de ensino pública federal. A coleta foi realizada com um notebook HP (*Hewlett-Packard Development Company*, Palo Alto, CA) e microfone *headset* da marca Logitech (Logitech, Fremont, CA), utilizando o software *PRAAT* versão 5.1.44 com taxa de amostragem de 44,100Hz.

As vozes foram editadas no *software SoundForge* versão 10.0 (Sony, Tokyo, Japan), no qual foram eliminados segundos iniciais e finais da emissão da vogal, devido a maior irregularidade nesses trechos, preservando-se o tempo mínimo de dois segundos para cada emissão. Foi realizada a normalização dos sinais, no controle *normalize* do *SoundForge*, no modo *peaklevel*, a fim de obter uma padronização na saída de áudio entre - 6 e 6 dB.

Esses sinais, inicialmente, foram classificadas com análise perceptivo-auditiva com a escala analógico visual (EAV) de acordo com o grau geral de intensidade do desvio vocal (grau 1 para voz saudável, grau 2 para voz com desvio leve e grau 3 para voz com desvio moderado).Do total de 93 sinais, 10 foram considerados normais, 70 apresentam desvio leve e 13 apresentam desvio moderado.

Não há, nessa base de dados, casos disponíveis de sinais classificados como grau geral 4 (desvio intenso). Quanto à presença do desvio vocal, 10 sinais foram considerados sem desvio,

19 sinais apresentaram rugosidade e 21 sinais apresentaram soprosidade. Os 43 sinais restantes, apresentaram desvios vocais não utilizados neste trabalho.

Para o estudo de caso 1, quatro classes de sinais foram consideradas: grau geral 1 (GG1), grau geral 2 (GG2), grau geral (GG3), grau geral 2 e grau geral 3 juntos (GG2 e GG3). A análise discriminante quadrática foi empregada a fim de investigar quatro casos de discriminação: $GG1 \times (GG2 e GG3)$, $GG1 \times GG2$, $GG1 \times GG3$, e $GG2 \times GG3$.

Para o estudo de caso 2, três classes de sinais foram consideradas: sinais de vozes saudáveis ou sem desvio (SDL), sinais de vozes com rugosidade (RUG) e sinais de vozes com soprosidade (SOP). A análise discriminante quadrática foi empregada a fim de investigar três casos de discriminação: SDL x RUG, SDL x SOP e RUG x SOP.

4.2 – Metodologia

A metodologia empregada segue o modelo de diagrama em blocos representado na Figura 4.1.

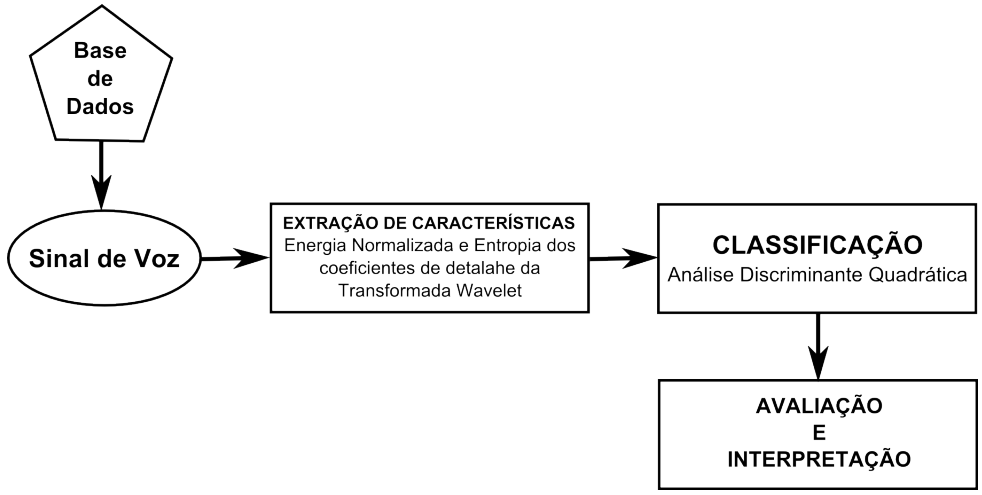


Figura 4.1 – Diagrama em blocos da metodologia empregada.

Os sinais de voz selecionados da base de dados são analisados por meio da energia normalizada e da entropia dos coeficientes de detalhe da transformada wavelet discreta, obtidas em nove níveis de resolução.

Como os sinais de voz foram amostrados a 44100 amostras/s, considera-se para análise frequências até 22050 Hz. Uma vez que nove níveis de resolução são considerados, as medidas de energia e entropia extraídas fornecem informações dos sinais em diferentes faixas de frequência. A Tabela 4.1 apresenta as faixas de frequência dos coeficientes de detalhe para cada um dos níveis de resolução considerados.

Sinais de vozes desviadas podem apresentar alterações em sua frequência fundamental. Como a frequência fundamental pode atingir valores entre 80 e 250 Hz, como apresentado na

Tabela 4.1 – Níveis de resolução e suas respectivas faixas de frequência para os coeficientes de detalhes da transformada wavelet.

Nível de Resolução	Faixa de Frequência (Hz)
1	11025 a 22050
2	5512,5 a 11025
3	2756,25 a 5512,5
4	1378,12 a 2756,25
5	689,06 a 1378,12
6	344,53 a 689,06
7	172,26 a 344,53
8	86,13 a 172,26
9	43,06 a 86,13

Figura 2.7, a escolha de nove níveis de resolução da transformada wavelet, visa contemplar essa faixa de frequência, de modo que o sistema possa ser empregado tanto na análise de vozes infantis, quanto de adultos.

As características extraídas, em cada nível de resolução, são utilizadas individualmente e combinadas entre si para a classificação do grau da intensidade do desvio vocal ou da qualidade vocal predominante.

Para tanto, empregou-se uma função de análise discriminante quadrática com validação cruzada, que é um procedimento em que são variados os conjuntos de treino e teste do classificador com os sinais da base de dados, com 10 subconjuntos de forma estratificada, considerando as medidas individualmente de energia e entropia e de forma combinada 2 a 2, 3 a 3, 4 a 4,..., 16 a 16 e as 17 conjuntamente.

A fim de diminuir a disparidade entre as quantidades de sinais utlizados nesta pesquisa, foram utlizados grupos equivalentes de sinais para cada caso de discriminação, e para os casos em que houve a necessidade de se dividir a classificação em vários grupos, foi tirada a média de todas as medidas obtidas.

Para cada subconjunto, 90% dos sinais foram usados para teste e 10% para treino. A escolha dos sinais se deu de forma aleatória e sem repetição.

A classificação foi realizada com Análise Discriminante Linear (LDA) e Quadrática (QDA), implementadas em ambiente Matlab v.7.12.0. Entretanto, as melhores medidas foram obtidas através da análise discriminante quadrática. Assim sendo, os resultados apresentados nesta pesquisa utilizam este modelo de classificação.

4.3 – Descrição do Classificador

A análise discriminante é uma técnica da estatística multivariada utilizada para classificar objetos em dois ou mais grupos. A ideia básica é a obtenção de uma combinação linear das características observadas que apresente maior poder de discriminação entre populações. Esta combinação linear é denominada função discriminante [7] [6] [76].

A função discriminante tem a propriedade de minimizar as probabilidades de baixas taxas de classificação, quando as populações são normalmente distribuídas com média μ e variância σ^2 conhecidas.

Entretanto, em geral, a média e a variância das populações não são conhecidas, resultando na necessidade de estimação desses parâmetros. Dessa forma, pode-se assumir que as populações têm uma mesma matriz de covariâncias ou não. Quando a regra de classificação assume que as variâncias das populações são iguais, as funções discriminantes são ditas lineares (LDA) e, quando não, são funções discriminantes quadráticas (QDA) [7] [6] [76].

Na Figura 4.2, é apresentado um exemplo da aplicação da análise discriminante linear em um espaço de duas características. A função discriminante linear realiza a transformação dos dados de ambas as classes para o sub-espaço LDA, no qual é traçado um hiperplano de separação entre as mesmas [7] [6] [26].

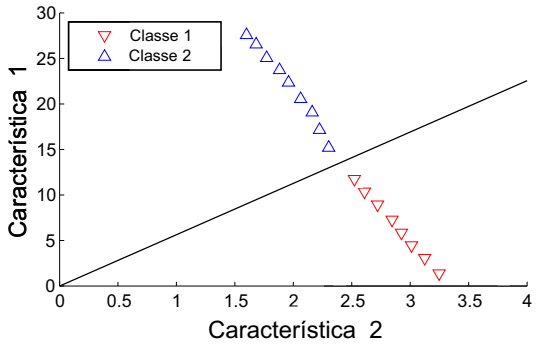


Figura 4.2 – Função discriminante linear em um espaço de características arbitrário. Fonte: [12].

Na Figura 4.3, é apresentado um exemplo da aplicação da análise discriminante quadrática em um espaço de duas características. A função discriminante quadrática busca uma curva não linear que proporcione a maior separabilidade entre as classes.

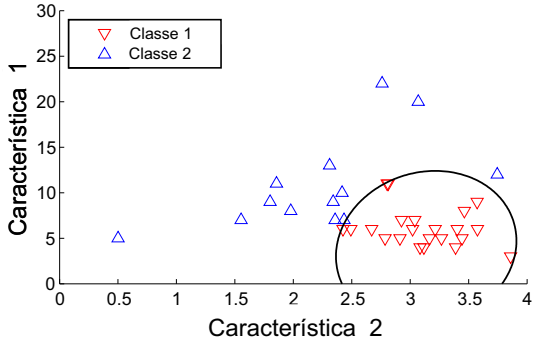


Figura 4.3 – Função discriminante quadrática em um espaço de característica arbitrário. Fonte: [12].

4.4 – Avaliação e Interpretação

Para mensurar a precisão dos classificadores empregados em cada estudo de caso, três medidas comumente empregadas são utilizadas: acurácia, sensibilidade e especificidade. Essas medidas estão relacionadas à capacidade de um classificador em diagnosticar uma doença em um paciente doente (Verdadeiro Positivo - VP) ou saudável (Falso Positivo - FP), ou, ainda, diagnosticar um estado saudável em um paciente saudável (Verdadeiro Negativo - VN) ou doente (Falso Negativo - FN) [24]. Cada um desses parâmetros aparecem na chamada matriz de confusão (Tabela 4.2), que representa o resultado obtido no classificador.

Tabela 4.2 – Matriz de confusão em um teste de detecção da presença/ausência de doença.

