Universidade Federal do Ceará
Centro de Tecnologia
Departamento de Engenharia de Teleinformática
Programa de Pós-Graduação em Engenharia de Teleinformática

Raphael Torres Santos Carvalho

Transformada Wavelet na Detecção de Patologias da Laringe

Fortaleza – Ceará
Março 2012

Raphael Torres Santos Carvalho

Transformada Wavelet na Detecção de Patologias da Laringe

Dissertação de Mestrado apresentado à Coordenação do Curso de Pós-Graduação em Engenharia de Teleinformática da Universidade Federal do Ceará como parte dos requisitos para obtenção do grau de ***Mestre em Engenharia de Teleinformática****.*

Área de Concentração: Sinais e Sistemas

Orientador : Prof. Dr. Charles Casimiro Cavalcante

Fortaleza – Ceará
Março 2012

Resumo

A quantidade de métodos não invasivos de diagnóstico tem aumentado devido à necessidade de exames simples, rápidos e indolores. Por conta do crescimento da tecnologia que fornece os meios necessários para a extração e processamento de sinais, novos métodos de análise têm sido desenvolvidos para compreender a complexidade dos sinais de voz. Este trabalho de dissertação apresenta uma nova ideia para caracterizar os sinais de voz saudável e patológicos baseado em uma ferramenta matemática amplamente conhecida na literatura, a Transformada *Wavelet* (WT). O conjunto de dados utilizado neste trabalho consiste de 60 amostras de vozes divididas em quatro classes de amostras, uma de indivíduos saudáveis e as outras três de pessoas com nódulo vocal, edema de Reinke e disfonia neurológica. Todas as amostras foram gravadas usando a vogal sustentada /a/ do Português Brasileiro. Os resultados obtidos por todos os classificadores de padrões estudados mostram que a abordagem proposta usando WT é uma técnica adequada para discriminação entre vozes saudável e patológica, e apresentaram resultados similares ou superiores a da técnica clássica quanto à taxa de reconhecimento.

Palavras-chaves: Reconhecimento de Voz, Extração de Características, Transformada *Wavelet*, Nódulo Vocal, Edema de Reinke, Disfonia Neurológica.

Abstract

The amount of non-invasive methods of diagnosis has increased due to the need for simple, quick and painless tests. Due to the growth of technology that provides the means for extraction and signal processing, new analytical methods have been developed to help the understanding of analysis of the complexity of the voice signals. This dissertation presents a new idea to characterize signals of healthy and pathological voice based on one mathematical tools widely known in the literature, Wavelet Transform (WT). The speech data were used in this work consists of 60 voice samples divided into four classes of samples: one from healthy individuals and three from people with vocal fold nodules, Reinke's edema and neurological dysphonia. All the samples were recorded using the vowel /a/ in Brazilian Portuguese. The obtained results by all the pattern classifiers studied indicate that the proposed approach using WT is a suitable technique to discriminate between healthy and pathological voices, since they perform similarly to or even better than classical technique, concerning recognition rates.

Keywords: Voice Recognition, Feature Extraction, Wavelet Transform, Artificial Neural Networks, Vocal Fold Nodules, Reinke's Edema, Neurological Dysphonia.

Dedico este trabalho a minha família
pelo constante apoio, incentivo
e admiração.

Agradecimentos

Agradeço primeiramente a Deus por todas as bençãos derramadas durante toda minha vida.

À minha família pelo carinho, apoio e incentivo que me permitiram chegar até aqui.

À minha amada Natália, pela paciência, compreensão, carinho e incentivo para que eu pudesse concluir esta importante etapa da minha vida.

Ao Professor Dr. Charles Casimiro Cavalcante, meu orientador, pela indicação, apoio, confiança em mais essa etapa, paciência, dedicação e disponibilidade apresentadas durante este trabalho e também pelas condições que me proporcionou na realização deste trabalho.

Ao Professor Dr. Paulo César Cortez, pelo apoio, pela amizade, pela confiança depositada e por ter possibilitado o início desta jornada.

Aos demais professores e funcionários do Departamento de Engenharia de Teleinformática que de forma direta ou indireta participaram do desenvolvimento deste trabalho.

À CAPES pelo suporte financeiro.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

Transformada *Wavelet*

$\psi(x)$	*Wavelet* Mãe
τ	Parâmetros de translação
a	Parâmetro de escalonamento
$\phi(x)$	Função de escalonamento *wavelet*
$h_\psi(t)$	Coeficiente passa-alta da função *wavelet*
$h_\phi(k)$	Coeficiente passa-baixa da função *wavelet*
$A(j_0, k)$	Coeficiente de aproximação (baixa frequência)
$D(j, k)$	Coeficiente de detalhes (alta frequência)
j_0	Escala inicial da decomposição *wavelet*
t	Variável de tempo contínuo
k	Variável de tempo discreto
E_j	Energia dos coeficientes no nível de decomposição j
$E(k)$	Energia de cada amostra de tempo k
E_{total}	Energia total do sinal
p_j	Energia *wavelet* relativa (distribuição de probabilidade da energia dos coeficientes *wavelet*)
$S_{WT}(p)$	Entropia *wavelet* total
$S_{WT}(p\|q)$	Entropia *wavelet* relativa

Reconhecimento de Padrões

$\mathbf{x}$	Vetor de características (Padrão)
$\mathbf{X}$	Espaço de características

$p(\mathbf{x}\|w_j)$	Função de densidade de probabilidade condicional do padrão $\mathbf{x}$ dado a classe w_j
$P(w_j)$	Probabilidade de ocorrência *a priori* da classe w_j
$P(w_j\|\mathbf{x})$	Probabilidade *a posteriori* de ocorrer a classificação na classe w_j dado o vetor $\mathbf{x}$
$P(w_j)$	Probabilidade *a priori*
p($\mathbf{x}$)	Função de densidade de probabilidade da mistura de classes
$\widehat{w}(\mathbf{x})$	Regra de decisão Bayes
$\varphi(.)$	Função de ativação não linear dos neurônios de uma rede neural
θ	Limiar ou *bias* de um neurônio
$\mathbf{w}$	Vetor de pesos associado a um neurônio da camada oculta
$\mathbf{m}$	Vetor de pesos associado a um neurônio da camada de saída
$\mathbf{W}$	Matriz de pesos da camada oculta
$\mathbf{M}$	Matriz de pesos da camada de saída
$\mathbf{u}$(t)	Vetor de ativações dos neurônios ocultos na iteração t
$\mathbf{z}$(t)	Saídas dos neurônios da camada oculta
$\mathbf{a}$(t)	Vetor de ativações dos neurônios da camada de saída na iteração t

Lista de Siglas

LP Lâmina Propria

RS Espaço de Reinke (do inglês *Reinke's Space*)

FT Transformada de Fourier (do inglês *Fourier Transform*)

STFT Transformada de Fourier de Tempo Curto (do inglês *Short Time Fourier Transform*)

WT Transformada *Wavelet* (do inglês *Wavelet Transform*)

CWT Transformada *Wavelet* Contínua (do inglês *Continuous Wavelet Transform*)

DWT Transformada *Wavelet* Discreta (do inglês *Discrete Wavelet Transform*)

RWE Entropia *Wavelet* Relativa (do inglês *Relative Wavelet Entropy*)

WP Pacotes *Wavelet* (do inglês *Wavelet Packet*)

DFT Transformada Discreta de Fourier (do inglês *Discret Fourier Transform*)

RP Reconhecimento de Padrões

RNA Rede Neural Artificial

MLP Perceptron Multicamadas (do inglês *Multilayer Perceptron*)

MSE Erro Quadrático Médio (do inglês *Mean Squared Error*)

AVC Acidente Vascular Cerebral

ELA Esclerose Lateral Amiotrófica

PCA Análise por Componentes Principais (do inglês *Principal Components Analysis*)

CVA Análise por Variáveis Canônicas (do inglês *Canonical Variate Analysis*)

LDA Análise por Discriminante Linear (do inglês *Linear Discriminant Analysis*)

NN Vizinho mais próximo (do inglês *Nearest Neighborhood*)

KNN k-Vizinhos mais próximo (do inglês *K-Nearest Neighborhoods*)

ELM *Extreme Learning Machine*

WSS Estacionário no sentindo amplo (do inglês *Wide-Sense Stationary*)

Capítulo 1

Introdução

A voz é um dos principais meios de comunicação humano e, como um sinal acústico, contém informações importantes sobre algumas características individuais. A estrutura biomecânica vocal, em associação com variáveis aerodinâmicas, desempenha um papel importante na produção da voz e está ligada às mudanças na qualidade vocal (SCALASSARA *et al.*, 2009).

A qualidade da voz depende do modo de fechamento e abertura da glote e da vibração das pregas vocais. Certas alterações laríngeas impedem que as pregas vocais tenham uma vibração glotal harmônica. Os principais fatores que determinam a vibração vocal são (ZWETSCH *et al.*, 2006):

i. posição da prega vocal ou a extensão em que as pregas vocais são aduzidas (abertas) ou abduzidas (fechadas);

ii. mioelasticidade, ou o grau de elasticidade das pregas vocais (determinado pela posição e grau de tensão decorrente da contração do músculo vocal);

iii. nível de pressão do ar através das pregas vocais.

As alterações das pregas vocais, devido à presença de patologias na laringe, causam mudanças significativas em seus padrões vibratórios, afetando a qualidade da voz. Certas alterações laríngeas podem causar dificuldade no diagnóstico, pois são muito semelhantes, apesar de apresentarem origens e alterações fisiopatológicas diferentes. Essas alterações podem ser classificadas como físicas, neuromusculares, traumáticas e psicogênicas, e todas afetam diretamente a qualidade da voz. Algumas

alterações laríngeas representam a grande maioria dos atendimentos de pacientes com alterações da voz. Estas alterações são: nódulo vocal, cisto vocal, pólipo vocal, edema de Reinke e sulco vocal. Na maioria dos casos, tais alterações laríngeas produzem uma rouquidão com características típicas de cada uma quando analisadas por médicos otorrinolaringologistas ou fonoaudiólogos (ZWETSCH *et al.*, 2006).

Desta forma, abre-se a possibilidade de avaliar a qualidade vocal e as patologias que a modificam a partir da análise das características do sinal acústico vocal.

1.1 Motivação

Para conseguir avaliar a qualidade vocal do paciente, os médicos utilizam diversas técnicas. Algumas são técnicas subjetivas, baseadas na audição da voz pelo médico especialista, sendo essas totalmente dependentes da experiência do profissional. Outras técnicas permitem a inspeção direta da laringe e a visualização do comportamento vibratório das pregas vocais através do uso de técnicas laringoscópicas como a videolaringoscopia direta e a videoestroboscopia.

A videolaringoscopia direta é um exame realizado pelo médico com o objetivo de visualizar a laringe utilizando uma microcâmera (PARRAGA, 2002). Esse exame é efetuado em centro cirúrgico, sob anestesia geral, possibilitando uma visualização direta da laringe e de eventuais lesões estruturais mínimas que não tenham sido identificadas na videolaringoscopia normal. A videolaringoscopia direta é utilizada para orientar a correção cirúrgica das patologias laríngeas e das pregas vocais.

A videoestrobocopia permite a visualização do comportamento vibratório das pregas vocais (PARRAGA, 2002). As cordas vocais vibram em uma velocidade que não permite a sua visualização pelo olho humano. Na videoestroboscopia da laringe, utiliza-se uma luz estroboscópica que torna a vibração aparente ao exame, permitindo visualizar pequenas alterações nas cordas vocais que não são identificáveis no exame convencional.

Essas técnicas, embora sejam mais objetivas, são consideradas invasivas e causam desconforto aos pacientes, além de utilizarem instrumentos sofisticados e caros como fontes de luz especial, instrumentos endoscópicos e câmeras de vídeo especializadas (FALCÃO *et al.*, 2008). Essas técnicas visuais resultam em uma avaliação qualitativa, cujos resultados são difíceis de serem quantificados e que necessitam do conhecimento e da experiência do especialista.

A simplicidade e a natureza não invasiva têm feito das técnicas de processamento digital de sinais, por meio da análise acústica, uma eficiente ferramenta para o diagnóstico das desordens provocadas por patologias na laringe, classificação de doenças da voz e acompanhamento da evolução de tratamentos. Em relação aos métodos tradicionais, as técnicas baseadas na análise acústica apresentam como vantagens:

i. propiciar um exame mais confortável ao paciente;

ii. fornecer uma avaliação quantitativa da doença;

iii. possibilitar o desenvolvimento de sistemas automáticos de auxílio ao diagnóstico por computador por um baixo custo.

E, portanto, podem ser aplicadas como técnicas auxiliares aos métodos baseados na inspeção direta das cordas vocais, diminuindo a regularidade dos exames mais invasivos (FALCÃO *et al.*, 2008).

Assim, a análise acústica tradicional é uma ferramenta essencial e familiar aos médicos e fonoaudiólogos. As medidas derivadas das características do processo natural de produção da fala como: freqüência fundamental (*pitch*), formantes do sinal de voz, perturbações de freqüência (*jitter*), perturbações de amplitude (*shimmer*), energia do sinal, análise cepstral e técnicas de predição linear, constituem o modelo tradicional de análise acústica. Entretanto, as mudanças morfológicas provocadas pela presença das patologias causam alterações nos movimentos vibratórios das pregas vocais e no sinal acústico. Em casos em que a voz do paciente tem um alto nível de desordem devido a uma patologia severa, há uma grande dificuldade em se obter essas medidas, tornando essas técnicas ineficientes.