	Doença	
Resultado	**Presente**	**Ausente**
Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: [12]

A medida de acurácia (Ac) mede a taxa de classificação correta global, refletindo a capacidade do classificador de identificar corretamente quando há e quando não há a presença do distúrbio. A acurácia é definida como a relação entre o número de casos corretamente classificados e todos os casos apresentados ao classificador [24]:

$$Ac = \frac{VP + VN}{VP + VN + FP + FN}. \tag{4.1}$$

A medida de sensibilidade (Sen) mede a capacidade do classificador em identificar a presença do distúrbio quando ele de fato existe, sendo definida pela relação entre o número de casos corretamente classificados como presença do distúrbio e a quantidade total de casos com o distúrbio:

$$Sen = \frac{VP}{VP + FN}. \tag{4.2}$$

A medida de especificidade (Esp) mede a capacidade do classificador em identificar corretamente a ausência do distúrbio quando de fato ele não existe, sendo definida pela relação entre o número de casos corretamente classificados como saudáveis e a quantidade total de casos de estado saudável:

$$Esp = \frac{VN}{VN + FP}. \tag{4.3}$$

O classificador apresenta bom desempenho caso seja capaz de obter altos valores para acurácia, sensibilidade e especificidade.

Dessa forma, a discriminação entre classes atinge maior precisão. A representação das medidas de sensibilidade e especificidade é mais clara quando se trata da discriminação

entre uma classe saudável e uma classe patológica. Quando há a discriminação entre classes disfônicas, é necessário que seja definido, no classificador, qual grupo de sinais terá sua correta classificação medida pela sensibilidade e qual grupo terá sua correta classificação medida pela especificidade [12].

Na Tabela 4.3, estão apresentados todos os casos de discriminação considerados nesta pesquisa, relacionando as medidas de sensibilidade e especificidade com cada classe envolvida.

Tabela 4.3 – Níveis de resolução e suas respectivas faixas de frequência para os coeficientes de detalhes da transformada wavelet.

Casos de Discriminação	**Sensibilidade**	**Especificidade**
$GG1 \times (GG2 e GG3)$	Taxa de correta classificação dos casos de vozes alteradas (GG2 e GG3)	Taxa de correta classificação dos casos saudáveis (GG1)
$GG1 \times GG2$	Taxa de correta classificação dos casos de vozes com desvio leve (GG2)	Taxa de correta classificação dos casos saudáveis (GG1)
$GG1 \times GG3$	Taxa de correta classificação dos casos de vozes com desvio moderado (GG3)	Taxa de correta classificação dos casos saudáveis (GG1)
$GG2 \times GG3$	Taxa de correta classificação dos casos de vozes com desvio moderado (GG3)	Taxa de correta classificação casos de vozes com desvio leve (GG2)
$SDLxRUG$	Taxa de correta classificação dos casos de rugosidade (RUG)	Taxa de correta classificação dos casos saudáveis (SDL)
$SDLxSOP$	Taxa de correta classificação dos casos de soprosidade (SOP)	Taxa de correta classificação dos casos saudáveis (SDL)
$RUGxSOP$	Taxa de correta classificação dos casos de soprosidade (SOP)	Taxa de correta classificação dos casos de rugosidade (RUG)

4.5 – Considerações Finais do Capítulo

Neste capítulo foi apresentada a metodologia empregada no desenvolvimento desta pesquisa, quais a características utilizadas, o modelo de classificador empregado, bem como quais estudos de caso foram considerados.

Para cada caso de classificação, as características dos coeficientes de energia e entropia foram consideradas de maneira individual e combinada, a fim de se obter a melhor taxa de acurácia, como será tratado no capítulo que se segue, onde serão apresentados todos os resultados obtidos nesta pesquisa.

Capítulo 5

Resultados

Neste capítulo são apresentados os resultados obtidos nesta pesquisa, de acordo com a metodologia adotada, apresentada no capítulo anterior, que tem como objetivo investigar o potencial discriminativo da energia normalizada e da entropia dos coeficientes de detalhe da transformada wavelet discreta, no Estudo de Caso 1 (Análise acústica do grau de intensidade do desvio vocal e no Estudo de Caso 2 (Análise acústica da qualidade vocal predominante (rugosidade e soprosidade).

Antes de apresentar os resultados para cada estudo de caso considerado, são apresentados os modelos de escolha adotados, para se chegar à decisão de que ordem da wavelet de Daubechies a ser utilizada.

5.1 – Teste das Ordens da Wavelet de Daubechies

A fim de investigar qual a ordem da família wavelet a ser utilizada em cada estudo de caso, foram feitos os testes de classificação como são apresentados nas subseções seguintes.

5.1.1 – Teste para o Estudo de Caso 1

Foram testadas as 45 ordens da wavelet de Daubechies, a fim de investigar qual ordem dessa wavelet apresentava melhor desempenho, para o estudo de caso 1, apresentado no capítulo anterior.

Elas foram testadas quanto a capacidade de classificação voz Saúdavel *x* Voz Desviada (GG1 *x* GG2 e 3), utilizando-se das nove medidas de energia e os resultados obtidos são apresentados na Figura 5.1.

Para selecionar a wavelet mãe adequada, bem como a ordem do filtro, que representasse bem as desordens vocais presentes nos sinais de voz analisados, utilizou-se o critério adotado por Salhi [66], que considera a Daubechies de ordem 40 adequada por proporcionar a análise de uma porção significativa do espectro wavelet. Esta escolha também considera o fato desta wavelet demonstrar um alto desempenho em testes de escuta informal.

Pela Figura 5.1, observa-se que a partir da ordem 12, a família Daubechies já apresenta resultados próximos aos da Db40. Espera-se que para ordens acima de 12, já se obtenha um bom desempenho. Optou-se, no entanto, por empregar a Db40 pelos motivos acima citados.

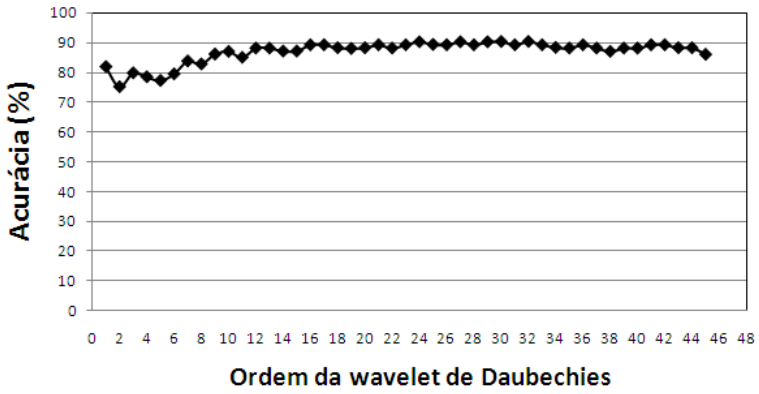


Figura 5.1 – Classificação GG1 x GG2 e 3 para as 45 Wavelets de Daubechies.

5.1.2 – Teste para o Estudo de Caso 2

As 45 wavelets da família de Daubechies, foram testadas a fim de investigar qual ordem dessa Wavelet apresentava melhor desempenho, para o estudo de caso 2, apresentado no capítulo anterior.

Elas foram testadas quanto a capacidade de classificação voz com rugosidade *x* Voz com soprosidade (RUG *x* SOP), utilizando-se das nove medidas de energia e os resultados obtidos são apresentados na Figura 5.2.

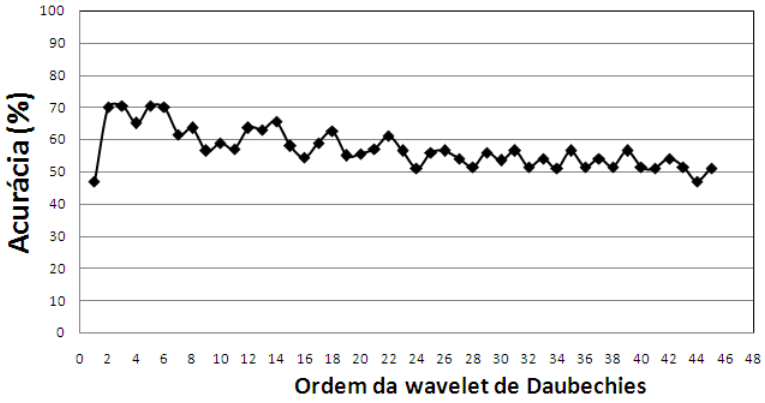


Figura 5.2 – Classificação RUG x SOP para as 45 Wavelets de Daubechies.

Como na literatura não foi encontrado esse caso de classificação, a wavelet foi escolhida pelo seu desempenho, observando-se a maior acurácia com menor desvio e, desta forma, chegamos a wavelet de Daubechies de ordem 5, para ser utilizada para este estudo de caso. O comprimento do filtro tem influência importante na determinação do melhor desempenho. Filtros de menor ordem são indicados para extração de detalhes, e é mais eficaz em detectar alterações sutis, como é o caso desta classificação, onde os sinais possuem características parecidas.

5.2 – Classificação no Estudo de Caso 1: Análise Acústica do Grau de Intensidade do Desvio Vocal

Os resultados referentes à classificação realizada no Estudo de Caso 1 são apresentados nesta seção. A classificação foi realizada de acordo com a metodologia apresentada no capítulo anterior.

O processo de classificação dos sinais foi realizado considerando-se quatro casos distintos: GG1 *x* (GG2 e GG3), GG1 *x* GG2, GG1 *x* GG3 e GG2 *x* GG3. O primeiro caso de classificação pode ser utilizado para triagem vocal, onde o sinal de voz é considerado ou não, desviado. Os demais casos de classificação, podem ser utilizados para monitoramento da terapia fonoaudiológica.

A análise de desempenho foi realizada usando como medida a acurácia, que é definida como a relação entre o número de casos corretamente classificados e todos os casos apresentados ao classificador.

Inicialmente, a energia e a entropia dos coeficientes de detalhe das wavelets foram avaliadas individualmente para cada nível de resolução. Em seguida, as características foram agrupadas em um único vetor, contendo 18 medidas (9 de energia e 9 de entropia). Por fim, as características foram combinadas 2 a 2, 3 a 3, 4 a 4,..., 16 a 16 e as 17, e selecionada a combinação que apresentou a melhor taxa de classificação.

Entre os classificadores LDA e QDA, o segundo foi escolhido por ter apresentado os melhores resultados na maioria dos casos considerados. A seguir, são apresentados os resultados obtidos para cada caso de discriminação considerado.

Classificação Grau Geral 1 x (Grau Geral 2 e Grau Geral 3)

Na Tabela 5.1, estão apresentados os melhores valores obtidos de acurácia (Ac), sensibilidade (Sen) e especificidade (Esp) em sinais de voz de crianças com grau geral 1 (grau geral normal - GG1) e crianças com voz alterada, compreendendo o conjunto dos casos com Grau Geral leve (Grau Geral 2 - GG2) e Grau Geral moderado (Grau Geral 3 - GG3).

Tabela 5.1 – Classificação GG1 x (GG2 e 3)

	Medidas	**Ac (%)**	**Sen (%)**	**Esp (%)**
Individual	E_4	$73,89 \pm 9,26$	$83,33 \pm 12,08$	$84,44 \pm 15,82$
	H_3	$65 \pm 10,36$	$53,33 \pm 15,51$	$76,67 \pm 13,48$
Combinada	Todas	$48,33 \pm 5,68$	$12,22 \pm 6,57$	$84,44 \pm 8,67$
	$E_1E_4E_5H_5H_7$	$98,89 \pm 1,12$	$98,89 \pm 1,12$	$98,89 \pm 1,12$

Para este caso de classificação, o nível de energia E1 e de entropia H3 foram os responsáveis pela separação entre as classes, de maneira individual. A combinação das medidas de energia e entropia resultou no aumento considerável da acurácia, atingindo 98,89% de acerto.

Classificação Grau Geral 1 *x* Grau Geral 2

Na Tabela 5.2, estão apresentados os melhores valores obtidos de acurácia (Ac), sensibilidade (Sen) e especificidade (Esp) em sinais de voz de crianças com grau geral 1 (grau geral normal - GG1) e crianças com Grau Geral 2 (grau geral Leve - GG2).

Tabela 5.2 – Classificação GG1 *x* GG2.

	Medidas	**Ac (%)**	**Sen (%)**	**Esp (%)**
Individual	E_4	$72,86 \pm 10,09$	$78,57 \pm 12,86$	$67,14 \pm 14,90$
	H_1	$64,29 \pm 8,41$	$47,14 \pm 14,80$	$80 \pm 12,61$
Combinada	Todas	$49,29 \pm 2,15$	$1,43 \pm 1,43$	$97,14 \pm 2,83$
	$E_1E_3E_4H_4$	$97,86 \pm 1,67$	$97,14 \pm 2,86$	$98,57 \pm 1,43$

Avaliando as características de forma individual, neste caso, destacam-se a energia no nível 4 (E4), onde se encontra a faixa de normalidade vocal e a entropia no nível 1 (H1), evidenciando a presença de distúrbios da voz nas frequências mais altas. A combinação das medidas de energia e entropia resultou no aumento considerável da acurácia, atingindo 97,86% de acerto, apesar de, neste caso, serem leves os desvios vocais.

Classificação Grau Geral 1 *x* Grau Geral 3

Na Tabela 5.3, estão apresentados os melhores valores obtidos de acurácia (Ac), sensibilidade (Sen) e especificidade (Esp) em sinais de voz de crianças com grau geral 1 (grau geral normal - GG1) e crianças com Grau Geral 3 (grau geral Moderado - GG3).

Tabela 5.3 – Classificação GG1 x GG3.

	Medidas	**Ac (%)**	**Sen (%)**	**Esp (%)**
Individual	E_1	$81,25 \pm 7,93$	$90,00 \pm 10,00$	$75,00 \pm 15,28$
	H_2	$78,33 \pm 7,48$	$90,00 \pm 10,00$	$70,00 \pm 13,34$
Combinada	Todas	$53,61 \pm 6,45$	$54,17 \pm 8,63$	$55,00 \pm 9,67$
	$E_1E_4H_5H_8$	$100,00 \pm 0$	$100,00 \pm 0$	$100,00 \pm 0$

Pode-se observar que, no caso de classificação GG1 x GG3, o nível de energia E1 e de entropia H2 foram os responsáveis pela separação entre as classes, de maneira individual, evidenciando a presença de distúrbios da voz nas frequências mais altas. A combinação das medidas de energia e entropia resultou no aumento considerável da acurácia, atingindo 100% de acerto.

Classificação Grau Geral 2 x Grau Geral 3

Na Tabela 5.4, estão apresentados os melhores valores obtidos de acurácia (Ac), sensibilidade (Sen) e especificidade (Esp) em sinais de voz de crianças com grau geral 2 (grau geral Leve - GG2) e crianças com Grau Geral 3 (grau geral Moderado - GG3).

Neste último caso de classificação, o nível de energia E6, onde se encontram as frequências mais graves do sinal de voz e o nível de entropia H1, onde se encontram as

Tabela 5.4 – Classificação GG2 x GG3.

	Medidas	**Ac (%)**	**Sen (%)**	**Esp (%)**
Individual	E_6	$69,99 \pm 9,82$	$66,66 \pm 12,59$	$67,5 \pm 12,73$
	H_1	$64,72 \pm 9,03$	$61,67 \pm 12,80$	$68,33 \pm 14,28$
Combinada	Todas	$5,005 \pm 2,54$	$20,00 \pm 13,34$	$90,00 \pm 10,00$
	$E_6H_1H_2H_4$	$96,11 \pm 3,41$	$96,67 \pm 3,34$	$95,00 \pm 5,00$

frequências mais agudas, foram os responsáveis pela separação entre as classes, de maneira individual. Isso evidencia a presença de distúrbios da voz tanto nas frequências mais altas, quanto a presença de desvios que diferenciam os graus em frequências mais baixas. A combinação das medidas de energia e entropia resultou no aumento considerável da acurácia, atingindo 96,11% de acerto, apesar de ambas as classes se tratarem de vozes desviadas.

5.2.1 – Discussão dos Resultados

É possível observar que a combinação das medidas resultou num aumento significativo da acurácia e reduziu o vetor de características de 18 para, 4 ou 5 medidas, no máximo.

A análise por faixa de frequência, por meio da energia normalizada e entropia dos coeficientes de detalhe da transformada wavelet, permitiu detectar a presença dos desvios vocais em sinais de vozes infantis.

As medidas combinadas apresentaram maior poder discriminatório, reduzindo a dimensionalidade dos dados e aumentando a acurácia.

Observa-se que a energia do nível de resolução 1 (E1) aparece em todas as combinações, exceto para a classificação entre os graus GG2 e GG3, ressaltando a presença de ruído nas altas frequências nas vozes desviadas. Como nesse nível estão presentes as componentes de mais alta frequência do sinal de voz, justifica-se o emprego da taxa de amostragem 44100 amostras/s. Faz-se necessário estudos que detalhem o comportamento dos desvios vocais nessa faixa de frequência.

Os níveis de energia 4, 5 e 6, em que situam-se o primeiro e segundo formantes da vogal /ε/ [4] são predominantes na discriminação, indicando a variabilidade destes parâmetros com o grau do desvio vocal.

5.3 – Classificação no Estudo de Caso 2: Análise Acústica da Qualidade Vocal Predominante

Os resultados referentes à classificação realizada no Estudo de Caso 2 são apresentados nesta Seção. A classificação foi realizada de acordo com a metodologia apresentada no capítulo anterior.

O processo de classificação dos sinais foi realizado considerando-se três casos distintos: Voz Normal *x* RUG, Voz Normal *x* SOP e RUG *x* SOP.

A análise de desempenho foi realizada usando como medida a acurácia que é definida como a relação entre o número de casos corretamente classificados e todos os casos apresentados ao classificador.

Inicialmente, a energia e a entropia dos coeficientes de detalhe das wavelets foram avaliadas individualmente para cada nível de resolução. Em seguida, as características foram agrupadas em um único vetor, contendo 18 medidas (9 de energia e 9 de entropia). Por fim, as características foram combinadas 2 a 2, 3 a 3, 4 a 4,..., 16 a 16 e as 17, e selecionada a combinação que apresentou a melhor taxa de classificação.

Entre os classificadores LDA e QDA, o segundo foi escolhido por ter apresentado os melhores resultados na maioria dos casos considerados. A seguir, os resultados obtidos para cada caso de discriminação considerado.

Classificação Voz Normal x RUG

Na Tabela 5.5, estão apresentados os melhores valores obtidos de acurácia (Ac), sensibilidade (Sen) e especificidade (Esp) em sinais de voz de crianças sem desvio vocal e crianças com desvio vocal do tipo rugosidade.

Tabela 5.5 – Classificação Voz Normal *x* RUG.

	Medidas	**Ac (%)**	**Sen (%)**	**Esp (%)**
Individual	E_1	$72,50 \pm 9,70$	$90,00 \pm 10,00$	$55,00 \pm 16,51$
	H_3	$60,00 \pm 9,83$	$80,00 \pm 12,65$	$40 \pm 14,84$
Combinada	Todas	$52,50 \pm 3,93$	$5,00 \pm 5,00$	$100,00 \pm 0$
	$E_1E_3E_4E_6H_5H_6$	$95 \pm 3,33$	100 ± 0	$90 \pm 6,67$

Nesta classificação, o nível 1 de Energia (E1) e o nível 3 de Entropia (H3), foram os níveis responsáveis pela separação, normal x rugoso. Apesar da qualidade vocal rugosidade estar mais associada aos níveis 4 e 5 [35], e esses mesmos níveis conterem as faixas de frequência para sinais de vozes normais, o algoritmo de classificação foi capaz de realizar a separação entre as classes, nos níveis de mais alta frequência, onde o sinal de voz disfônico mostrou características que não estavam presentes no sinal de voz saudável.

Na combinação, o uso de todas as medidas confundiu o classificador. Foram necessários apenas quatro níveis de energia e dois de entropia, proporcionando cerca de 35% a mais de 20%, comparado à maior acurácia individual, obtida com H3.

Classificação Voz Normal *x* SOP

Na Tabela 5.6, estão apresentados os melhores valores obtidos de acurácia (Ac), sensibilidade (Sen) e especificidade (Esp) em sinais de voz de crianças sem desvio e crianças com a qualidade vocal soprosidade.

Para este caso de classificação, o nível 3 de Energia (E3) e o nível 3 de Entropia (H3), foram os níveis responsáveis pela separação, normal *x* soproso. Observando-se a combinação

Tabela 5.6 – Classificação Voz Normal x SOP.

	Medidas	**Ac (%)**	**Sen (%)**	**Esp (%)**
Individual	E_3	$80,00 \pm 8,17$	$80,00 \pm 12,65$	$80,00 \pm 12,65$
	H_3	$70,,00 \pm 8,77$	$55,00 \pm 15,81$	$85,00 \pm 11,67$
Combinada	Todas	$50,00 \pm 3,93$	$5,00 \pm 5,00$	$95,00 \pm 5,00$
	$E_1E_2E_3H_7$	$100,00 \pm 0$	$100,00 \pm 0$	$100,00 \pm 0$

de características de energia e entropia observa-se a presença da qualidade vocal soprosidade entre os níveis de resolução 2 e 3 evidenciando a presença de um desvio vocal de mais alta frequência.

Na combinação, apenas um nível de entropia (H7), combinado a três níveis de energia de alta frequência proporcionaram os melhores resultados.

Classificação RUG *x* SOP

Na Tabela 5.7, estão apresentados os melhores valores obtidos de acurácia (Ac), sensibilidade (Sen) e especificidade (Esp) em sinais de voz de crianças com grau geral 1 (grau geral normal - GG1) e crianças com Grau Geral 3 (grau geral Moderado - GG3).

Tabela 5.7 – Classificação RUG x SOP

	Medidas	**Ac (%)**	**Sen (%)**	**Esp (%)**
Individual	E_5	$65,00 \pm 6,68$	$50,00 \pm 10,55$	$80,00 \pm 8,17$
	H_1	$60,00 \pm 6,67$	$45,00 \pm 11,67$	$75,00 \pm 8,34$
Combinada	Todas	$70,00 \pm 5,00$	$45,00 \pm 11,68$	$95,00 \pm 5,00$
	$E_2E_5E_8E_9H_1H_3H_4$	$90,00 \pm 5,53$	$100,00 \pm 0$	$80,00 \pm 13,34$

Apesar dos sinais rugosos possuírem algum grau de soprosidade e os sinais soprosos apresentarem também algum grau de rugosidade, a classificação proposta nesta pesquisa obteve 90% de separação entre as classes.