Novas técnicas para extração de características de sinais de vozes patológicas têm sido propostas com o objetivo de contornar essas dificuldades em obter as medidas lineares e classificar esses sinais eficientemente. Dentre os trabalhos estudados para esse pesquisa, tem-se: medidas de entropia relativa (SCALASSARA *et al.*, 2009), decomposição autoregressiva e rastreamento de pólo (SCALASSARA *et al.*, 2007), discriminante local e algoritmo genético (HOSSEINI; ALMASGANJ; DARABAD, 2008), teoria de sistema dinâmico (ALONSO *et al.*, 2005), as quais foram recentemente aplicadas a vários tipos de tarefas de classificação de voz patológica.

No processo de extração de característica baseada na FT, as características que são extraídas tem resolução tempo-frequência fixa por causa da limitação inerente à Transformada Discreta de Fourier (DFT), tornando difícil a classificação de amostras de voz usando essas características. Recentemente Transformada *Wavelet* Discreta (DWT) e Pacotes *Wavelet* (WP) (FONSECA; PEREIRA, 2009; PARRAGA, 2002) foram utilizados para extração de características devido às suas capacidades de multi-resolução. As características foram selecionadas a partir dos coeficientes *wavelet* de alta energia. Embora essas características oferecessem a vantagem de pegar alta freqüência de um sinal variando lentamente, sofriam com um problema de mudança da variância. Uma pequena mudança no sinal pode causar uma grande variação nos coeficientes *wavelet*, alterando assim as características extraídas.

Neste trabalho, é proposta uma técnica que é invariante à mudança de variância para sinais de alta frequência e, portanto, pode ser utilizada para extrair características de sinais de voz saudável e patológica baseada na DWT. Para avaliar o desempenho dessa técnica, comparou-se os resultados obtidos por diversos classificadores de padrões conhecidos na literatura.

1.2 Objetivos

O objetivo geral desta dissertação, bem como seus objetivos específicos, são detalhados a seguir.

1.2.1 Objetivo Geral

Esta dissertação tem por principal objetivo o desenvolvimento de um procedimento de extração de características dos sinais de fala para a diferenciação entre as vozes normal e patológica utilizando Transformada *Wavelet* Discreta (DWT).

A principal contribuição deste trabalho consiste em desenvolver uma abordagem utilizando DWT que permite classificar com alta probabilidade o tipo de patologia presente na laringe, possibilitando desenvolver um sistema de auxílio ao diagnóstico precoce e acompanhamento do tratamento das patologias estudadas.

1.2.2 Objetivos Específicos

O objetivo geral da dissertação apresentado anteriormente, por ser bastante amplo, dá margem ao surgimento de vários objetivos menores e mais específicos,

a saber:

i. Avaliação das características extraídas dos sinais de voz usando a Transformada *Wavelet* (WT);

ii. Proposição de um método modificado para extração de características de sinais de vozes patológicas e normais usando a Transformada *Wavelet* (WT) que seja invariante aos deslocamentos do sinal de voz;

iii. Avaliação do desempenho de classificadores neurais em relação aos demais encontrados na literatura para detecção de vozes patológicas;

iv. E desenvolvimento de um sistema de classificação baseado em redes neurais para vozes patológicas e normais.

1.3 Produção Científica

Ao longo do desenvolvimento desta dissertação os seguintes artigos científicos foram publicados:

1. **Raphael T. S. Carvalho**, Charles C. Cavalcante e Paulo C. Cortez (2011), "Wavelet Transform and Artificial Neural Networks Applied to Voice Disorders Identification", publicado no *Third World Congress on Nature and Biologically Inspired Computing* (NaBIC 2011), Salamanca-Espanha.

2. **Raphael T. S. Carvalho**, Charles C. Cavalcante e Paulo C. Cortez (2011), "Detecção de Doenças da Laringe usando Transformada Wavelet e Redes Neurais Artificiais", publicado no *X Congresso Brasileiro de Inteligência Computacional* (CBIC 2011), Fortaleza-CE.

1.4 Estrutura da Dissertação

O restante desta dissertação está organizada segundo os capítulos abaixo.

Capítulo 2 - Neste capítulo são apresentados os conceitos relacionados a fisiologia da voz e as principais características das patologias da laringe estudadas neste trabalho, destacando as principais causas e como a presença dessas patologias interfere no sinal de voz.

Capítulo 3 - 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. As características extraídas com essa ferramenta também são explicadas em maiores detalhes.

Capítulo 4 - Neste capítulo são discutindo os fundamentos de reconhecimento de padrões utilizados nesta dissertação, bem como uma proposta para o diagnóstico de patologias da laringe usando Transformada Wavelet. São discutidos os fundamentos da extração, seleção de características e a classificação de padrões, destacando os classificadores utilizando neste trabalho.

Capítulo 5 - Neste capítulo inicialmente são descritos o conjunto de dados utilizado e a metodologia das simulações. Em seguida são apresentados os resultados obtidos pelas simulações realizadas utilizando o *software* MATLAB.

Capítulo 6 - Por fim, são apresentadas as conclusões do estudo realizado nesta dissertação e as propostas para trabalhos futuros.

Capítulo 2

Fisiologia da Voz e Patologias

Neste Capítulo é dada uma noção básica sobre a fisiologia da voz, ou seja, o modo de produção da voz. Além disso, são descritas as principais características das patologias da laringe abordadas neste trabalho, destacando as principais causas para o aparecimento das mesmas e como a presença dessas interferem no sinal de voz.

2.1 Fisiologia da Voz

A fala é uma das capacidades ou aptidões que os seres humanos possuem de comunicação, manifestando seus pensamentos, opiniões e sentimentos através dos vocábulos. Consiste no principal sinal entre os distintos sinais abordados pela linguagem natural, como por exemplo, ideogramas, gestos, gritos, trejeitos e outros tipos de linguagem corporal (FANT, 1973).

Existem duas principais fontes de características da fala específicas aos locutores, as físicas e as adquiridas (ou aprendidas). As características físicas relacionam-se principalmente ao trato vocal, estrutura formada pelas cavidades que vão das pregas vocais até os lábios e o nariz (KENT, 1993). A Figura 2.1 ilustra o conjunto de órgãos que formam o trato vocal e compõem o sistema de produção da fala.

A produção da voz humana depende de uma conjunto de vários mecanismos, como a ressonância do trato vocal, as pressões subglotal e supraglotal, as características biomecânicas dos tecidos do trato vocal e a oscilação do trato vocal duração a fonação (SCALASSARA *et al.*, 2009). As pregas vocais vibram com a passagem de ar vindo dos pulmões. Essa vibração das pregas vocais produz um

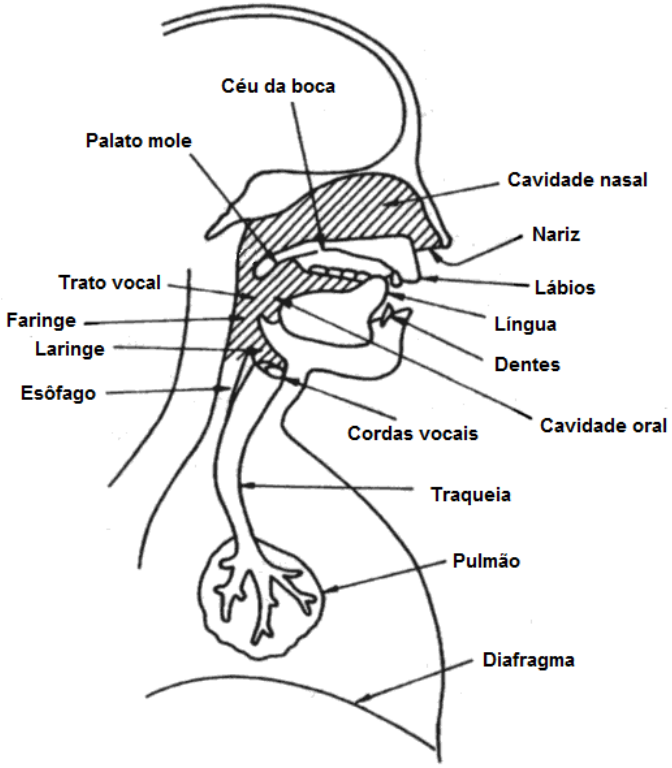


Figura 2.1: Sistema de produção da fala (KENT, 1993).

som fraco e constituído de poucos harmônicos, que é expandido quando passa pelas cavidades de ressonância (laringe, faringe, boca e nariz) e ganha "forma" final quando é articulado através de movimentos de língua, lábios, mandíbula, dentes e palato (BOONE; MCFARLANE, 2003).

A passagem do ar pelas cavidades do trato vocal altera o espectro do som devido às ressonâncias. Assim, o trato vocal funciona como um filtro gerando picos de amplitude no espectro de frequência conhecidos como formantes. Através da análise espectral da fala produzida é possível estimar a forma do trato vocal (KENT, 1993).

De acordo com Hirano (1988), a vibração do trato vocal determina a qualidade da voz. Em uma simples representação, o trato vocal consiste de cinco camadas de estrutura complexa: o epitélio, as camadas superficial, intermediária e profunda da Lâmina Propria (LP) da mucosa e o músculo vocal, conforme mostrado na Figura 2.2.

A camada superficial da LP consiste principalmente de uma substância amorfa muito maleável conhecido como Espaço de Reinke (RS). A camada intermediária da LP é composta principalmente de fibras elásticas, enquanto que a camada de fundo

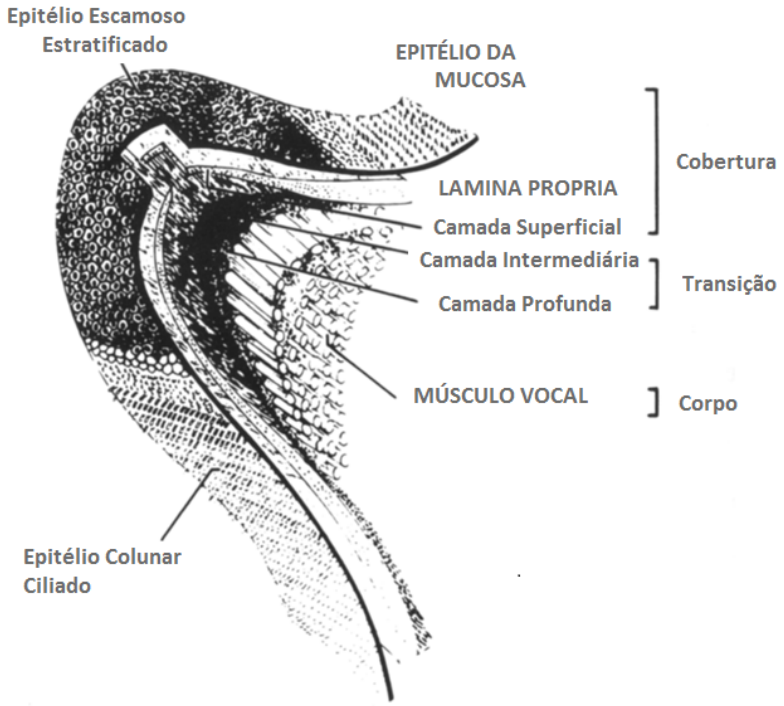


Figura 2.2: Representação do trato vocal, destacando as cinco camadas dessa estrutura (HIRANO, 1988).

é composta basicamente por fibras colágenas. A estrutura formada pelas camadas intermediária e profunda é chamada de ligamento vocal (HIRANO, 1988).

A estrutura das pregas vocais é frequentemente descrito pelo conceito de cobertura-corpo. Isto sugere que as pregas podem ser divididas em duas camadas de tecido, com diferentes propriedades mecânicas. A camada de cobertura é formada de tecido maleável, não-contrátil das mucosas, que serve como um invólucro em torno do corpo de camada. Em contraste, a camada de corpo é composto de fibras musculares (tireoaritenóideo) e algum tecido ligamentoso (STORY, 2002).

Assim, o movimento das pregas vocais pode ser observado como a sobreposição dessas duas componentes principais: a dinâmica do músculo vocal (corpo) e um dos epitélio e lâmina própria superficial (cobertura), conhecida como a onda mucosal. Este processo é descrito como uma onda que se propaga sobre o tecido de cobertura durante o ciclo de fonação, consistindo em um deslocamento dos tecidos em relação ao corpo (SCALASSARA *et al.*, 2009).

Na imagem do videolaringoscópio, Figura 2.3, a fase de fechamento deve ser observado pelo contato total entre as bordas das duas pregas vocais resultantes de

perturbações produzidas pelas ondas que viajam sobre as estruturas de cobertura respectiva, e produzindo o som da voz.

Figura 2.3: Imagem videolaringoscópica mostrando a fase de fechamento das pregas vocais. O contato pleno entre as duas fronteiras pregas vocais podem ser observados, resultando na produção de som da voz (SCALASSARA *et al.*, 2009).