Como se pode observar na Tabela 5.7, no vetor resultante da combinação das medidas, as características de alta frequência das disfonias foram detectadas pelo níveis de energia E1 e pelos níveis de entropia E2 e H1, enquanto que os níveis de energia E8 e E9, foram os responsáveis pela separação entre as classes, no tocante a detecção das características das disfonias presentes nas baixas frequências.

Na discriminação entre rugosidade e soprosidade foi necessário uma quantidade maior de níveis, comparado aos demais casos de classificação. Este aspecto, em parte, pode ser justificado devido ao fato de que estas características não serem encontradas de forma única nos sinais de vozes. Nenhum sinal é apenas soproso ou apresenta só rugosidade. O predomínio da qualidade vocal é soproso ou rugoso, mas cada sinal apresenta um certo grau de cada um. Dessa forma, são necessários mais parâmetros para distinguir um do outro.

5.3.1 – Discussão dos Resultados

É possível observar que a combinação das medidas resultou num aumento significativo da acurácia e reduziu o vetor de características de 18 para, no máximo, 7 medidas, para todos os casos de classificação considerados neste estudo de caso.

A análise por faixa de frequência por meio da energia normalizada e entropia dos coeficientes de detalhe da transformada wavelet permitiu detectar a presença dos desvios vocais em sinais de vozes infantis.

Para a separação entre classes desviados fez-se necessário a utilização de faixa de frequências mais ampla, se comparado a classificação entre um classe normal e uma classe desviada.

As medidas combinadas apresentaram maior poder discriminatório, reduzindo a dimensionalidade dos dados e aumentando a acurácia.

Capítulo 6

Considerações Finais

Esta pesquisa investiga a aplicabilidade da classificação de disturbios da voz, utilizando-se das medidas de energia e entropia dos coeficientes de detalhe da transformada wavelet em nove níveis de resolução. As medidas foram empregadas, de maneira individual e combinadas, a fim de ser obter a maior acurácia possível.

Dois estudos de caso foram considerados para validar o uso dessas medidas e os resultados obtidos mostram o poder de discriminação das medidas de energia e entropia combinadas, mesmo quando os sinais possuem graus de desvio próximo, e até mesmo entre classes desviadas.

Para o estudo de caso 1, em todas as classificações, os valores de acurácia foram superiores a 95%: Classificação Grau Geral 1 *x* (Grau Geral 2 e Grau Geral 3) - 98,89%; Classificação Grau Geral 1 *x* Grau Geral 2 - 97,86%; Classificação Grau Geral 1 *x* Grau Geral 3 - 100%; Classificação Grau Geral 2 *x* Grau Geral 3 - 96,11%.

Para o estudo de caso 2, em todas as classificações, os valores de acurácia foram superiores a 90%: Classificação Voz Normal *x* Rugosidade - 95%; Classificação Voz Normal *x* Soprosidade - 100%; Classificação Rugosidade *x* Soprosidade - 90%.

Desta forma, conclui-se que o uso das medidas de energia e entropia dos coeficientes de detalhe da transformada wavelet mostra-se como uma técnica promissora que pode ser considerada para ser empregada como uma ferramenta para análise acústica de desvios vocais.

Como principais contribuições deste trabalho destacam-se: a análise dos sinais por faixa de frequência individual e de forma combinada; a aplicação do método na avaliação da qualidade vocal de vozes infantis e a análise dos desvios vocais por alteração das frequências formantes.

Para a continuidade deste trabalho sugere-se: estender a análise dos formantes para os sinais de vozes com o desvio vocal tensão; investigar o poder discriminatório das medidas de energia e entropia dos coeficientes de aproximação da transformada wavelet; e aplicar outros modelos matemáticos na extração das características, como por exemplo, as wavelets packets.

Referências Bibliográficas

[1] Z. Camargo, "Avaliação objetiva da voz," *Carrara-Angelis E, Furia LB, Mourão LF, Kowalski LP. A atuação da Fonoaudiologia no câncer de cabeça e pescoço. São Paulo: Lovise*, pp. 175–92, 2000.

[2] J. Hirschberg, P. Dejonckere, M. Hirano, K. Mori, H.-J. Schultz-Coulon, and K. Vrtička, "Voice disorders in children," *International journal of pediatric otorhinolaryngology*, vol. 32, pp. S109–S125, 1995.

[3] J. K. Casper and R. Leonard, *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Lippincott Williams & Wilkins, 2006.

[4] I. C. S. Lilian E. Minikel Brod, "As vogais orais do português brasileiro na fala infantil e adulta: uma análise comparativa.," *Journal of Voice*, vol. 16, no. 1, pp. 111–130, 2013.

[5] A. Parraga, "Aplicação da transformada wavelet packet na análise e classificação de sinais de vozes patológicas," Master's thesis, Porto Alegre, Brasil., 2002.

[6] S. L. N. C. Costa, *Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas*. PhD thesis, Rio de Janeiro, Brasil, 2008.

[7] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 10, pp. 1943–1953, 2006.

[8] M. L. Meredith, S. M. Theis, J. S. McMurray, Y. Zhang, and J. J. Jiang, "Describing pediatric dysphonia with nonlinear dynamic parameters," *International journal of pediatric otorhinolaryngology*, vol. 72, no. 12, pp. 1829–1836, 2008.

[9] Y. Zhang and J. J. Jiang, "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," *Journal of Voice*, vol. 22, no. 1, pp. 1–9, 2008.

[10] P. Henríquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Díaz-de María, "Characterization of healthy and pathological voice through measures based on

nonlinear dynamics," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 6, pp. 1186–1195, 2009.

[11] L. Salhi and A. Cherif, "Selection of pertinent acoustic features for detection of pathological voices," in *Modeling, Simulation and Applied Optimization (ICMSAO), 2013 5th International Conference on*, pp. 1–6, IEEE, 2013.

[12] V. J. D. Vieira, "Avaliação de distúrbios da voz por meio de análise de quantificação de recorrência," Master's thesis, Instituto Federal de Educação Ciência e Tecnologia da Paraíba, Curso de Pós-Graduação em Engenharia Elétrica, João Pessoa, 2014.

[13] M. K. Arjmandi and M. Pooyan, "An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine," *Biomedical Signal Processing and Control*, vol. 7, no. 1, pp. 3–19, 2012.

[14] E. S. Fonseca, R. C. Guido, P. R. Scalassara, C. D. Maciel, and J. C. Pereira, "Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 571–578, 2007.

[15] S. Correia, W. Costa, and S. Costa, "Detecção automática de patologias laríngeas usando a transformada wavelet discreta," in *Anais do 11th Brazilian Congress on Computational Intelligence (CBIC)*, 2013.

[16] Z. Mahmoudi, S. Rahati, M. M. Ghasemi, V. Asadpour, H. Tayarani, and M. Rajati, "Classification of voice disorder in children with cochlear implantation and hearing aid using multiple classifier fusion," *Biomedical engineering online*, vol. 10, no. 1, p. 3, 2011.

[17] R. T. S. Carvalho, *Transformada Wavelet na detecção de patologias da laringe*. PhD thesis, Universidade Federal do Ceará, 2012.

[18] L. C. Rodrigues, *Uma nova família de filtros digitais para classificação de dados com aplicações ao pré-diagnóstico de patologias na laringe*. PhD thesis, Universidade de São Paulo, 2012.

[19] N. C. d. Almeida, "Sistema inteligente para diagnóstico de patologias na laringe utilizando máquinas de vetor de suporte," 2010.

[20] L. M. d. Souza, *Detecção inteligente de patologias na laringe baseada em máquinas de vetores de suporte e na transformada wavelet*. PhD thesis, Universidade de São Paulo, 2011.

[21] E. S. Fonseca, *Wavelets, Predição Linear e LS-SVM aplicados na análise e classificação de sinais de vozes patológicas*. PhD thesis, Universidade de São Paulo, 2008.

[22] C. D. P. Crovato, “Classificação de sinais de voz utilizando a transformada wavelet packet e redes neurais artificiais,” Master’s thesis, Universidade Federal do Rio Grande do Sul. Escola de Engenharia. Programa de Pós-Graduação em Engenharia Elétrica, 2004.

[23] S. Costa, S. Correia, H. Falcão, N. Almeida, and F. Assis, “Uso da entropia na discriminação de vozes patológicas,” in *II Congresso de Inovação da Rede Norte Nordeste de Educação Tecnológica, João Pessoa, Paraíba*, 2007.

[24] W. C. A. Costa, *Análise dinâmica não linear de sinais de voz para detecção de patologias laríngeas*. PhD thesis, Campina Grande, Universidade Federal de Campina Grande, 2012.

[25] G. Niedzielska, “Acoustic analysis in the diagnosis of voice disorders in children,” *International journal of pediatric otorhinolaryngology*, vol. 57, no. 3, pp. 189–193, 2001.

[26] M. Behlau, *Voz: O livro do Especialista*, vol. 1 of *Optics and Photonics*. Rio de Janeiro, Brasil.: Revinter, 2001.

[27] J. A. d. F. D. B. F. Benjamin Pereira dos Santos Siqueira, Priscila Lemos Kallás, “Características dos sons das vogais do português falados no brasil,” *Incitel*, 2013.

[28] M. E. Dajer, “Padrões visuais de sinais de voz através de técnica de análise de não-linear.,” Master’s thesis, São Paulo, Brasil., 2006.

[29] D. M. B. M Hirano, “Exame videoestroboscópico da laringe.,” 1993.

[30] L. W. Lopes, I. L. B. Lima, L. N. A. Almeida, D. P. Cavalcante, and A. A. F. de Almeida, “Severity of voice disorders in children: correlations between perceptual and acoustic data,” *Journal of Voice*, vol. 26, no. 6, pp. 819–e7, 2012.

[31] M. Behlau and P. Pontes, “Avaliação global da voz,” *São Paulo, EPPM*, 1990.

[32] R. H. Colton, J. K. Casper, and R. Leonard, “Compreendendo os problemas de voz: uma perspectiva fisiológica ao diagnóstico e ao tratamento,” 1996.

[33] R. H. G. Martins, C. B. Hidalgo Ribeiro, B. M. Z. Fernandes de Mello, A. Branco, and E. L. M. Tavares, “Dysphonia in children,” *Journal of Voice*, vol. 26, no. 5, pp. 674–e17, 2012.

[34] J. Lopes, S. Freitas, R. Sousa, J. Matos, F. Abreu, and A. Ferreira, “A medida hnr: sua relevância na análise acústica da voz e sua estimação precisa,” *Jornadas sobre Tecnologia e Saúde, Guarda, Portugal*, 2008.

[35] P. A. L. Pontes, V. P. Vieira, M. I. R. Gonçalves, and A. A. L. Pontes, “Características das vozes roucas, ásperas e normais: análise acústica espectrográfica comparativa,” *Rev Bras Otorrinolaringolia*, vol. 68, no. 2, pp. 182–8, 2002.

[36] J. Martens, H. Versnel, and P. H. Dejonckere, "The effect of visible speech in the perceptual rating of pathological voices," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 133, no. 2, pp. 178–185, 2007.

[37] P. H. Dejonckere and J. Lebacq, "Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology," *ORL*, vol. 58, no. 6, pp. 326–332, 1996.

[38] I. Guimarães, "A ciência e a arte da voz humana," *Alcoitão, Escola Superior de Saúde de Alcoitão*, 2007.

[39] S. Simberg, A. Laine, E. Sala, and A.-M. Rönnemaa, "Prevalence of voice disorders among future teachers," *Journal of voice*, vol. 14, no. 2, pp. 231–235, 2000.

[40] S. M. R. Pinho, P. Pontes, S. M. Pinho, and P. Pontes, "Escala de avaliação perceptiva da fonte glótica: Rasat," *Vox Brasilis*, vol. 8, no. 3, pp. 11–3, 2002.

[41] A. G. Gift, "Visual analogue scales: measurement of subjective phenomena.," *Nursing research*, vol. 38, no. 5, pp. 286–287, 1989.

[42] M. E. Cline, J. Herman, E. R. Shaw, and R. D. Morton, "Standardization of the visual analogue scale.," *Nursing research*, vol. 41, no. 6, pp. 378–379, 1992.

[43] R. Yamasaki, S. Leão, G. Madazio, M. Padovani, R. Azevedo, and M. Behlau, "Correspondência entre escala analógico-visual e a escala numérica na avaliação perceptivo-auditiva de vozes," in *XVI Congresso Brasileiro de Fonoaudiologia*, pp. 24–27, 2008.

[44] W. Koenig, H. K. Dunn, and L. Y. Lacy, "The sound spectrograph," *The Journal of the Acoustical Society of America*, vol. 18, no. 1, pp. 19–49, 1946.

[45] G. P. Jotz, *Configuração laríngea, análise perceptiva auditiva e computadorizada da voz de crianças institucionalizadas do sexo masculino*. PhD thesis, Universidade Federal de Sao Paulo. Escola Paulista de Medicina, 1997.

[46] R. Kent and C. Read, "The acoustic analysis of speech," 1992.

[47] G. Fant, *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*, vol. 2. Walter de Gruyter, 1971.

[48] M. S. Behlau, *Uma análise das vogais do português brasileiro falado em São Paulo: perceptual, espectrográfica de formantes e computadorizada de freqüência fundamental*. Escola Paulista de Medicina, 1984.

[49] M. S. Behlau, O. Tosi, and P. A. d. L. Pontes, "Determinação da freqüência fundamental e suas variações em altura," *Acta Awho*, vol. 4, no. 1, pp. 5–10, 1985.

[50] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*, vol. 100. Prentice-hall Englewood Cliffs, 1978.

[51] M. M. Sondhi, "New methods of pitch extraction," *Audio and Electroacoustics, IEEE Transactions on*, vol. 16, no. 2, pp. 262–266, 1968.

[52] A. M. Noll, "Cepstrum pitch determination," *The journal of the acoustical society of America*, vol. 41, no. 2, pp. 293–309, 1967.

[53] J. E. Markel and A. H. Gray, *Linear prediction of speech*. Springer-Verlag New York, Inc., 1982.

[54] P. Ladefoged, "Vowels and consonants," *Phonetica*, vol. 58, pp. 211–212, 2001.

[55] M. Behlau, S. Pontes, P. A. de Lima, M. M. Ganança, and O. Tosi, "Análise espectrográfica de formantes das vogais do português brasileiro," *Acta Awho*, vol. 7, no. 2, pp. 74–85, 1988.

[56] M. Akay, *Time Frequency andWavelets in Biomedical Signal Processing*. IEEE Press series in biomedical Engineering, 1998.

[57] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[58] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE signal processing magazine*, vol. 8, no. LCAV-ARTICLE-1991-005, pp. 14–38, 1991.

[59] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989.

[60] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1997.

[61] Y. Y. Tang, *Wavelet theory and its application to pattern recognition*. World Scientific, 2000.

[62] V. I. B. Sablón, L. R. Mendez, and Y. Iano, "A transformada wavelet no processamento e compressão de imagens," *Revista Ciência e Tecnologia*, vol. 6, no. 9, 2010.

[63] I. Daubechies *et al.*, *Ten lectures on wavelets*, vol. 61. SIAM, 1992.

[64] I. J. Sanches, *Compressão sem perdas de projeções de tomografia computadorizada usando a transformada Wavelet*. PhD thesis, Universidade Federal do Paraná, 2001.

[65] G. Kaiser, *A friendly guide to wavelets*. Birkhäuser: Boston, 1994.

[66] L. Salhi, M. Talbi, and A. Cherif, “Voice disorders identification using hybrid approach: Wavelet analysis and multilayer neural networks,” in *Proceedings of World Academy of Science, Engineering and Technology*, vol. 35, pp. 2070–374, 2008.

[67] J. Nayak, P. S. Bhat, R. Acharya, and U. Aithal, “Classification and analysis of speech abnormalities,” *ITBM-RBM*, vol. 26, no. 5, pp. 319–327, 2005.

[68] P. Kukharchik, D. Martynov, I. Kheidorov, and O. Kotov, “Vocal fold pathology detection using modified wavelet-like features and support vector machines,” in *Proceedings of 15th European Signal Processing Conference*, 2007.

[69] J. Macedo, “Transformadas e decomposição de sub-bandas,” *Departamento de Informática da Universidade do Minho & Faculdade de Engenharia da Universidade Católica de Angola.*

[70] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Başar, “Wavelet entropy: a new tool for analysis of short duration brain electrical signals,” *Journal of neuroscience methods*, vol. 105, no. 1, pp. 65–75, 2001.

[71] O. Farooq and S. Datta, “Phoneme recognition using wavelet based features,” *Information Sciences*, vol. 150, no. 1, pp. 5–15, 2003.

[72] R. T. Vieira, N. Brunet, S. C. Costa, S. Correia, B. G. A. Neto, and J. M. Fechine, “Combining entropy measures and cepstral analysis for pathological voices assessment,” *Journal of Medical and Biological Engineering*, vol. 32, no. 6, pp. 429–435, 2012.

[73] R. Behroozmand and F. Almasganj, “Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients speech signal with unilateral vocal fold paralysis,” *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 474–485, 2007.

[74] L. Zunino, D. Perez, M. Garavaglia, and O. Rosso, “Wavelet entropy of stochastic processes,” *Physica A: Statistical Mechanics and its Applications*, vol. 379, no. 2, pp. 503–512, 2007.

[75] C. E. Shannon, “A mathematical theory of communication,” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[76] G. Fant, *Speech acoustics and phonetics: Selected writings*, vol. 24. Springer Science & Business Media, 2006.

[77] Marco Soares, “Informações técnicas,” 2009. Access date: 05 jan. 2015.

APÊNDICES

APÊNDICE A

Análise dos Formantes

Este apêndice traz, de maneira detalhada, a análise dos desvios vocais infantis por meio da análise dos formantes.

A seção a seguir, apresenta a metodologia empregada nesta pesquisa.

A.1 – Metodologia

A metodologia empregada neste estudo esta decrita Figura A.1.

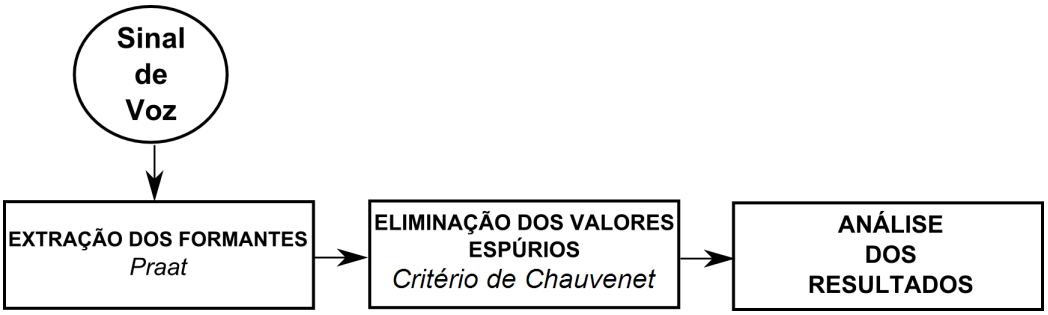


Figura A.1 – Diagrama em blocos da metodologia empregada.

A base de dados utilizada é a mesma em todo o trabalho. Os sinais de voz da base de dados tem os seus formantes extraídos através do software *Praat* (**Apêndice - B**). O *Praat* é um *software* open source que integra várias funcionalidades de análise de sinais de voz. Este *software* proporciona um vasto número de métodos de medição, nomeadamente de perturbação da frequência fundamental, amplitude e de ruído, bem como análise espectral. Permite ainda a utilização de métodos não convencionais como a síntese de voz e redes neuronais. Este *software* é essencialmente orientado para as áreas de análise acústica de sinais de voz disfônica, como a Terapia da Fala.

Os formantes extraídos de cada segmento do sinal de voz são submetidos ao critério de Cheauvenet, para eliminação dos valores duvidosos ou espúrios. O critério de Cheauvenet especifica que um valor medido pode ser rejeitado se a probabilidade **m** de obter o desvio em relação à média é menor que $\frac{1}{2n}$ com base na razão do desvio em relação ao desvio padrão para vários valores de **n** conforme este critério (**Apêndice - C**).Em seguida, é calculada a média

dos valores dos formantes de todos os segmento de cada sinal de voz e esses valores são comparados à valores presentes na literatura e entre si, quanto a presença ao não do desvio vocal [77].

A.2 – Resultados

Nesta pesquisa,a análise dos formantes foi dividida em dois estudos de caso: sinal de voz saudável *x* sinal de voz com desvio vocal (rugosidade e/ou sorposidade) e sinal de voz com rugosidade x sinal de voz com soprosidade. Para o estudo de caso 1, os valores de F1, F2 e F3, para o grupo de sinais que apresentam algum desvio vocal (rugosidade e/ou soprosidade) apresentam valores superiores quando comparado ao grupo de sinais com voz normal, o que evidencia, uma alteração dos formantes do sinal de voz na presença de algum tipo de distúrbio da voz.

Pra o estudo de caso 2, quando comparamos o grupo de sinal de voz com soprosidade, com o grupo de sinal de voz com rugosidade, os valores dos três primeiros formantes, para o grupo com soprosidade apresentam-se mais elevados, mais agudos do que o grupo com rugosidade. Desta forma, podemos justificar esta elevação nos valores dos formantes, na presença de ar turbulento, presente no desvio vocal soprosidade, que pode estar atrelada a um fechamento glótico insuficiente.