Durante a respiração, as pregas vocais abertas permitem a passagem de ar através da glote, como mostrado na Figura 2.4. As imagens de vídeolaringoscópio foram coletadas no Departamento de Otorrinolaringologia e no Departamento de Cirurgia de Cabeça e Pescoço da Faculdade de Medicina de Ribeirão Preto, Estado de São Paulo, Brasil (SCALASSARA *et al.*, 2009).

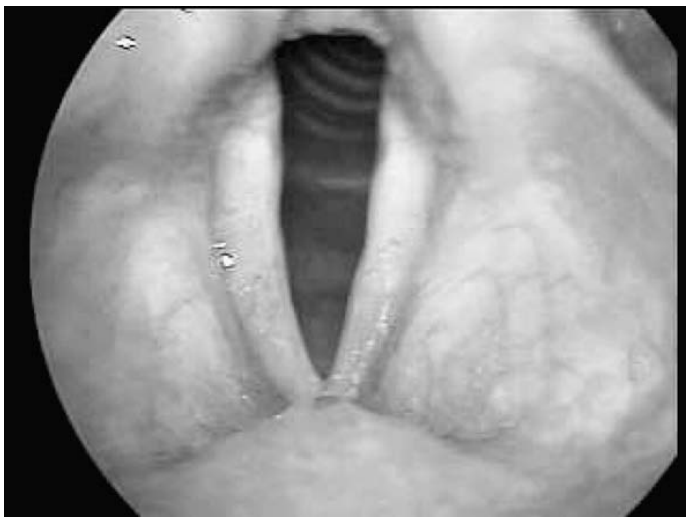


Figura 2.4: Imagem videolaringoscópica mostrando as pregas vocais abertas durante a respiração, permitindo a passagem de ar através da glote.(SCALASSARA *et al.*, 2009).

2.2 Patologias

Nesta Seção, são descritas as patologias abordadas neste trabalho, suas principais características, as principais causas e como elas interferem na produção do sinal de voz.

2.2.1 Nódulos Vocais

Os nódulos vocais são as lesões benignas mais recorrente nos pacientes das clínicas de voz, tanto em crianças quanto adultos. O uso excessivo da voz e os hábitos vocais inadequados são os fatores determinantes no desenvolvimento de nódulos nas pregas vocais, sendo frequentemente visto em pessoas que utilizam a voz profissionalmente, como professores, palestrantes e cantores, bem como em crianças. Outros fatores predisponentes têm sido destaque na gênese dos nódulos nas pregas vocais, como obstrução nasal, rinossinusite recorrente, insuficiência velofaríngea, hipoacusia e refluxo gastroesofágico (MARTINS *et al.*, 2010). Os principais sintomas da presença de nódulos são rouquidão, falta de ar, fadiga vocal, desconforto na garganta e redução na extensão vocal durante o dia.

O nódulo vocal, geralmente, ocorre em ambos os lados da prega vocal, estritamente simétrica na fronteira do terço anterior e médio da prega vocal, conforme mostrando na Figura 2.5, sendo geralmente imóvel durante a fonação. A lesão está confinada à camada superficial da lâmina própria e é o resultado da colisão traumática e constante das pregas vocais, causados pela sobrecontração dos músculos da laringe intrínseca durante a fonação. De acordo com diversos estudos, a formação dos nódulos ocorre principalmente no ponto médio da prega vocal membranosa, onde as forças de impacto são os maiores e estes são geralmente bilateral (SCALASSARA *et al.*, 2009).

Durante a fase final da vibração das pregas, a presença de nódulos na camada exterior do tecido das pregas vocais inibe-as de serem completamente dobradas umas sobre as outras, conforme a Figura 2.6. Consequentemente, o fechamento da glote é inacabado, acrescentando ar turbulento ao sinal de voz. A fim de reduzir este efeito, os indivíduos aumentam a tensão muscular e a pressão subglótica, aumentando as forças de colisão vocal (HILLMAN *et al.*, 1990).

O tratamento do nódulo vocal depende do tamanho, da forma e do tecido do nódulo. Os nódulos moles e recentes podem desaparecer com o repouso vocal, que

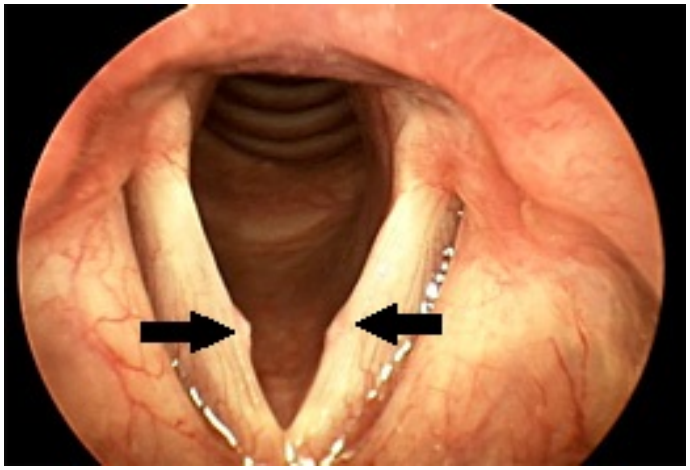


Figura 2.5: Imagem videolaringoscópica mostrando a presença de nódulos nas pregas vocais, indicados pelas setas (SULICA, 2009).

serve para amolecer e dissolver inchaço associado com fonotrauma. Os fibrosos e mais antigos costumam ser retirados com a cirurgia, sendo o paciente indicado a um acompanhamento fonoaudiológico em muitos casos até mesmo antes da cirurgia.

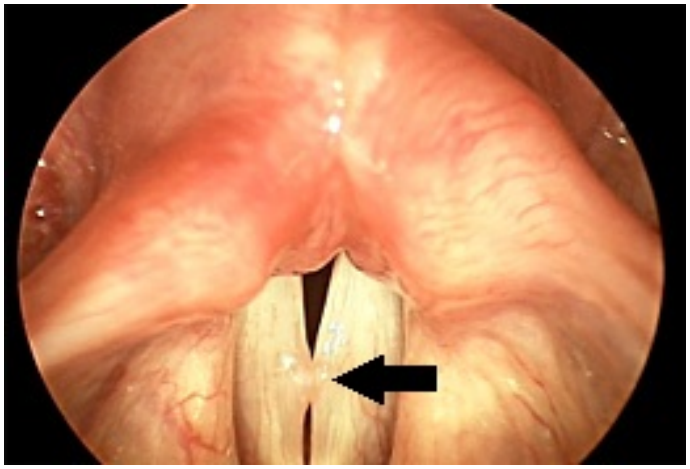


Figura 2.6: Imagem videolaringoscópica mostrando o fechamento das pregas vocais na presença de nódulos, indicados pela seta (SULICA, 2009).

O tratamento com o fonoaudiólogo é conhecido como fonoterapia e tem como objetivo promover a reabsorção dos nódulos corrigindo o desvio funcional e tornar o paciente mais consciente das circunstâncias e hábitos de uso da voz que levaram ao problema, encontrando estratégias de uso da voz que será menos problemático. Quando a terapia vocal é em crianças, é importante salientar que a orientação e o auxílio dos pais é determinante para o tratamento, além de conscientizar a criança do abuso vocal e identificar a causa do problema.

O sucesso do tratamento com a reabsorção dos nódulos vai depender de diversos fatores que vão desde a qualidade do profissional e dos exercícios propostos até a dedicação do paciente. O procedimento cirúrgico é raramente utilizado, pois existe uma grande probabilidade de reabsorção durante a fonoterapia, sendo os nódulos do tipo cônicos pontiagudos os que encontram maior resistência a absorção. Nos casos em que se opta pelo tratamento cirúrgico, a fonoterapia deve ser ministrada no pós-operatório, a fim de modificar os ajustes laríngeos inadequados e trabalhar as questões comportamentais.

2.2.2 Edema de Reinke

O edema de Reinke é uma doença da laringe na qual ocorre um inchaço generalizado das pregas vocais devido ao acúmulo de líquido na camada superficial da lâmina própria (NEVES; NETO; PONTES, 2004). Recebe este nome por se localizar no espaço anatômico com o nome em homenagem ao anatomista alemão Friedrich Reinke que foi o primeiro a investigar a anatomia das pregas vocais. Ele descreveu a frouxa camada subepitelial das pregas vocais, que é limitada acima pela linha arqueada superior e abaixo pela linha arqueada inferior na junção do epitélio cilíndrico com o escamoso (KLEINSASSER, 1997).

É uma doença que pode ocorrer em ambas as pregas vocais ou ser limitada a uma prega vocal, geralmente no início da doença. Nessa doença, o líquido acumulado no forro da submucosa do Espaço de Reinke (RS) faz com que a cobertura das pregas vocais fique menos rígida e mais maciça, fazendo com que a prega vocal aumente sua espessura e se projete para o interior da laringe como uma estrutura de vibração, conforme a Figura 2.7, geralmente evoluindo para uma irritação crônica das pregas vocais, modificando a permeabilidade capilar dos tecidos (HIRANO, 1981).

Essa doença está frequentemente associada com fatores etiopatogênicos, como tabagismo, sinusite ou infecção do trato respiratório superior, e com o uso excessivo da voz, também conhecido com fonotrauma (ABREU, 1999). O refluxo gastroesofágico (RGE) ou irritação persistente também é considerado um fator que contribui para o edema de Reinke (KLEINSASSER, 1997). Entretanto, conforme citado por Greene (1989) e Scalassara *et al.* (2009), ainda não há evidências se as condições alérgicas ou medicamentos também sejam fatores que contribuem para esta doença.

O edema de Reike é uma lesão laríngea benigna e não existe comprovação de que

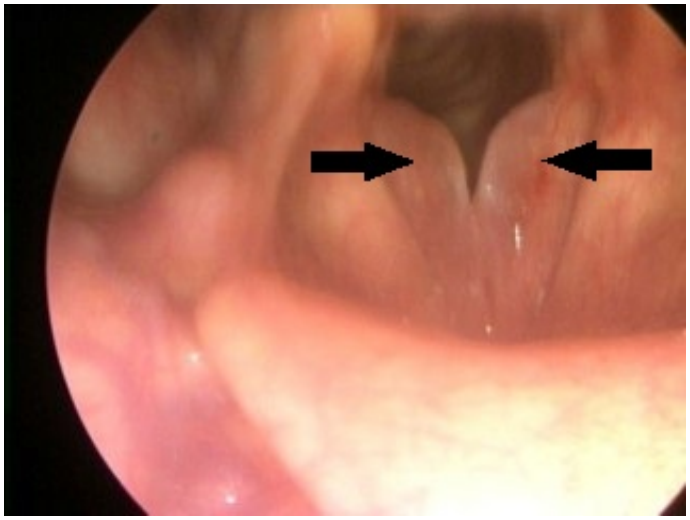


Figura 2.7: Imagem videolaringoscópica mostrando as pregas vocais com edema de Reinke, indicado pelas setas (SULICA, 2009).

tenha potencial para transformar-se em câncer de laringe. A relação com o câncer, em alguns trabalhos científicos, deve-se ao fato dessa doença ocorrer geralmente em pacientes fumantes e o tabagismo, este sim, é um dos principais fatores etiológicos dos tumores desta região.

De um modo geral, o edema de Reinke ocorre gradualmente ao longo do tempo e inicialmente o paciente só percebe a voz em tom baixo, sendo a rouquidão um dos principais sintomas. Essa rouquidão é considerada até mesmo agradável no início. Com o tempo, o problema se agrava e a voz se torna grave, com o paciente podendo notar um aumento do esforço com a fala e dificuldade para respirar, conforme mostrado na Figura 2.8. Devido a esta característica gradual, os pacientes geralmente procuram um especialista somente após mudanças significativas na qualidade da voz.

Os pacientes com edema de Reinke apresentam algumas características alteradas, como a frequência fundamental e a intensidade da voz, qualidade vocal ruim, fadiga vocal devido tensão músculo-esquelética excessiva e relação fonte/filtro alterada (ABREU, 1999).

O diagnóstico é feito pelo otorrinolaringologista, analisando os sintomas relatados pelo paciente e correlacionando com os encontrados na videolaringoscopia e que são bastante característicos. Este exame é realizado no próprio consultório, sem necessidade de sedação ou anestesia. De acordo com Courey *et al.* (1995), o edema de Reinke em avaliação estroboscópica apresenta movimentos da onda mucosa maior

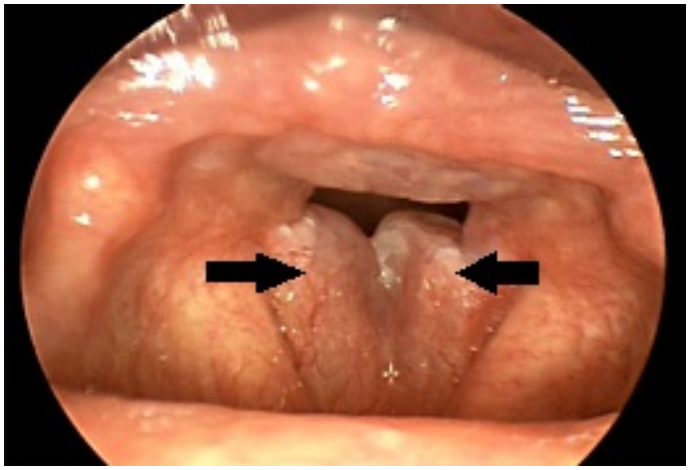


Figura 2.8: Imagem videolaringoscópica mostrando as pregas vocais com edema de Reinke grave e que pode causar dificuldade para respirar. Nesta foto, as pregas vocais estão abertas (SULICA, 2009).

que o normal e grande abertura posterior da glote. Hirano e Bless (1997) descrevem que a glote fica completamente fechada durante a vibração e que os movimentos das pregas vocais bilaterais são assimétricos e as vibrações sucessivas são aperiódicas. A amplitude da excursão horizontal é com freqüência pequena, mas a onda mucosa é em geral acentuadamente grande.

O tratamento do edema de Reinke depende da fase em se encontra a doença no momento do diagnóstico. Inicialmente, a medida é identificar e remover o fator irritativo, como o tabagismo, o abuso vocal ou o refluxo gastroesofágico. Nas fases iniciais, quando o edema é pequeno, pode-se tentar o tratamento conservador através da fonoterapia, que consiste em exercícios com fonoaudiólogo que estimulam a absorção do excesso de líquido pelo organismo. O objetivo, nesse último caso, é controlar a doença, ou seja, interromper sua progressão, já que ainda não existe um tratamento clínico que elimine totalmente o edema (ABREU, 1999).

Nos estágios mais avançados, o tratamento é essencialmente cirúrgico, através de microcirurgia da laringe, sendo mais recomendado principalmente quando o edema é muito grande, o que pode ocasionar o fechamento da laringe, ou quando o paciente é do sexo feminino e a voz está muito grave e desagradável, a ponto de interferir em sua autoestima. A cirurgia também é recomendada em casos de suspeita de lesão pré-maligna ou de câncer. O procedimento cirúrgico é realizado em hospital com anestesia geral. Normalmente não é necessária a internação hospitalar, permanecendo o paciente na sala de recuperação até ter condições de

alta e retornando a sua residência no mesmo dia.

Existem várias técnicas para realizar esta cirurgia, sendo que atualmente todas visam remover o edema com preservação da cobertura mucosa da prega vocal, visando manter a qualidade da voz o mais próximo possível da normalidade. O procedimento é realizado com microscópio cirúrgico e com instrumental próprio para microcirurgia de laringe, e consiste em realizar uma incisão e raspagem das pregas vocais e aspirar o fluido. Então, tanto se pode cauterizar o corte com laser como só repor a mucosa sobre o local que a região se regenera.

Após a cirurgia, pacientes com edema de Reinke bilateral acentuado permanecem afônicos por várias semanas e iniciam fonação com alguma dificuldade, necessitando de reeducação vocal com fonoaudiólogo. Em pacientes fumantes, o fato de voltar a fumar após a cirurgia tem grandes possibilidades de provocar a reincidência do problema. Portanto, para um melhor resultado, é imprescindível abandonar o hábito do fumo.

2.2.3 Disfonia Neurológica

Uma disfonia representa qualquer dificuldade para emissão vocal que impeça a produção natural da voz. Quando essa dificuldade tem origem neurológica, a disfonia é conhecida como Disfonia Neurológica.

Diversas doenças com origem neurológica afetam diretamente a produção da voz, sendo as mais conhecidas: Acidente Vascular Cerebral (AVC), Doença de Huntington, Doenças de Parkinson, Esclerose Lateral Amiotrófica (ELA), Mononeurite múltipla, Mitocondropatia, Distrofia de Duchenne, Distrofia Miotônica e Distonia Cervical.

Dependendo da localização da lesão neurológica têm-se manifestações diversas na fala, voz e linguagem do indivíduo. Os distúrbios neurológicos da fala podem ser definidos de acordo com o nível anatômico afetado: transtornos do neurônio motor superior, transtornos do neurônio motor inferior, transtornos do sistema extrapiramidal, transtornos do cerebelo, transtornos da junção neuromuscular e transtornos mistos (disfonia espástica) (ORTIZ; CARRILLO, 2008).

2.3 Resumo do Capítulo

Neste capítulo foram descritas a fisiologia de produção da voz e as principais características das patologias da laringe estudadas neste trabalho. Baseado nas diferenças fisiológicas entre as vozes saudáveis e patológicas, é necessária uma ferramenta matemática que permita extrair características suficientes para diferenciar e classificar as vozes dos pacientes. No próximo capítulo, é descrito a Transformada *Wavelet* (WT), suas propriedades e as características que serão extraídas do sinal de voz para classificar as patologias da voz descrita neste capítulo.

Capítulo 3

Transformada Wavelet

A Transformada *Wavelet* (WT) é uma ferramenta matemática de análise espaço-frequência que tem sido intensamente estudada para aplicações de processamento de sinais da fala durante as últimas décadas, principalmente devido às limitações da Transformada de Fourier (FT). Na análise por FT (OPPENHEIM; WILLSKY; NAWAB, 1996) assume-se que o sinal é estacionário no tempo, ou seja, que as suas propriedades estatísticas não variam em função do tempo.

Assim para utilizar a FT no processamento de sinais não-estacionários como os da fala, é necessário utilizar uma versão janelada do sinal com a suposição de estacionaridade durante esta janela. Essa versão modificada da FT é conhecido como Transformada de Fourier de Tempo Curto (STFT). Neste método, escolhendo-se uma janela de curta duração de tempo, a resolução em frequência também será pequena. Se aumentar a duração da janela, esta também aumentará. Ou seja, fixando-se o tamanho da janela, a resolução tempo-freqüência conseguida pela STFT é também fixada (FAROOQ; DATTA, 2003).

Para solucionar a duração fixa da janela, a análise por Transformada *Wavelet* (WT) utiliza uma janela de tamanho adaptativo, permitindo selecionar mais tempo para baixas freqüências e menos tempo para altas freqüências (RIOUL; VETTERLI, 1991). Essa análise pode ser usada para um sinal que tem componentes de alta frequência de curta duração e componentes de baixa frequência de longa duração, como no caso da fala (FAROOQ; DATTA, 2003).

A WT expande um sinal dentro de um conjunto completo de funções de base (geralmente é utilizado um conjunto de base ortogonal). Diferentemente das funções

de Fourier, as *wavelets* fornecem uma representação tempo-freqüência de forma simultânea, o que é de grande auxílio, pois em muitos casos é de interesse conhecer a ocorrência de um componente espectral num determinado instante, sendo este capaz de revelar aspectos importantes como limites, pontos de inflexão, descontinuidades e similaridade (COSTA, 2006).

Muitos dos avanços obtidos nos estudos utilizando WT foram desenvolvidos devido à cooperação de Ingrid Daubechies e Stephane Mallat. Daubechies (1992) desenvolveu uma família de *wavelets* com base compacta (*compact support*) e Mallat (1989) introduziu a WT no conceito de decomposição multirresolução de sinais, além da implementação da transformada rápida baseada em conceitos de filtragem.

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*.

3.1 Decomposição *Wavelet*

A WT consiste na decomposição de um sinal $x(t)$ através de uma família de bases, geralmente, reais e ortonormais. A função base usada na WT é localizada tanto no tempo como na freqüência. Todas as funções *wavelet* são versões geradas por dilatações e translações de uma função protótipo $\psi(t)$, também conhecida como *wavelet* 'mãe', de média zero e centrada na vizinhança de $t = 0$, dada por (DAUBECHIES, 1992):

$$\psi_{\tau,a}(t) = |a|^{-\frac{1}{2}} \cdot \psi\left(\frac{t-\tau}{a}\right), \tag{3.1}$$

em que os parâmetros τ e a são chamados parâmetros de translação e escalonamento respectivamente. O termo $|a|^{-1/2}$ é usado para normalização da energia, ou seja, para que $\|\psi_{\tau,a}(t)\| = \|\psi(t)\|$ para todo τ e a, assumindo que $\|\psi(t)\| = 1$, em que $\|\cdot\|$ é operador norma.

Para ser considerada uma *wavelet*, uma função também tem de atender as seguintes propriedades (DAUBECHIES, 1992):

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 $\int_{-\infty}^{\infty} |\psi(t)|^2 dt$ é finita.

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.

A Transformada *Wavelet* Contínua (CWT) de um sinal $x(t)$, em que $x \in \mathbf{L^2}$, é definida como a correlação entre a função $x(t)$ e a família wavelet $\psi_{\tau,a}(t)$ para cada τ e a, dada por (RIOUL; VETTERLI, 1991):

$$CWT(\tau, a) = a^{-\frac{1}{2}} \int x(t) \cdot \psi^* \left(\frac{t-\tau}{a} \right) dt, \tag{3.2}$$

em que o parâmetro de escalonamento a fornece a largura da *wavelet*, τ indica a posição e $\psi^*(t)$ é o complexo conjugado de $\psi(t)$. Tipicamente, a CWT é sobre-completa e uma amostragem apropriada pode ser usada para eliminar redundâncias (FAROOQ; DATTA, 2003).

A versão discreta da transformada podem ser obtida discretizando as dilatações e as translações. Neste caso, as funções *wavelets* para a Transformada *Wavelet* Discreta (DWT) podem ser ser representadas pela função *wavelet* mãe $\psi(t)$ com um conjunto discreto de parâmetros, $a_j = 2^j$ e $\tau_{j,k} = 2^j k$, com $j, k \in \mathbb{Z}$ (conjunto dos inteiros) dada por:

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

Essa família de funções constitue uma base ortonormal do Espaço de Hilbert $\mathbf{L^2}$ consistindo de sinais de energia finita. Para construir a *wavelet* mãe $\psi(t)$, é preciso determinar a função de escalonamento $\phi(t)$, que satisfaz a seguinte equação:

$$\phi(t) = \sqrt{2} \sum_k h_\phi(k) \phi(2t - k). \tag{3.4}$$

A função mãe $\psi(t)$ está relacionada com a função de escalonamento $\phi(t)$ através da seguinte equação:

$$\psi(t) = \sqrt{2} \sum_{k} h_{\psi}(k) \phi(2t - k), \tag{3.5}$$

em que

$$h_{\psi}(k) = (-1)^{k} h_{\phi}(1 - k). \tag{3.6}$$

Para formar um conjunto de funções de base, os coeficientes $h(k)$, também chamados de coeficientes da função *wavelet* (GONZALEZ; WOODS, 2008), precisam obedecer às seguintes condições: serem únicos, ortonormais e possuirem um certo grau de regularidade. Os coeficientes $h_{\psi}(t)$ (passa-baixa) e $h_{\phi}(k)$ (passa-alta) são de fundamental importância para a Transformada *Wavelet* Discreta (DWT).

O conhecimento dos coeficientes *wavelets* possibilita a utilização desta transformada sem necessariamente explicitar as formas reais das funções de escalonamento $\phi(t)$ e wavelet mãe $\psi(t)$ (COSTA, 2006). A decomposição *wavelet* pode ser definida como

$$x(t) = 2^{-N/2} \sum_{k} A(j_0, k) \phi_{j_0,k}(t) + 2^{-N/2} \sum_{j=j_0}^{\infty} \sum_{k} D(j, k) \psi_{j,k}(t), \tag{3.7}$$

em que $A(j, k)$ e $D(j, k)$ são conhecidos como coeficientes de aproximação (baixa frequência) e detalhes (alta frequência), respectivamente, e são definidos pelas equações

$$A(j_0, k) = 2^{-N/2} \sum_{i} x(i) \cdot \phi^{*} \left(2^{-j_0} i - k\right), \tag{3.8}$$

e

$$D(j, k) = 2^{-N/2} \sum_{i} x(i) \cdot \psi^{*} \left(2^{-j} i - k\right), \tag{3.9}$$

em que j_0 é a escala inicial, j é a escala atual da decomposição ($j = 0, 2, \ldots, N-1$), N é o número de escalas de decomposição e i, j e k são inteiros.

A DWT fornece uma representação não redundante do sinal e seus valores constituem os coeficientes de uma série *wavelet*. Estes 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 $\mathbf{L^2}$ (ROSSO *et al.*, 2001).

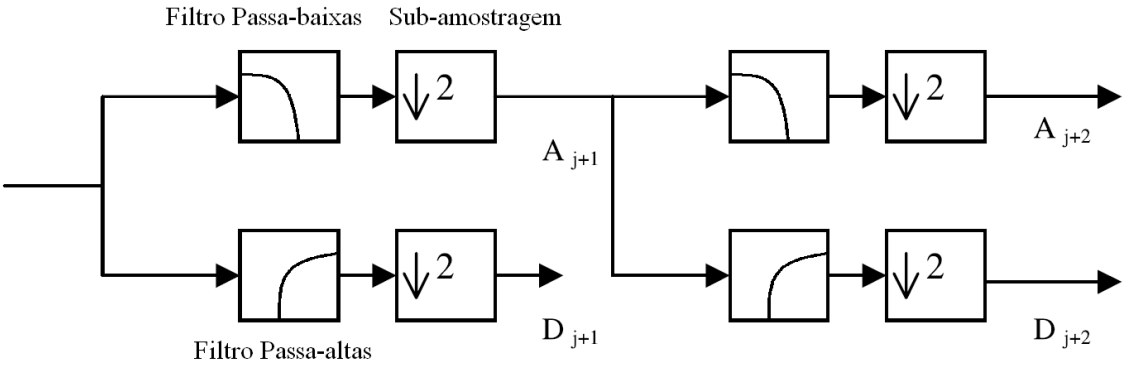


Figura 3.1: Decomposição de sinal usando DWT diádica (FAROOQ; DATTA, 2003).