Os resultados obtidos são apresentados nas Tabelas A.1, A.2 e A.3 e nas Figuras A.2, A.3 e A.4.

Tabela A.1 – Valores mínimo, máximo e médios dos formantes para sinais de voz saudável.

Sem desvio			
	F1	F2	F3
Mínimo	305,756	1.992,011	2.342,421
Máximo	3.171,572	3.774,072	3.973,041
Média	946,907	2.779,737	2.857,796

Tabela A.2 – Valores mínimo, máximo e médios dos formantes para sinais de voz com Rugosidade.

Rugosidade			
	F1	F2	F3
Mínimo	368,840	2.404,540	2.239,546
Máximo	3.188,790	3.456,290	5.277,375
Média	1.179,617	2.791,850	3.334,040

Tabela A.3 – Valores mínimo, máximo e médios dos formantes para sinais de voz com Soprosidade.

Soprosidade			
	F1	F2	F3
Mínimo	205,509	1.689,599	1.983,280
Máximo	3.568,698	5.129,342	5.433,368
Média	2.701,647	3.293,284	4.924,548

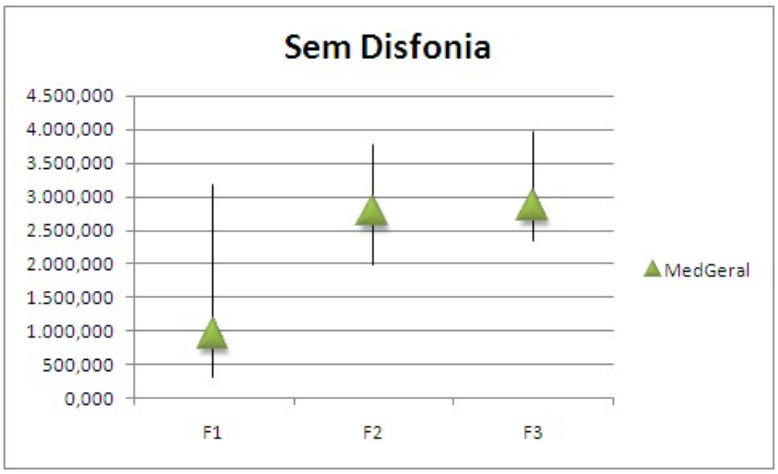


Figura A.2 – Gráfico dos valores médios dos formantes para crianças com voz saudável.

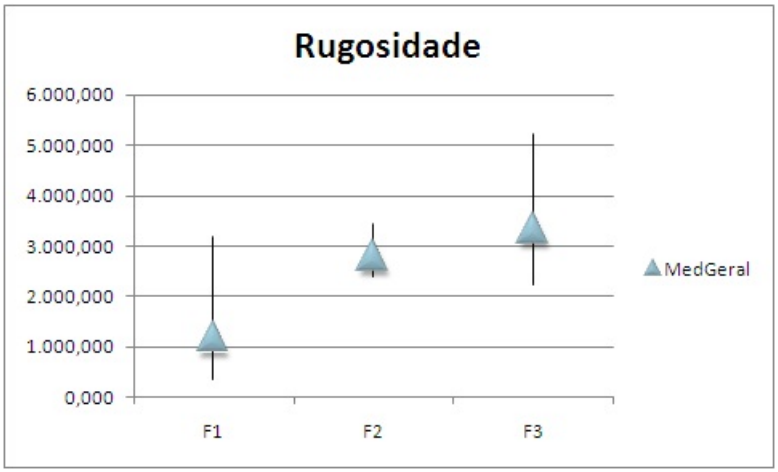


Figura A.3 – Gráfico dos valores médios dos formantes para sinais de voz com Rugosidade.

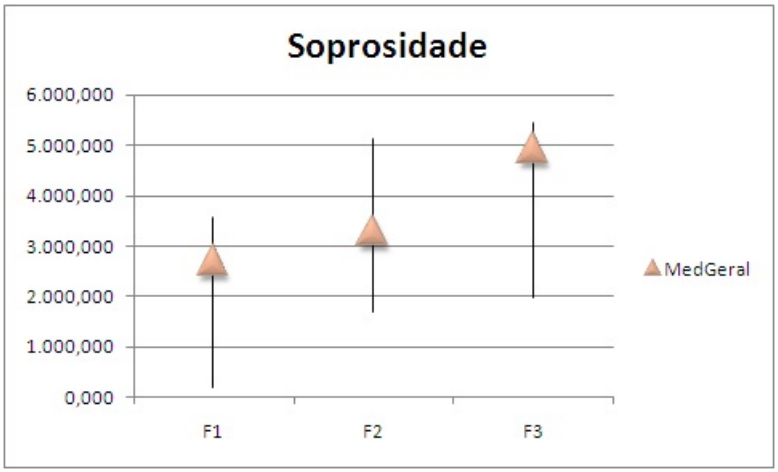


Figura A.4 – Gráfico dos valores médios dos formantes para sinais de voz com Soprosidade.

As Figuras A.5, A.6 e A.7, são do *software Praat*, e apresentam o espectro e espectrograma para um sinal de voz escolhidos, apenas, por serem representativo de cada grupo

utilizado neste experimento: uma voz sem desvio, uma voz com rugosidade e uma voz com soprosidade.

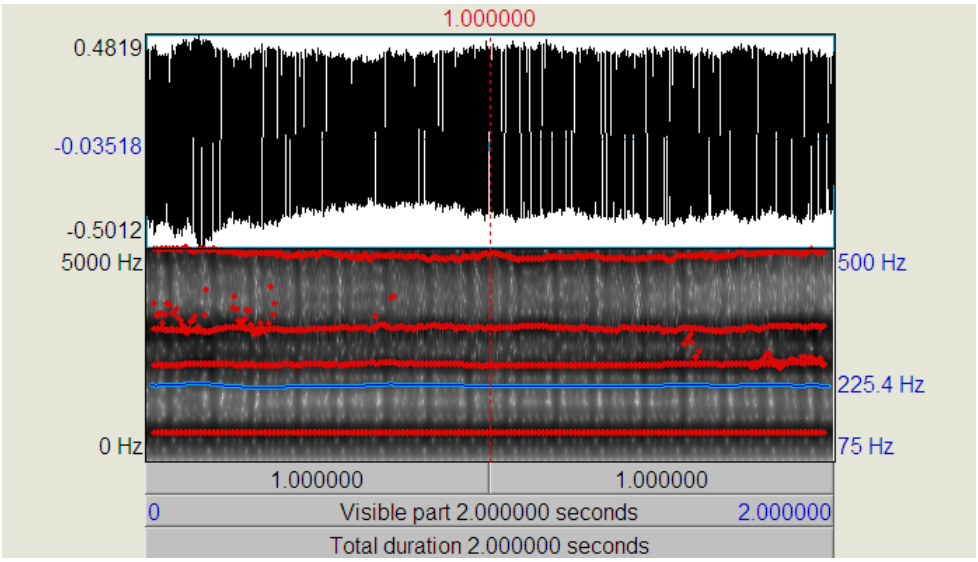


Figura A.5 – Espectro e Espectrograma de uma voz sem desvio.

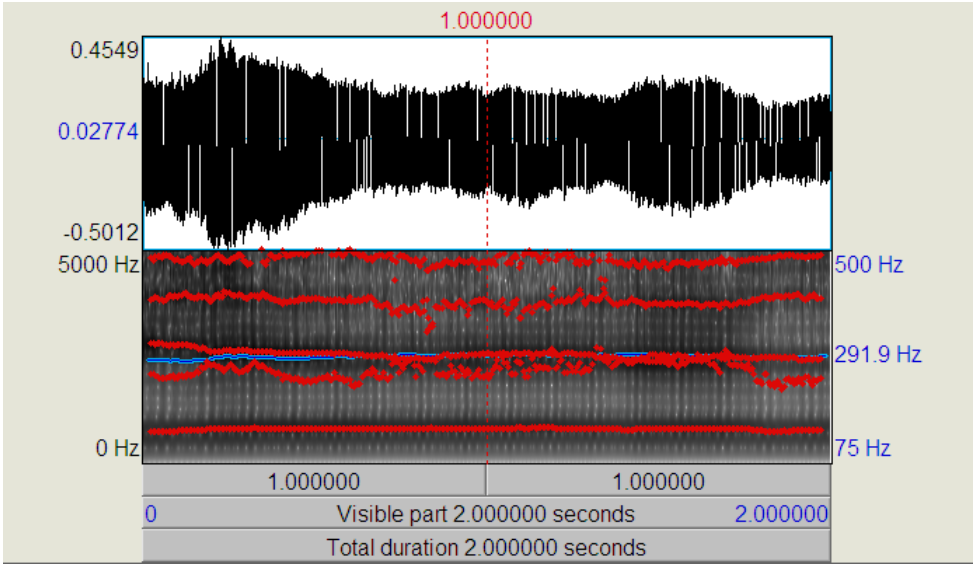


Figura A.6 – Espectro e Espectrograma de uma voz com Rugosidade.

Os valores apresentados neste estudo aproximam-se dos valores encontrados pela pesquisadora Behlau (1984). As diferenças entre os valores dos fomantes de vozes normais apresentados por Behlau e os presentes nesta pesquisa podem ser justificado devido à questão regional da fala, bem como, pelo modo de gravação das vozes e método de extração dos formantes.

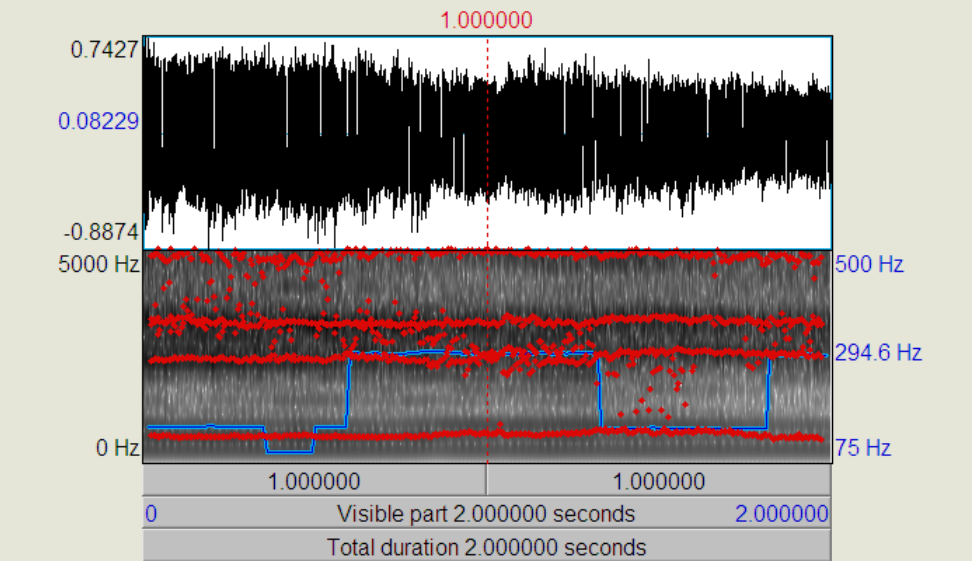


Figura A.7 – Espectro e Espectrograma de uma voz com Soprosidade.