A DWT também pode ser vista como um processo de filtragem do sinal, usando um filtro passa-baixas (função de escalonamento $\phi(t)$) e um filtro passa-altas (função *wavelet* mãe $\psi(t)$). Então, o primeiro nível de decomposição DWT de um sinal divide em duas faixas, uma versão passa-baixas e uma versão passa-altas do sinal. A versão passa-baixas fornece a representação aproximada do sinal, enquanto a passa-altas indica os detalhes ou variações de altas freqüências. O segundo nível de decomposição é executado sobre a versão passa-baixas do primeiro nível de decomposição, como mostrado na Figura 3.1. Então, a decomposição *wavelet* resulta em uma árvore cuja estrutura é dita recursiva (FAROOQ; DATTA, 2003).

3.2 Características Extraídas

Nesta seção, são descritas algumas características extraídas a partir dos coeficientes obtidos pela decomposição *wavelet* de um determinado sinal.

3.2.1 Energia *Wavelet*

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 (DWT). Então, dado um processo estocástico x(t), seu sinal associado é assumido ser dado pelos valores amostrados $\chi = \{x(n), n = 1, \ldots, M\}$. Os coeficientes *wavelet* obtidos da decomposição *wavelet* são dados por

$$D(j,k) = \langle \chi, 2^{j/2}\phi(2^j t - k) \rangle \tag{3.10}$$

com $j = 1, 2, \ldots, N$ 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 $2^j k$. 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 (ZUNINO *et al.*, 2006).

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

$$E_j = \sum_k |D(j,k)|^2 \tag{3.11}$$

e a energia de cada amostra de tempo k é

$$E(k) = \sum_{j=1}^{N} |D(j,k)|^2. \tag{3.12}$$

Consequentemente, a energia total do sinal pode ser obtida através da seguinte equação:

$$E_{total} = \sum_{j=1}^{N} \sum_k |D(j,k)|^2 = \sum_{j=1}^{N} E_j. \tag{3.13}$$

Normalizando os valores de energia em cada nível de decomposição pela energia total, obtém-se a energia *wavelet* relativa dada por

$$p_j = \frac{E_j}{E_{total}}, \tag{3.14}$$

para os níveis de decomposição $j = 1, 2, \ldots, N$. Os valores de energia *wavelet* relativa definem uma distribuição de probabilidade da energia, pois $\sum_j p_j = 1$. A distribuição p_j pode ser considerada como uma densidade na escala do tempo. Isto dá uma ferramenta adequada para detectar e caracterizar fenômenos específicos nos domínios do tempo e da freqüência.

3.2.2 Entropia *Wavelet*

Outra característica a ser extraída dos coeficientes da decomposição *wavelet* é a entropia *wavelet*. A entropia de Shannon (SHANNON, 1948) 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* total é definida como

$$S_{WT}(p) = \sum_{j=1}^{N} p_j \cdot \ln [p_j], \tag{3.15}$$

em que p_j é a distribuição de probabilidade para os níveis de decomposição $j = 1, 2, \ldots, N$ e ln é a operação matemática de logaritmo natural.

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. Na verdade, um processo muito ordenado pode ser pensado como um sinal mono-freqüência periódica (sinal com um espectro de banda estreita). Uma representação *wavelet* de tal sinal será muito resolvido em um único nível de decomposição *wavelet*, ou seja, todas as energias *wavelet* relativa serão quase zero, exceto para o nível de decomposição *wavelet* que inclui a frequência do sinal representativo. Para este nível especial a energia *wavelet* relativa será quase 1 e em consequência a entropia *wavelet* estará próxima de zero ou um valor muito baixo.

Um sinal gerado por um processo totalmente aleatório pode ser tomado como representando um comportamento muito desordenado. Este tipo de sinal terá uma representação *wavelet* com contribuições significativas de todas as bandas de freqüência. Além disso, pode-se esperar que todas as contribuições serão da mesma ordem. Consequentemente, a energia *wavelet* relativa será quase igual para todos os níveis de resolução e a entropia *wavelet* terá seus valores máximos.

3.2.3 Entropia *Wavelet* Relativa

Supondo que se tenham duas diferentes distribuições de probabilidade $\{p_j\}$ e $\{q_j\}$, com $\sum_j p_j = \sum_j q_j = 1$. Neste caso, estas podem ser vistas como distribuições de probabilidade da energia *wavelet* para dois segmentos de um sinal ou para dois diferentes sinais. Definindo a Entropia *Wavelet* Relativa (RWE), formalmente uma divergência de Kullback-Leibler, como:

$$S_{WT}(p|q) = \sum_{j=1}^{N} p_j \cdot \ln \left[\frac{p_j}{q_j}\right], \tag{3.16}$$

em que $q_j \neq 0$.

Esta característica fornece uma medida do grau de similaridade entre duas distribuições de probabilidade (mais precisamente da distribuição $\{p_j\}$ com respeito à distribuição $\{q_j\}$ tomada como uma distribuição de referência) (ROSSO *et al.*, 2001). Note que a RWE é positiva e desaparece somente se $p_j = q_j$ (COVER; THOMAS, 2006).

3.3 Resumo do Capítulo

Neste Capítulo foram descritos os conceitos básicos da Transformada *Wavelet* (WT) e as principais características extraídas utilizando os coeficientes da decomposição *wavelet*. De acordo com o exposto neste capítulo, a Transformada *Wavelet* Discreta (DWT) pode ser utilizada para extrair características dos sinais de vozes, permitindo classificar as amostras de voz em patológicas e saudáveis. No próximo capítulo, são descritos os classificadores para reconhecimento de padrões utilizados neste trabalho para classificar essas amostras de voz a partir destas características extraídas com a DWT.

Capítulo 4

Reconhecimento de Padrões: Fundamentos e Proposta

O Reconhecimento de Padrões (RP) trata da classificação de uma estrutura de dados através de um conjunto de propriedades ou características. O RP envolve técnicas para a atribuição dos padrões a suas respectivas classes, de forma automática ou com a menor intervenção humana possível. Um padrão é uma descrição de um objeto e a classe de padrões é uma família de objetos que compartilham uma mesma propriedade.

Exemplos de aplicações de RP são o reconhecimento de voz e impressão digital, identificação de caracteres, estrutura de íris, reconhecimento de palavras e escrita cursiva, reconhecimento de formas, supervisão de processos, detecção de falha em máquinas e diagnósticos médicos.

O reconhecimento de padrões é composto pelas etapas de extração de características e classificação de padrões. Na etapa de extração de características os dados de entrada são representados em termos de medidas ou informações que possam ser utilizados facilmente na etapa de classificação. Por fim, na etapa de classificação, os padrões são classificados em função das características em comum entre estes.

Neste Capítulo são descritos os fundamentos do reconhecimento de padrões utilizados nesta dissertação, bem como uma proposta para o diagnóstico de patologias da laringe usando Transformada Wavelet. São discutidos os fundamentos da extração, seleção de características e a classificação de padrões.

4.1 Extração de Características

A extração de características é uma forma especial de redução dimensional. Quando a quantidade de dados de entrada para um algoritmo é considerada grande para o processamento ou existe informação notavelmente redundante (muitos dados e pouca informação), os dados deverão ser transformados em um conjunto reduzido de características e que seja mais representativo.

Assim, na extração de características, o espaço de dados de entrada é transformado num espaço de características que possui, geralmente, uma dimensão menor que a do espaço de dados original, ou seja, é representado por um número reduzido de características efetivas.

Na modelagem, as características extraídas são agrupadas em um vetor conhecido como padrão $\mathbf{x}_i$, representado por

$$\mathbf{x}_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}, \tag{4.1}$$

em que cada componente x_j representa a característica j extraída num conjunto total de n características. Este vetor também é definido através da notação $\mathbf{x}_i = (x_1; x_2; \ldots; x_n)^T$, em que ^{T} indica a operação de transposição de matrizes ou vetores.

O espaço de características é definido na forma de uma matriz $\mathbf{X}$, com n linhas (características) e p colunas (padrões), em função dos vetores $\mathbf{x}_i$ através da notação $\mathbf{X} = (\mathbf{x}_1; \mathbf{x}_2; \ldots; \mathbf{x}_p)$, que será utilizada nesta dissertação. O espaço de característica $\mathbf{X}$ é também representado pela matriz

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}. \tag{4.2}$$

Uma vez extraída as características, o problema de RP consiste em obter uma função discriminante que permita separar as diferentes classes presentes no espaço de características.

4.2 Dimensionalidade do Espaço de Características

O termo dimensionalidade é atribuído ao número de características de uma representação de padrões, ou seja, à dimensão do espaço de características. Essa está diretamente relacionada à separação e organização dos padrões em classes.

A adição de uma nova característica ao espaço fornece uma nova informação a ser usada pela classificação. A contínua adição de características não implica necessariamente na melhoria da classificação, pois isso gera um aumento da dimensionalidade e torna a classificação mais difícil (DUDA; HART; STORK, 2001). Além disso, quando o conjunto de dados não é muito grande, um espaço de características reduzido pode geralmente resultar em um sobre-dimensionamento das amostras de treinamento, gerando resultados indesejados para novos dados (MARQUES, 2005).

Assim, é necessário e mais eficiente selecionar as características que melhor discriminam as classes do problema estudado, ao invés de utilizar muitas características para realizar a classificação. Essa seleção pode ocorrer com base no desempenho do classificador ou na obtenção de outro critério de separabilidade.

A redução de informações redundantes ou irrelevantes dos dados pode contribuir não somente para a redução do esforço computacional dos processos de extração de características e classificação (TANG *et al.*, 2005), mas, em alguns casos, também pode melhorar a precisão do classificador com essa redução (JAIN; ZONGKER, 1997).

Os principais métodos para redução do espaço de características são:

- **combinação de características -** é o método mais geral e consiste em obter

um novo espaço de características a partir da combinação ou transformação das características existentes. Os métodos existentes são divididos em lineares e não lineares. No método linear, o novo espaço é gerado a partir de combinação linear das variáveis de entradas. Exemplos clássicos desses métodos são Análise por Componentes Principais (PCA), Análise por Discriminante Linear (LDA) (TANG *et al.*, 2005) e Análise por Variáveis Canônicas (CVA) (HOTELLING, 1936). Exemplos de métodos não lineares são as redes neurais artificais (HAYKIN, 2001);

- **seleção de características -** é o processo de escolher um subconjunto do conjunto de características originais que melhor represente o problema estudado.

4.3 Classificação de Padrões

A classificação de padrões pode ser definida como a determinação de uma fronteira de decisão que consegue distinguir diferentes padrões em classes dentro de um espaço de características d-dimensional, em que d é o número de características. Então, a classificação pode ser realizada e pode ser entendida, de maneira geral, pela partição do espaço de atributos em um número finito de regiões de tal forma que objetos de uma mesma classe recaiam com maior frequência dentro de uma mesma região.

Os problemas de classificação podem ser categorizados em dois tipos (DUDA; HART; STORK, 2001):

- **classificação supervisionada -** um ou mais padrões de exemplos, conhecidos como conjunto de treinamento, de classes de objetos conhecidos a priori, são fornecidos como referência para classificação de padrões desconhecidos. Esse tipo de classificação é dividido em dois estágios: (i) aprendizado, correspondendo à etapa em que os parâmetros dos classificadores e os critérios de seleção são formulados a partir dos protótipos e (ii) reconhecimento, onde o sistema treinado é usado para classificar novos padrões desconhecidos;

- **classificação não-supervisionada -** não há necessidade de um conhecimento prévio dos rótulos ou classes das amostras para identificação de um padrão, geralmente utilizam-se técnicas de agregação de dados. Essas técnicas

baseiam-se frequentemente na minimização de um critério derivado de uma medida de similaridade entre as amostras.

Em meio aos dois tipos de classificadores, outros classificadores assumem que os dados podem ser modelados por distribuição estatística. Os classificadores paramétricos estimam os parâmetros das distribuições envolvidas, enquanto que os classificadores não paramétricos realizam a classificação com base em medidas não paramétricas das fontes de dados.

A seguir são descritos os fundamentos dos classificadores para RP utilizados neste trabalho para classificar as doenças da laringe.

4.3.1 *Naive* Bayes

O classificador é denominado ingênuo (*naive*) por assumir que os atributos são condicionalmente independentes, ou seja, a informação de um evento não é informativa sobre nenhum outro. A classificação bayesiana é um exemplo de abordagem supervisionada que utiliza os conceitos da teoria de decisão estatística para estabelecer fronteiras de decisões entre classes de padrões (RUSSELL *et al.*, 1996).

Seja $\Omega = \{w_1, \ldots, w_c\}$ um conjunto de c classes e $\mathbf{x}$ um vetor de atributos com n características (em que cada uma delas é uma variável aleatória), $p(\mathbf{x}|w_j)$ a função de densidade de probabilidade condicional do padrão (função de verossimilhança da classe w_k) $\mathbf{x}$ dada a classe w_j e $P(w_j)$ a probabilidade de ocorrência *a priori* da classe w_j.