A.2.1 – Discussão dos Resultados

A análise de desvios vocais através dos formantes, permitiu visualizar de maneira quantitativa, a presença dos desvios vocais em sinais de vozes infantis.

Os resultados obtidos nesta pesquisa, para os três primeiros formantes dos sinais de vozes infantis sem desvio, aproximam-se dos valores apresentados por [48] em sua pesquisa, validando o modelo de extração de formantes empregado nesta pesquisa.

Os resultados obtidos nesta pesquisa, fazem da análise dos formantes do sinal de voz, uma ferramenta de auxílio ao diagnóstico para profissionais da fala, bem como no acompanhamento da evolução de fonoterapia.

Os dados encontrados nos formantes F1, F2 e F3 foram difíceis de serem justificados pela literatura já que a correlação entre os formantes e os desvios vocais em crianças não foi encontrada.

APÊNDICE B

Utilizando o Praat para obtenção dos Formantes

Neste apêndice será detalhado como como foi realizada a extração dos formantes nesta pesquisa, utilizando o software Praat, versão 5.4.04.

B.1 – Passo a Passo da Obtenção dos Formantes

Nesta pesquisa a obtenção dos formantes se deu da seguinte forma:

1º Passo: Na barra de Ferramentas da janela inicial do programa, seleciona-se a opção “Open” e em seguida “ Read from file ... ” para carregar o sinal de voz que se deseja utilizar (Figura B.1).

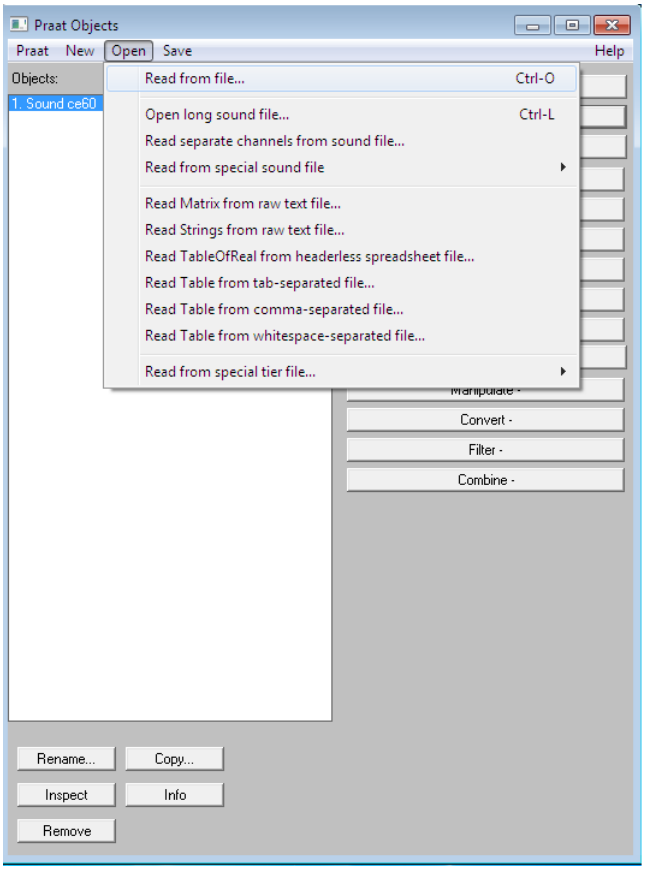


Figura B.1 – Janela Inicial do programa. Carregando o sinal de voz a ser utilizado.

2º Passo: Após o sinal de voz ser carregado, é selecionada a guia “Analyse Spectrum” para escolha do método de extração de formantes. Nesta pesquisa a opção escolhida foi o método “ To Formant (burg)...”, como pode ser visto na Figura B.2. Este algoritmo, inicialmente, encontra o número máximo de formantes definida em todo o intervalo entre 0 Hz e a formante máxima. Os formantes inicialmente encontrados podem, portanto, por vezes, ter frequências muito baixas (perto de 0 Hz) ou muito altas (perto do formante máximo). Para que você seja capaz de identificar a F1 e F2, todos os formantes abaixo de 50 Hz e todos os formantes acima do formante máximo, são eliminados.

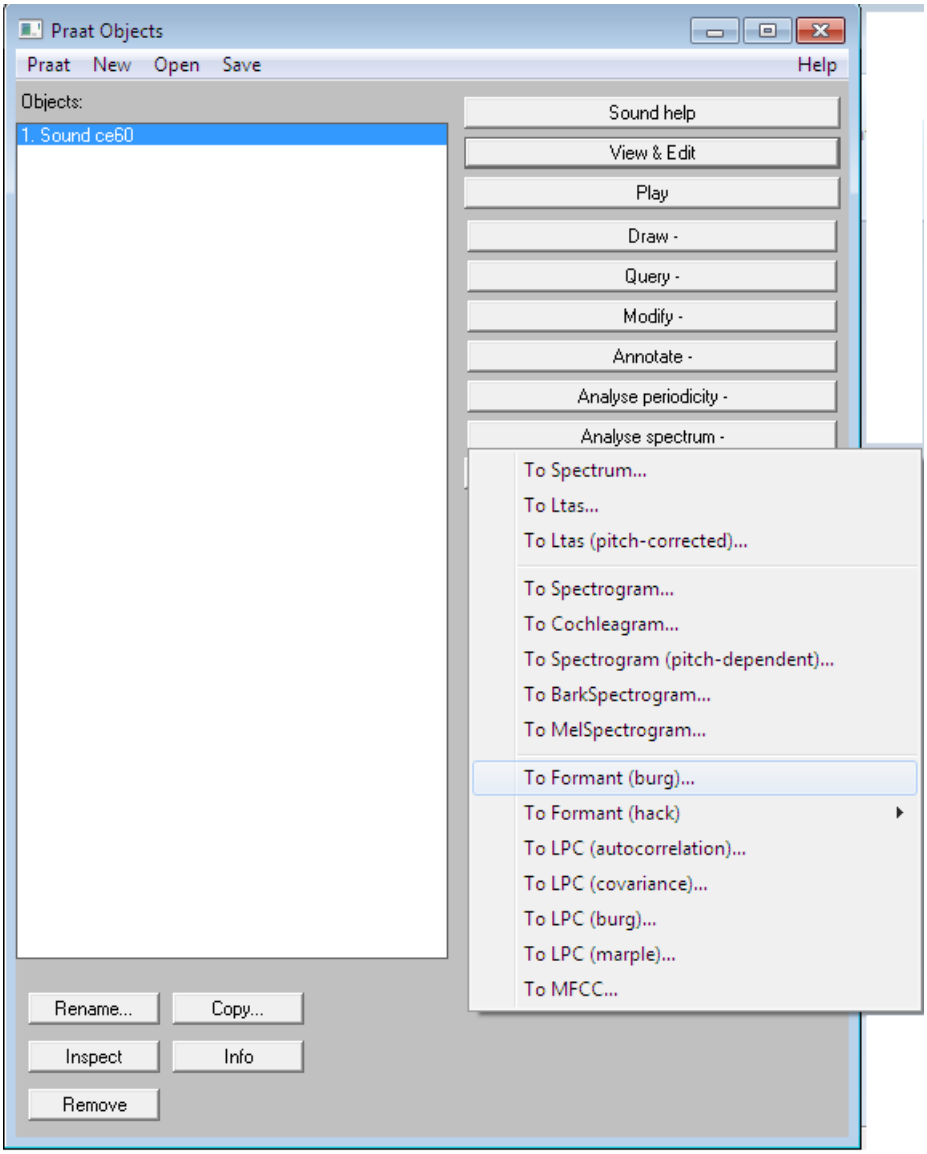


Figura B.2 – Escolha do método de extração dos formantes.

3º Passo: Aparecerá uma janela para configuração do método de extração dos formantes (Figura B.3). Cada opção deve ser configurada de modo a se obter as medidas corretamente.

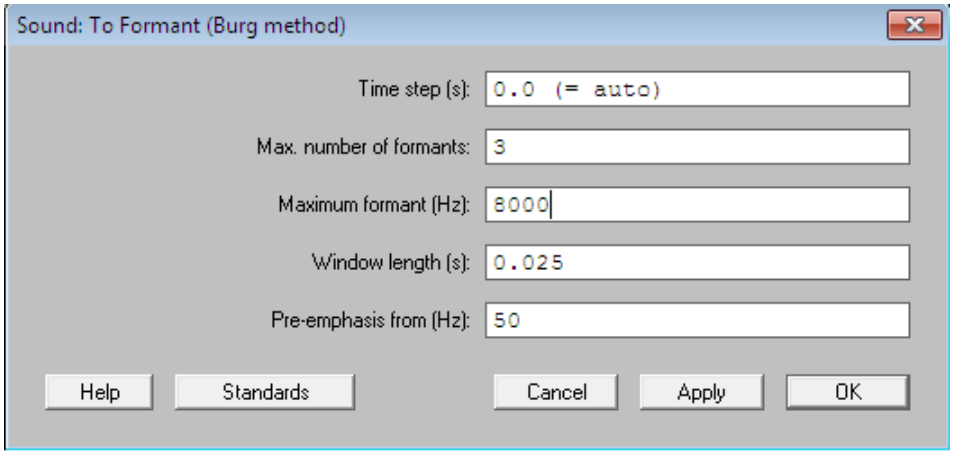


Figura B.3 – Configuração do método de extração dos formantes.

Time step (s): Define o intervalo entre os segmentos a serem analisados. Se o intervalo de tempo definido for de 0,0 segundo , o Praat irá utilizar um intervalo de tempo que é igual a 25% do comprimento da janela de análise.

Max number of formants: Esta opção permite que você determine o numero máximo de formantes por segmento do sinal de voz que você deseja extrair. O Praat permite que sejam extraídos no máximo 5 formantes por segmento.

Max formant (Hz): Este comando permite que você defina o teto máximo em Hertz que deseja usar. O valor padrão de 5500 Hz é adequado para adultos do sexo femino. Para um adulto do sexo masculino, deve-se usar 5000 Hz; se você usar 5500 Hz para um homem adulto , você pode acabar eliminando formantes presentes em regiões de baixa frequência. Em crianças, deve-se usar um valor em torno de 8000 Hz (experimento realizado em crianças com vogal sustentada).

Window length (s):

Este comando permite que se defina o tamanho da janela em segundos. O comprimento real é sempre duas vezes o valor determinado, pois o Praat utiliza uma janela de análise Gaussiana com lóbulos laterais abaixo -120 dB. Por exemplo , se o comprimento da janela é de 0,025 segundos, a duração da janela Gaussiana real é 0,050 segundos.