A regra de Bayes permite estabelecer a probabilidade *a posteriori* $P(w_j|\mathbf{x})$ de ocorrer a classificação na classe w_j dado o vetor $\mathbf{x}$ em função da probabilidade *a priori* $P(w_j)$:

$$P(w_j|\mathbf{x}) = \frac{p(\mathbf{x}|w_j) \cdot P(w_j)}{p(\mathbf{x})}, \tag{4.3}$$

em que a função de densidade de probabilidade do vetor aleatório $\mathbf{x}$ é:

$$p(\mathbf{x}) = \sum_{j=1}^{c} p(\mathbf{x}|w_j) \cdot P(w_j). \tag{4.4}$$

A regra de decisão de Bayes afirma que, dado $\mathbf{x}$, devemos atribuí-lo à classe w_j, de tal forma que $P(w_j|\mathbf{x})$ seja máxima.

Sendo $p(\mathbf{x})$ um fator de normalização na equação 4.3, e denotando $\widehat{w}$ a regra de decisão, para o caso particular de duas classes apenas, podemos escrever a regra de decisão de Bayes como:

$$\widehat{w}(\mathbf{x}) = \begin{cases} w_1, & \text{se } p(\mathbf{x}|w_1) \cdot P(w_1) > p(\mathbf{x}|w_2) \cdot P(w_2), \\ w_2, & \text{se } p(\mathbf{x}|w_1) \cdot P(w_1) < p(\mathbf{x}|w_2) \cdot P(w_2). \end{cases}$$

4.3.2 Vizinho mais Próximo

O exemplo mais simples de um classificador não paramétrico é o classificador do Vizinho mais próximo (NN) (DUDA; HART; STORK, 2001). Este é um método intuitivo em que cada padrão desconhecido a ser classificado é atribuído à classe do vizinho mais próximo dentro do espaço de treinamento.

Essa proximidade entre as amostras é baseada na distância que um padrão desconhecido possui em relação às classes definidas em um espaço de treinamento. Essa distância pode ser calculada de diversas formas, entre as quais, podem-se destacar as distâncias de Hamming, euclidiana e quarteirão (*city-block*) (GONZALEZ; WOODS, 2008).

Este método é altamente sensível ao ruído, gerando erros de classificação para regiões do espaço característico em que há uma mistura de várias classes. Para melhorar o desempenho, a abordagem do NN pode ser estendida para o método do k-vizinhos mais próximo

4.3.3 *k*-Vizinhos mais próximo (KNN)

O classificador k-Vizinhos mais próximo (KNN) é uma extensão da técnica do vizinho mais próximo (DUDA; HART; STORK, 2001). Conforme seu nome indica, para cada padrão desconhecido determinam-se os k vizinhos mais próximos. Neste caso, ao invés de assumir a classe do vizinho mais próximo, k vizinhos próximos são determinados, e a classe assumida é aquela que possui a maioria desses vizinhos.

O parâmetro k é determinado para cada tipo de aplicação. Geralmente é estabelecido um valor pequeno ($k = 3$ e 5 são os valores mais comuns para este parâmetro), porém na realidade o valor de k depende da quantidade de dados disponíveis. A melhor maneira de determinar o valor de k é através de experimentação (DUDA; HART; STORK, 2001).

4.3.4 Redes Neurais Artificiais

Uma Rede Neural Artificial (RNA) pode ser vista como um modelo matemático composto de muitos elementos computacionais não lineares, chamados de neurônios, operando em paralelo e conectados por ligações caracterizadas por diferentes pesos (ARAÚJO, 2004).

Um simples neurônio v_k, calcula a soma das entradas (x_1, x_2, ..., x_i) ponderadas pelos pesos (w_{k1}, w_{k2}, ..., w_{ki}) que cada conexão possui e o bias (b_k), direciona este resultado para uma função de ativação não linear $\varphi(.)$ para produzir uma saída simples y_k denominada nível de ativação daquele neurônio. Um modelo de um neurônio pode ser observado na Figura 4.1.

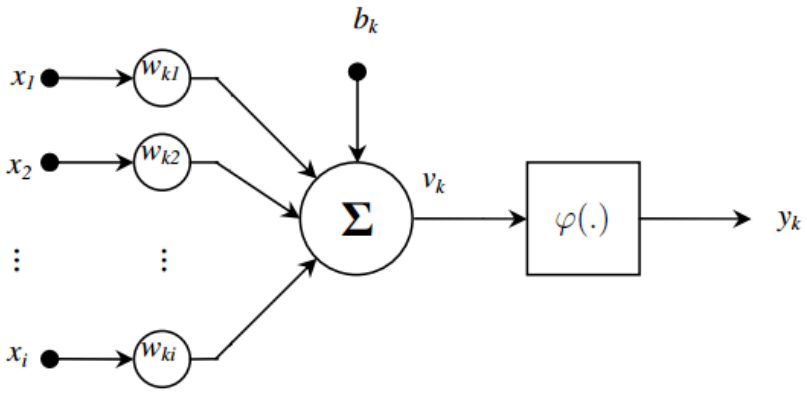


Figura 4.1: Modelo não linear de um neurônio (HAYKIN, 2001).

Modelos de redes neurais são especificados pela topologia da rede, características dos neurônios e regras de aprendizagem ou treinamento. O termo topologia refere-se à estrutura da rede como um todo, especificando como as entradas, as saídas e as camadas escondidas são interconectadas (HAYKIN, 2001). Neste trabalho, é utilizada a topologia do Perceptron Multicamadas (MLP).

A rede MLP é uma das RNAs mais utilizadas para separar dados não-linearmente separáveis. Para iniciar o processo de aprendizagem da rede neural, faz-se necessária a seleção de um conjunto de amostras das classes (conjunto de treinamento) a serem reconhecidos pela rede e suas saídas desejadas correspondentes. Deve-se selecionar para treinamento amostras representativas de cada classe e um número suficiente destas amostras para que a rede possa aprender a identificar os padrões.

Basicamente, uma rede do tipo MLP apresenta três ou mais camadas de

neurônios, a saber, um conjunto de unidades sensoriais (nós de fonte) que constituem a camada de entrada, uma ou mais camadas ocultas e uma camada de saída. A sua topologia é completamente interconectada na direção da camada de entrada para a de saída sem retroalimentação. O sinal de entrada se propaga através da rede, camada por camada, até a saída (HAYKIN, 2001).

A definição do número de neurônios das camadas de entrada e saída é realizada de acordo com o problema em questão. O número de neurônios da camada intermediária, ou mesmo o número de camadas intermediárias, é definido de forma intuitiva, não havendo portanto, uma regra que defina o seu número. Se a quantidade de neurônios escolhidos for pequena, pode acontecer com que alguns neurônios especializem-se em características não úteis, tais como ruído. Se o número de neurônios for insuficiente, pode acontecer da rede não conseguir aprender os padrões desejados (SILVA; SPATTI; FLUAZINO, 2010).

O neurônio individual é o bloco construtivo de cada camada, que é caracterizado principalmente por sua função da ativação. A função de ativação mais comumente utilizada é a função logística e é definida como (HAYKIN, 2001)

$$f(x) = \frac{1}{1 + exp[-\beta \cdot x]}. \tag{4.5}$$

Outra função de ativação bastante utilizada é a função tangente hiperbólica, e é definida como

$$f(x) = \frac{1 - exp[-\beta \cdot x]}{1 + exp[-\beta \cdot x]}, \tag{4.6}$$

em que $\beta > 0$ está associado com a inclinação da função em relação ao ponto de inflexão (SILVA; SPATTI; FLUAZINO, 2010).

As redes MLPs são projetadas para aproximar uma relação entre entrada e saída não conhecida, através dos pesos de cada conexão, via regras de aprendizagem. Uma característica de grande importância deste modelo é o aprendizado supervisionado baseado em duas etapas: propagação e adaptação.

A propagação ocorre na fase de treinamento da rede e consiste em fornecer à rede um conjunto de estímulos (padrões de entradas) e a saída desejada correspondente ao padrão de entrada apresentado. Nesta fase, o primeiro padrão de entrada é propagado até a saída. Durante este passo os pesos sinápticos não mudam de valor.

Na fase de adaptação, o sinal do erro é computado (resultado da diferença

entre a saída desejada e saída real da rede) e transmitido de volta para cada neurônio da camada intermediária que contribuiu para a saída obtida. Sendo assim, cada neurônio da camada intermediária recebe somente uma parte do erro total, conforme a contribuição relativa que o neurônio obtém na saída gerada. Este processo repete-se camada por camada, até que cada neurônio da rede receba o seu peso correspondente. Tal processo é conhecido como retropropagação do erro, pois, o aprendizado baseia-se na propagação retroativa do erro, contra a direção das conexões sinápticas da rede (HAYKIN, 2001).

Os pesos existentes nas conexões entre os neurônios são atualizados de acordo com o erro recebido pelo neurônio associado. Esta atualização é um processo iterativo em que a rede ajusta seus pesos até que a informação do ambiente seja aprendida. O processo de aprendizagem termina quando a saída obtida pela rede neural, para cada um dos padrões de entrada, for próxima o bastante da saída desejada, de forma que a diferença entre ambas seja aceitável. Esta diferença é obtida pelo cálculo do Erro Quadrático Médio (MSE).

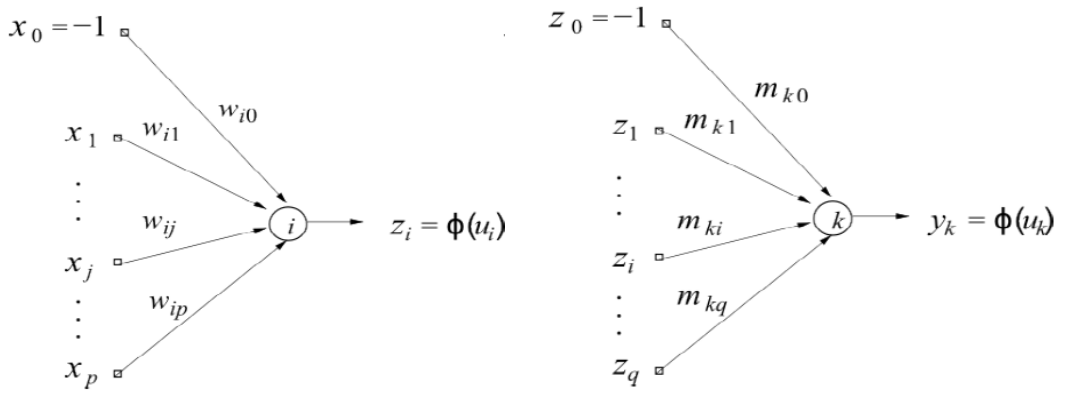
4.3.5 *Extreme Learning Machine*

Muitos algoritmos de treinamento de RNAs são baseados em gradiente descendente. Esses algoritmos são geralmente lentos e convergem facilmente para mínimos locais. Nesses algoritmos, o treinamento é realizado iterativamente para conseguir uma melhor generalização, o que resulta em longos intervalos de tempo para treinar a rede. Além disso, na maioria dos algoritmos tradicionais de treinamento, tais como *backpropagation*, todos os parâmetros das redes, ou seja, todos os pesos e *bias* devem ser ajustados durante o treinamento.

A arquitetura *Extreme Learning Machine* (ELM) é uma rede neural do tipo *feedforward*, isto é sem realimentação, com apenas uma camada de neurônios ocultos (HUANG; ZHU; SIEW, 2006).

Esta arquitetura é semelhante a da rede Perceptron Multicamadas (MLP), mostrado na Figura 4.2, em que os neurônios da camada oculta (primeira camada de pesos sinápticos) são representados na Figura 4.2(a), enquanto os neurônios da camada de saída (segunda camada de pesos sinápticos) são representados na Figura 4.2(b).

O vetor de pesos $\mathbf{w}$ associado a cada neurônio i da camada oculta é representado como



(a) (b)

Figura 4.2: (a) Neurônio da Camada Oculta. (b) Neurônio da Camada de Saída.

$$\mathbf{w}_i = \begin{pmatrix} w_{i0} \\ w_{i1} \\ \vdots \\ w_{ip} \end{pmatrix} = \begin{pmatrix} \theta_i \\ w_{i1} \\ \vdots \\ w_{ip} \end{pmatrix}, \tag{4.7}$$

em que θ_i é o limiar ou *bias* associado ao neurônio i. Os neurônios desta camada são chamados de neurônios ocultos por não terem acesso direto saída da rede, onde são calculados os erros de aproximação. De modo semelhante, o vetor de pesos $\mathbf{m}$ associado a cada neurônio k da camada de saída é definido por

$$\mathbf{m}_k = \begin{pmatrix} m_{k0} \\ m_{k1} \\ \vdots \\ m_{kq} \end{pmatrix} = \begin{pmatrix} \theta_k \\ m_{k1} \\ \vdots \\ m_{kq} \end{pmatrix}, \tag{4.8}$$

em que θ_k é o *bias* associado ao neurônio de saída k.

No algoritmo ELM não é necessário ajustar os pesos e os bias dos neurônios ocultos na etapa de treinamento. Esses pesos w_{ij}, $i = \{1; \ldots; q\}$ e $j = \{0; \ldots; p\}$ são inicializados com valores aleatórios utilizando, como por exemplo, as distribuições de probabilidade uniforme ou normal. Os pesos da camada oculta da rede neural podem ser definidos pela matriz de pesos $\mathbf{W}$, com q linhas e $p + 1$ colunas

$$\mathbf{W} = \begin{pmatrix} w_{10} & w_{11} & \dots & w_{1p} \\ w_{20} & w_{21} & \dots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{q0} & w_{q1} & \dots & w_{qp} \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_q^T \end{pmatrix}, \tag{4.9}$$

em que a i-ésima linha da matrix $\mathbf{W}$ é composta pelo vetor de pesos do i-ésimo neurônio oculto.

Dado um vetor de entrada $\mathbf{x}(t)$, na iteração t, o primeiro passo é calcular as ativações dos neurônios da camada oculta como

$$u_i(t) = \sum_{j=0}^{p} w_{ij} x_j(t) = \mathbf{w}_i^T \mathbf{x}(t) \;\; i = 1, \dots, q, \tag{4.10}$$

em que q indica o número de neurônios da camada escondida. Essa operação sequencial pode ser realizada de uma única vez utilizando a notação matricial. Neste caso, tem-se que o vetor de ativações $\mathbf{u}(t)$ dos neurônios ocultos na iteração t é calculado como

$$\mathbf{u}(t) = \mathbf{W}\mathbf{x}(t). \tag{4.11}$$

Em seguida, as saídas correspondentes dos neurônios da camada oculta são calculadas, em notação matricial, como

$$\mathbf{z}(t) = \phi(\mathbf{u}(t)) = \phi(\mathbf{W}\mathbf{x}(t)), \tag{4.12}$$

em que ϕ é a função de ativação e pode ser definida pelas funções logística e tangente hiperbólica (HAYKIN, 2001). Assim, para as N amostras de treinamento, $\mathbf{x}(t)$, $t = \{1; \dots; N\}$, tem-se um vetor $\mathbf{z}(t)$ correspondente, que pode ser disposto como uma coluna de uma matriz $\mathbf{Z}$. Esta matriz terá dimensão q linhas por N colunas:

$$\mathbf{Z} = [\mathbf{z}(1)|\mathbf{z}(2)| \dots |\mathbf{z}(N)]. \tag{4.13}$$

Os pesos da camada de saída são determinados analiticamente a partir da matriz $\mathbf{Z}$. Para determiná-los, inicialmente considera-se que para cada vetor de entrada

$\mathbf{x}(t)$, $t = \{1; \ldots; N\}$ existe um vetor de saídas desejadas $\mathbf{d}(t)$ correspondente. Estes N vetores podem ser organizados ao longo das colunas de uma matriz $\mathbf{D}$, que terá dimensão m linhas e N colunas:

$$\mathbf{D} = [\mathbf{d}(1)|\mathbf{d}(2)|\ldots|\mathbf{d}(N)]. \tag{4.14}$$

O cálculo dos pesos da camada de saída pode ser entendido como o cálculo dos parâmetros de um mapeamento linear entre a camada oculta e a camada de saída (MELO, 2011), em que o vetor $\mathbf{z}(t)$ é a entrada da camada de saída na iteração t e o vetor $\mathbf{d}(t)$ é a saída. Assim, busca-se determinar a matriz $\mathbf{M}$ que melhor represente a transformação

$$\mathbf{d}(t) = \mathbf{M}\mathbf{z}(t). \tag{4.15}$$

Para isso, utiliza-se o método dos mínimos quadrados, também conhecido como método da pseudoinversa (BROOMHEAD; LOWE, 1988). Assim, usando as matrizes $\mathbf{Z}$ e $\mathbf{D}$, a matriz de pesos $\mathbf{M}$ é definida por

$$\mathbf{M} = \mathbf{D}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1} = \mathbf{H}^\dagger\mathbf{D}, \tag{4.16}$$

em que $\mathbf{H}^\dagger = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ é a matriz inversa generalizada de Moore-Penrose (HUANG; ZHU; SIEW, 2006).

Uma vez determinadas as matrizes de pesos $\mathbf{W}$ e $\mathbf{M}$, a rede ELM está pronta para uso. Na etapa de teste da rede ELM, as ativações dos neurônios da camada de saída são calculadas por

$$\mathbf{a}(t) = \mathbf{M}\mathbf{z}(t), \tag{4.17}$$

em que m é o número de neurônios de saída.

Na rede ELM, assume-se que os neurônios de saída usam a função identidade como função de ativação, ou seja, as saídas destes neurônios são iguais às suas ativações e definidas como

$$y_k(t) = \phi(a_k(t)) = a_k(t) \tag{4.18}$$

Por generalização adequada entende-se a habilidade da rede em utilizar o conhecimento armazenado nos seus pesos e limiares para gerar saídas coerentes para novos vetores de entrada, ou seja, vetores que não foram utilizados durante o treinamento. A generalização é considerada boa quando a rede, durante o treinamento, é capaz de aprender adequadamente a relação entrada-saída do mapeamento de interesse (MELO, 2011).

Baseado no que foi descrito anteriormente, o algoritmo ELM pode ser resumido em três etapas básicas:

- inicializar aleatoriamente os pesos de entrada e os bias dos neurônios da camada oculta;
- calcular a matriz de saída da camada oculta;
- obter a matriz de pesos da camada de saída a partir da matriz modificada, utilizando a matriz generalizada inversa de Moore-Penrose.

4.4 Extração de características: proposta

A técnica de extração de características proposta para detectar patologias na laringe é baseada numa modificação da decomposição *wavelet* padrão proposta por Farooq e Datta (2003) e consiste em dividir o *frame* de áudio em 4 sub-*frames* de duração de 6 ms para acomodar oscilações rápidas e permitem acompanhar a evolução temporal da energia média por amostra em cada banda de frequência. Como o sinal de voz é aproximadamente estacionário por um período de até 10 ms, devido à limitação física de movimento nas articulações de produção da fala, qualquer redução na duração do quadro não é útil.

A decomposição *wavelet* é aplicada em cada sub-*frame* e a energia dos coeficientes *wavelet* em cada faixa de freqüência é calculada. Esta energia é normalizada pelo número de amostras na faixa correspondente, resultando desse modo uma energia média por amostra em cada faixa. A normalização é essencial porque cada faixa terá um número diferente de amostras. Estas energias médias por amostra, para diferentes faixas, são usadas como características para classificação. As características extraídas em um sub-*frame* base não fornecem somente a energia em cada faixa, mas fornecem também uma ideia da variação temporal da energia em cada faixa.

A quantidade de características extraídas depende do nível de decomposição de cada sub-*frame*. Uma decomposição nível N de um sub-*frame* extrai $N + 1$ características referentes aos N coeficientes de detalhes e 1 coeficiente de aproximação. Desta forma, para cada *frame*, extrai-se um total de $4N + 4$ características. Por exemplo, para uma decomposição nível 3 de um sub-*frame* são extraídas 4 características e, assim, para um *frame* tem-se um total de 16 características. Essas características são agrupadas em um vetor de atributos que será utilizado para classificação.

A Tabela 4.1 apresenta um resumo do algoritmo proposto.

Tabela 4.1: Resumo do algoritmo usado para a análise do sinal de voz.

i. Avaliação manual das amostras de voz

- i.1 Eliminação dos fonemas indesejados
- i.2 Seleção das partes estacionárias Estacionário no sentindo amplo (WSS)

ii. Pré-processamento do sinal

- ii.1 Normalização da amplitude

iii. Decomposição *wavelet*
Para cada frame de áudio

- iii.1 Dividir em 4 sub-*frames*
- iii.2 Aplicar a decomposição wavelet em cada sub-*frame*
- iii.3 Calcular a energia média em cada faixa de frequência de cada sub-*frame*
- iii.4 Agrupar as energias médias em um vetor de atributos

4.5 Resumo do Capítulo

Neste Capítulo foram brevemente descritos o reconhecimento de padrões, as características dos classificadores utilizados neste trabalho, além da proposta de extração de características. No capítulo seguinte são descritos o conjunto de dados utilizados, a metodologia de simulação e os resultados obtidas pela técnica proposta para extração de características usando a Transformada *Wavelet* Discreta (DWT) em comparação com a técnica padrão.

Capítulo 5

Simulações Computacionais: Desempenho e Análise

Neste Capítulo são mostrados os resultados experimentais de análise e classificação dos padrões obtidos a partir da extração das características com a decomposição *wavelet* padrão e proposta. Na primeira seção é apresentado o conjunto de dados com as quatro classes das amostras de vozes saudáveis e patológicas utilizado nesta dissertação. Na próxima seção, é descrita a metodologia de simulação para obtenção dos resultados desse trabalho. Na Seção 3, é apresentado uma análise das quatro classes de amostras de vozes. Em seguida, os principais resultados obtidos através dos algoritmos de reconhecimento de padrões descritos no Capitulo 4 são apresentados em duas seções analisando a influência da classe de disfonia neurológica no desempenho de classificação.

5.1 Conjunto de Dados

O conjunto de dados utilizado nos testes é formado por amostras de voz coletadas de 60 voluntários de ambos os sexos, com idade entre 18 e 90 anos, divididos em 4 grupos de aproximadamente mesmo tamanho.

O primeiro grupo é composto de amostras de voz de pessoas saudáveis sem patologias na voz. O segundo grupo é composto por pessoas com nódulos vocais em diferentes estágios de evolução da doença (SCALASSARA *et al.*, 2007). O terceiro grupo é composto por amostras de pessoas com edema de Reinke. E o último grupo é composto por pacientes que apresentam diferentes distúrbios

neurológicos, como AVC (Acidente Vascular Cerebral), Doença de Huntington, Doenças de Parkinson, ELA (Esclerose Lateral Amiotrófica), Mononeurite múltipla, Mitocondropatia, Distrofia de Duchenne, Distrofia Miotônica e Distonia Cervical, que são agrupados como pacientes com disfonia neurológica.

Essas amostras de voz são parte de um banco de dados de voz do Grupo de Bioengenharia da Escola de Engenharia de São Carlos da Universidade de São Paulo, Brasil. Esses sinais foram coletados ao longo dos últimos 10 anos e tiveram o consentimento dos pacientes para o uso em diversos estudos (ROSA; PEREIRA; GRELLET, 2000; SCALASSARA *et al.*, 2007; SCALASSARA *et al.*, 2009).

Os pacientes submetidos à análise foram diagnosticados por médicos do Departamento de Otorrinolaringologia e do Departamento de Cabeça e Pescoço do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, Estado de São Paulo, Brasil, por meio de vídeo-laringoscópio e luz estroboscópica. Os indivíduos do grupo de pessoas saudáveis também foram diagnosticados para provar a ausência de qualquer patologia nas cordas vocais.

Os dados foram gravados usando um protocolo similar ao apresentado por Uloza, Saferis e Uloziene (2005), no qual os indivíduos foram convidados a produzir a vogal sustentada /a/ em um tom e nível de intensidade confortáveis por cerca de 5 segundos. O microfone usado neste procedimento estava de acordo com os padrões estabelecidos pela Sociedade Brasileira de Fonoaudiologia. Este foi colocado a uma distância de cinco centímetros da boca da pessoa. Coletas consecutivas foram realizadas, selecionado o sinal com menor variabilidade da voz (SCALASSARA *et al.*, 2009).

Como apresentado em Lass (1979), as vogais são geralmente usadas em estudos de patologias da voz porque as cordas vocais vibram durante a pronuncia de uma vogal. Além disso, a avaliação acústica da função laríngea relaciona-se com a adequação das vibrações vocais sustentadas. Portanto, a fim de coletar os dados, foi utilizado o fonema sustentado /a/ em Português para avaliar os parâmetros acústicos das amostras de voz.

No momento da aquisição da voz, verificaram se o indivíduo poderia lidar com o intervalo de fonação e, em caso negativo, ele foi convidado a parar. Esse procedimento foi importante, pois a manutenção do enunciado provoca um aumento da freqüência fundamental e uma estabilidade artificial sobre a sua produção (ROSA; PEREIRA; GRELLET, 2000).

A fim de evitar a influência dos fenômenos de transição, o início e o término do sinal de voz adquiridos foram descartados. Assim, garantindo que esses trechos do sinal não influenciaram o resultado final. Após esta etapa, a amplitude do sinal foi normalizada de acordo com seu valor máximo absoluto para eliminar a influência de diferentes níveis de som a partir dos sinais coletados. Todas as amostras de vozes foram quantizadas em 16 bits com amplitude e gravada em mono-canal. A freqüência de amostragem foi de 22.050 Hz.

Quatro exemplos dos sinais de voz desse banco de dados, cada um referente a uma das classes, são mostrados na Figura 5.1. O primeiro é um típico sinal de voz saudável para a vogal sustentada /a/, o segundo é um sinal de voz gerado por um paciente portador do Edema de Reinke, o terceiro é um sinal de voz gerado por um paciente portador de Nódulo Vocal e o último é de um sinal caracterizado pela presença de Disfonia Neurológica no paciente.