Pre-emphasis from (hz):

Este comando aplica uma pré-ênfase ao sinal. Ele permite que o espectro do sinal se torne mais plano e dessa forma possibilite uma melhor análise dos formantes. Se este valor é de 50 Hz, as frequências abaixo de 50 Hz não são amplificadas , as frequências de cerca de 100 Hz são amplificados em 6 dB , as frequências de cerca de 200 Hz são amplificados por 12 dB , e assim por diante.

4º Passo:

O passo anterior irá gerar o arquivo “Formante NomeDoArquivoDeÁudio” como ilustrado na Figura B.4. Com o arquivo selecionado, utlizamos a guia “Tabulete” e em seguida o comando

"List" (Figura B.5), para abrir o arquivo que contém os Formates de todos os segmentos extráidos do sinal de voz, para o número de Formantes determinado no Passo 3 Figura B.6.

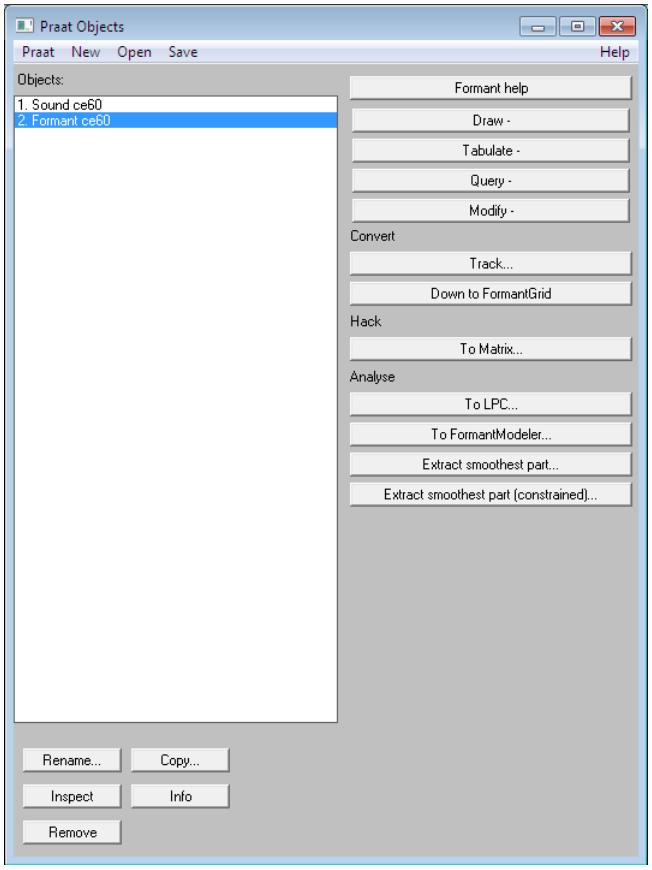


Figura B.4 – Arquivo gerado pelo passo anterior contendo os Formantes extraídos.

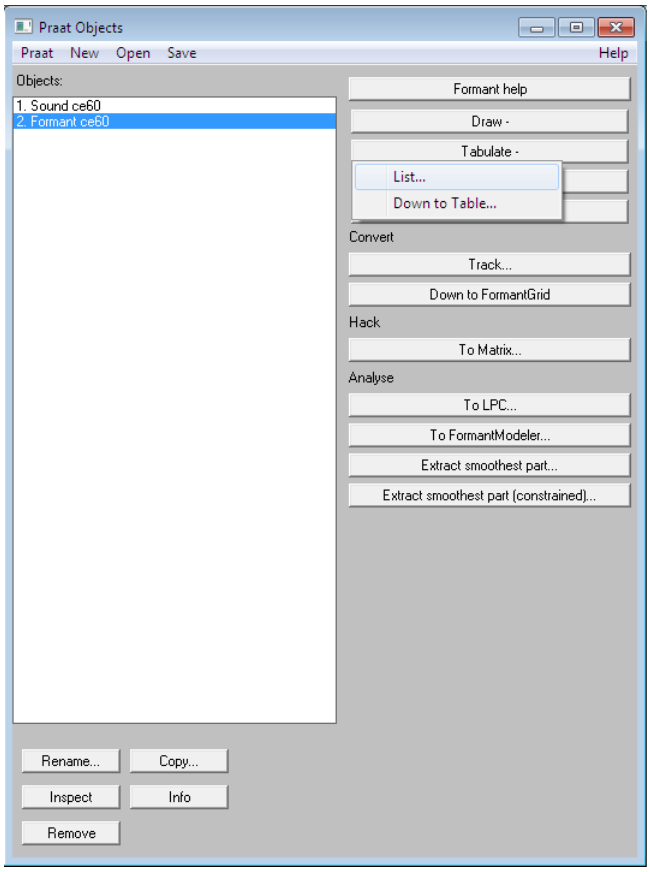
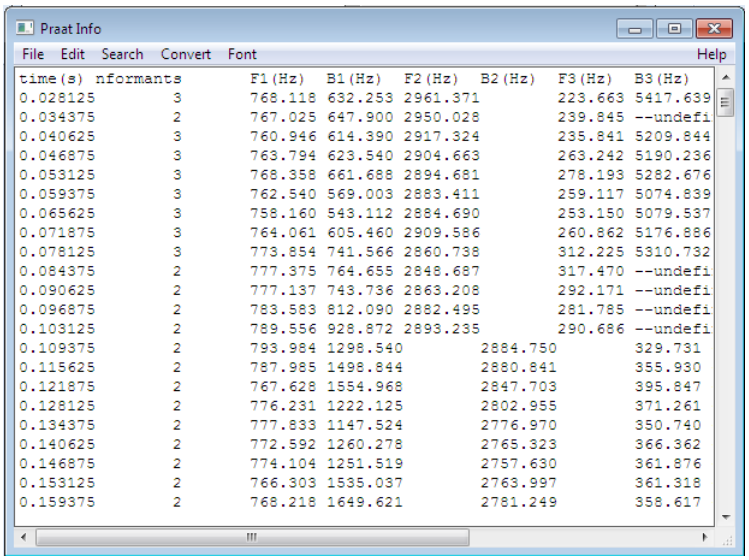


Figura B.5 – Comando para abrir o arquivo que contém os Formantes.



time(s)	nformants	F1(Hz)	B1(Hz)	F2(Hz)	B2(Hz)	F3(Hz)	B3(Hz)
0.028125	3	768.118	632.253	2961.371		223.663	5417.639
0.034375	2	767.025	647.900	2950.028		239.845	--undefi
0.040625	3	760.946	614.390	2917.324		235.841	5209.844
0.046875	3	763.794	623.540	2904.663		263.242	5190.236
0.053125	3	768.358	661.688	2894.681		278.193	5282.676
0.059375	3	762.540	569.003	2883.411		259.117	5074.839
0.065625	3	758.160	543.112	2884.690		253.150	5079.537
0.071875	3	764.061	605.460	2909.586		260.862	5176.886
0.078125	3	773.854	741.566	2860.738		312.225	5310.732
0.084375	2	777.375	764.655	2848.687		317.470	--undefi
0.090625	2	777.137	743.736	2863.208		292.171	--undefi
0.096875	2	783.583	812.090	2882.495		281.785	--undefi
0.103125	2	789.556	928.872	2893.235		290.686	--undefi
0.109375	2	793.984	1298.540		2884.750		329.731
0.115625	2	787.985	1498.844		2880.841		355.930
0.121875	2	767.628	1554.968		2847.703		395.847
0.128125	2	776.231	1222.125		2802.955		371.261
0.134375	2	777.833	1147.524		2776.970		350.740
0.140625	2	772.592	1260.278		2765.323		366.362
0.146875	2	774.104	1251.519		2757.630		361.876
0.153125	2	766.303	1535.037		2763.997		361.318
0.159375	2	768.218	1649.621		2781.249		358.617

Figura B.6 – Arquivo com os Formantes gerados

APÊNDICE C

Utilizando o Critério de Chauvenet

Neste apêndice será detalhado como são excluídos os valores anôma-los pelo critério de *Chauvenet*.

C.1 – Critério de *Chauvenet*

Quando se realiza uma sequência de **n** medições de um mesmo objeto, é possível a ocorrência de alguns resultados estranhamente fora da distribuição estatística esperada.

Os resultados anômalos ou espúrios podem afetar sensivelmente a média e comprometer a exatidão do processo. É razoável, portanto, algum critério para seu descarte. Mas isso não deve ser visto como regra geral. Resultados inesperados às vezes podem ser decisivos no estudo de certos fenômenos [77].

O critério de *Chauvenet* é um dos métodos mais simples e mais usados para indicar os resultados a desprezar.

Seja uma seqüência de **n** medições que estatisticamente seguem o comportamento comum da distribuição normal. Então pode-se rejeitar resultados cujas probabilidades sejam menores que

$$\frac{1}{2n} \tag{C.1}$$

Isso significa que resultados considerados “bons” estão dentro de uma faixa cuja probabilidade é:

$$1-\frac{1}{2n} \tag{C.2}$$

As faixas de probabilidades são dadas em termos de desvio-padrão. Assim, para cada valor de **n** pode ser calculada a probabilidade conforme Eq. C.2 e, por integração matemática da função de densidade da distribuição normal, determinado um coeficiente **C** correspondente ao número de desvios padrão para a faixa de valores considerados aceitáveis.

A Tabela C.1 apresenta os valores de **C** para alguns valores de **n**.

Portanto, a faixa de valores aceitáveis é dada por:

Tabela C.1 – Critério de Chauvenet para rejeição de valor medido.

n	C
3	1,38
4	1,54
5	1,65
6	1,73
7	1,80
8	1,87
9	1,91
10	1,96
15	2,13
20	2,24
25	2,33
50	2,57
100	2,81
300	3,14
500	3,29
1000	3,48

Fonte: [77]

$$media \pm C + \sigma, \tag{C.3}$$

onde σ é o desvio padrão

Valores fora dessa faixa podem ser descartados segundo o critério de *Chauvenet*.

Exemplo numérico

A tabela C.2 dá os resultados de uma série hipotética de **n** = 10 medidas da massa de um determinado corpo.

Calculam-se a média e o desvio padrão da amostra segundo fórmulas estatísticas:

Média:

$$\overline{x} = \frac{\sum xi}{n} \backsim 2,68 \tag{C.4}$$

Desvio Padrão:

$$s = \sqrt{\frac{(x_i - \overline{x})^2}{n-1}} \backsim 0,76 \tag{C.5}$$

Para **n** = 10, o coeficiente do critério de Chauvenet é C = 1,96 segundo a Tabela 2.1. Multiplicando pelo desvio padrão, temos:

$C \cdot s = 1,96 \cdot 0,76 \approx 1,49$

Portanto, os valores confiáveis devem estar entre:

Tabela C.2 – Tabela com valores para série hipotética.

Medida	Valor
01	2,41
02	2,42
03	2,43
04	2,43
05	2,44
06	2,44
07	2,45
08	2,46
09	2,47
10	2,48

Fonte: [77]

2,68 - 1,49 = 1,19 e 2,68 + 1,49 = 4,17

Segundo o critério, pode-se rejeitar a medida 10 (4,85) da Tabela 4.1.