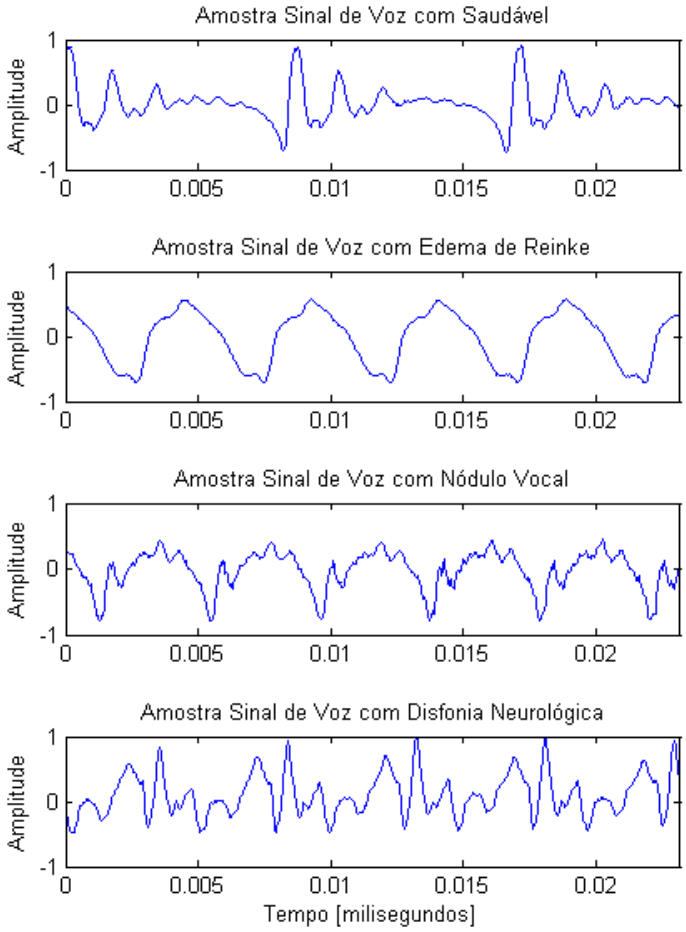


Figura 5.1: Exemplos de sinais de voz para vogais sustentadas /a/.

5.2 Metodologia de Simulação

As simulações realizadas neste trabalho são executadas em um notebook da marca Lenovo com processador Intel Core i3 de 2,40 GHz e 3 GB de memória RAM com sistema operacional Windows 7. Todas as simulações foram realizadas utilizando-se o software Maltab versão 2010b.

O intuito dessa avaliação é realizar um estudo estatístico do desempenho de classificação dos sinais de vozes patológicas e saudáveis baseados nas características extraídas utilizando a decomposição *wavelet* padrão e proposta, e a partir da análise desses resultados, validar essa ferramenta matemática como eficiente para este tipo de estudo.

O sinal de voz é não-estacionário e ruidoso, mas pode ser considerado estacionário para períodos de tempo entre 10 e 30 milissegundos (MAKHOUL, 1975; DELLER; PROAKIS; HANSEN, 2000; CARVALHO, 2009). Assim, antes da etapa de extração de características, implementa-se o janelamento do sinal utilizando uma janela retangular que é movida ao longo do sinal de voz sem sobreposição entre *frames* adjacentes. O tamanho dessa janela nos testes é de 24 milissegundos (ms), ou seja, de 512 pontos de áudio por *frame* para um taxa de amostragem de 22.050Hz.

Na etapa de extração de características utilizando a decomposição *wavelet* padrão e proposta, varia-se a quantidade de características extraídas por *frame* de áudio. Em ambas as técnicas, varia-se a quantidade de níveis de decomposição de 3 a 7, sendo o nível de decomposição *wavelet* 7 o que gera a menor banda de frequência, de 0-172,4 Hz. Qualquer outra decomposição não aumenta o desempenho de reconhecimento, pois o conteúdo de baixa freqüência é insignificante e, portanto, não tem informações discriminatórias. A quantidade de características extraídas depende do nível de decomposição. Para uma decomposição nível N são extraídas $N + 1$ características referente aos N coeficientes de detalhes e 1 coeficiente de aproximação (MALLAT, 1989). Assim, para a variação de 3 a 7 níveis de decomposição, a quantidade de características extraídas varia de 4 a 8.

Uma vez extraídos os atributos usando as duas técnicas, são utilizados os classificadores estudados na Seção 4.3 para para obter uma função discriminante para separar as diferentes classes presentes no espaço de características.

Para as redes neurais MLP e ELM, é utilizada uma configuração com duas camadas, sendo uma camada oculta cujo número de neurônios q foi determinado

usando a Regra de Kolmogorov, $q = 2 \cdot n + 1$, em que n é o número de atributos usados na classificação. Diversos experimentos foram realizados utilizando as principais métricas conhecidas na literatura para determinação de q e a Regra de Kolmogorov obteve os melhores resultados médios, sendo, portanto, escolhida para obter os resultados deste trabalho. Os pesos dos neurônios dessas redes são inicializados com uma distribuição uniforme no intervalo $(0; 0.1)$ para acelerar a convergência da rede. Para rede MLP os parâmetros de treinamento são escolhidos empiricamente e são: 200 épocas de treinamento, MSE desejado de 10^{-4}, passo de aprendizagem de $0,01$ e fator de momento $0,2$.

Para os classificadores KNN e NN é utilizada a métrica euclidiana para o cálculo da distância. No k-Vizinhos mais próximo (KNN), o valor de k pode ser determinado empiricamente após diversas simulações variando esse parâmetro, e o valor $k = 5$ obteve os melhores resultados.

O desempenho dos classificadores é obtido de 20 simulações independentes, sob as mesmas condições, utilizando as amostras que são embaralhadas aleatoriamente e divididas em 80% para o conjunto de treinamento e 20% para o de teste. A avaliação da técnica é feita com base nas taxas de acerto média, máxima, mínima e desvio-padrão.

5.3 Análise das Características Extraídas

Para análise, selecionou-se manualmente um trecho de cada amostra de voz com características acústicas adequadas, a fim de eliminar fenômenos transitórios e variações da voz presentes na gravação. Os sinais foram pré-processados através de uma normalização min-max (HAN; KAMBER; PEI, 2006) para resultar em sinais com amplitudes entre 0 e 1. Para todas as amostras, selecionou-se cerca de 1 s de amostra de voz e dividiu-se em 40 partes de 512 pontos, de aproximadamente 24 ms cada.

Após o pré-processamento dos sinais de voz, aplicou-se a decomposição *wavelet* padrão sobre cada *frame* de 24 ms, extraindo a energia dos coeficientes *wavelet* conforme definido na Subseção 3.2.1. Inicialmente tentou-se classificar as amostras em saudável, nódulo vocal, edema de Reinke e disfonia neurológica sem o auxílio de um algoritmo de reconhecimento de padrão.

Para isso, foram utilizados os conceitos de energia *wavelet* relativa e entropia

wavelet, definidos nas Subseções 3.2.2 e 3.2.3. A partir da energia *wavelet* relativa foi estimado a distribuição de probabilidade de cada *frame* e, assim, calculada a entropia de Shannon (SHANNON, 1948). Essa métrica foi utilizada por ser um critério útil para analisar e comparar a distribuição de probabilidade, já que fornece uma medida da informação para qualquer distribuição de probabilidade.

Os resultados obtidos pela entropia *wavelet* não foram suficientes para descriminar as quatro classes. Com isso, calculou-se a Entropia *Wavelet* Relativa (RWE) entre as amostras de vozes patológicas e a saudável utilizando o conceito definido na Subseção 3.2.3. A RWE é calculada para cada par dos pedaços do sinal. Quatro casos são considerados para estimação da entropia: primeiro, amostras saudáveis versus amostras saudáveis (considerado como grupo de controle); segundo, amostras saudáveis versus amostras de nódulo vocal; terceiro, amostras saudáveis versus amostras de edema de Reinke; e por fim, amostras saudáveis versus amostras de disfonia neurológica.

Para cada amostra saudável, após todas as combinações de pedaços das amostras são usados para cada caso apresentado, a média dos valores da RWE é calculada. Portanto, existem 14 amostras saudáveis e quatro grupos, o número total de comparações é 56. Cada cálculo da média é executada através de 22400 valores, que é obtido pelas 40 partes de 24 ms do grupo do saudável por 40 pedaços das 14 amostras do grupo escolhido para análise.

Os resultados destes 56 valores médios para cada um dos casos: amostras saudável-saudável, amostras saudável-nódulo, amostras edema-saudável e amostras saudável-neurológica são apresentados na Figura 5.2. Os 14 pontos de cada caso são ordenados por sua avaliação, portanto, eles não estão emparelhados. Como pode ser visto na Figura 5.2, a média dos valores da RWE das amostras saudáveis pelas amostras de nódulo vocal e de Edema de Reinke são mais elevados do que a das estimativas saudável-saudáveis. Assim para valores de RWE maiores que 0.5 *bits*, é possível afirmar que existe a presença de uma patologia do tipo Edema de Reinke ou Nódulo Vocal. Entretanto, a diferença entre os dois grupos de patologias não é suficiente para classificar utilizando apenas essa métrica.

Além disso, na Figura 5.2 observa-se que para as amostras de disfonia neurológica tem RWE maior que as amostras saudáveis e menor que as amostras de Edema de Reinke e de Nódulo Vocal em 85% dos amostras.

Esses resultados obtidos pela RWE mostraram que essa métrica inicialmente

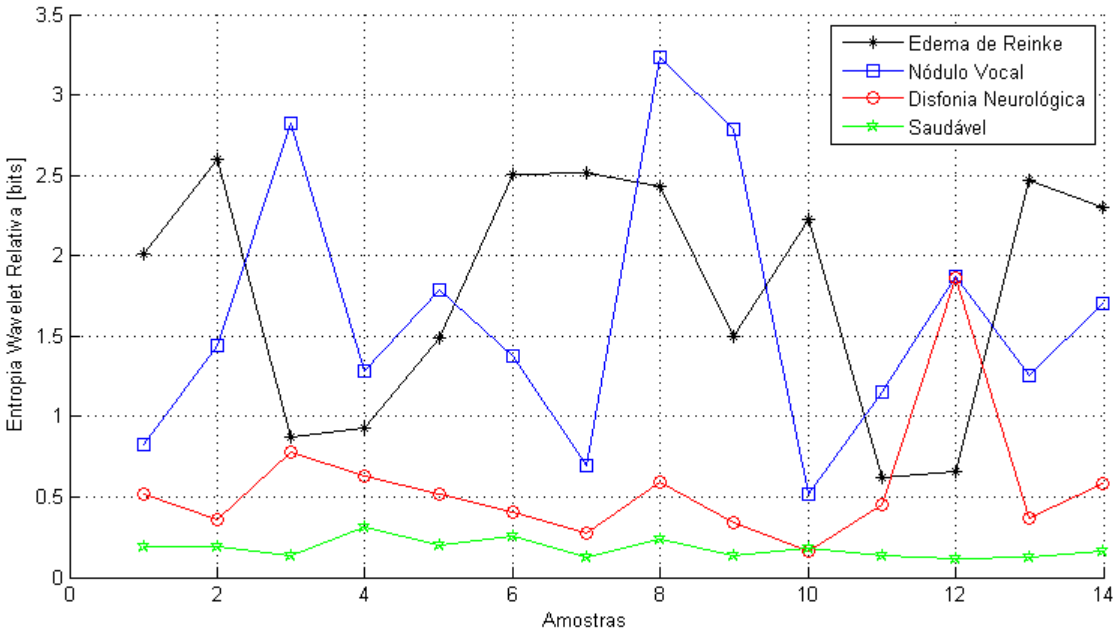


Figura 5.2: Média dos valores de RWE entre cada uma das amostras saudáveis por cada um dos quatro casos de estudo: primeiro, amostras saudáveis (grupo de controle); segundo, amostras de nódulos vocais; terceiro; amostras de edema de Reinke; e quarto, amostras de disfonia neurológica.

não é suficiente para classificar cada tipo de patologia, porém permite classificar as amostras de voz em saudável e não saudável. Assim nas próximas seções, primeiramente será utilizado os classificadores de padrões para separar as classes de Edema de Reinke, de Nódulo Vocal e Saudável. Em seguida, será incluída a classe de Disfonia Neurológica para avaliar o desempenho dos mesmos classificadores.

5.4 Desempenho sem classe de Disfonia Neurológica

Nesta seção, são apresentados os resultados obtidos pelos classificadores mostrados na Seção 4.3 sem utilizar a classe de amostras de pessoas com Disfonia Neurológica, conforme realizados nos trabalhos (SCALASSARA *et al.*, 2009; SCALASSARA *et al.*, 2007) com mesma banco de dados. Na primeira, são apresentados os resultados para a extração proposta. Em seguida, são apresentados os resultados obtidos pela extração de características padrão, comparando com os resultados obtidos da técnica proposta.

Como características para classificação, aplicou-se a decomposição *wavelet* proposta sobre cada *frame* de 24 ms, extraindo a energia dos coeficientes *wavelet* conforme definido na Subseção 3.2.1, variando a quantidade de características

extraídas em cada sub-*frame*. Para classificação, foram utilizados os classificadores mostrados na Seção 4.3.

5.4.1 Resultados: Técnica Proposta

Para avaliar a técnica proposta, as características extraídas foram aplicadas aos classificadores *Extreme Learning Machine* (ELM), KNN, Vizinho mais próximo (NN) e Perceptron Multicamadas (MLP). As estatísticas da taxa de acerto e o desvio padrão da taxa de acerto obtidas pela técnica proposta usando as amostras de voz saudável, nódulo vocal e edema de Reinke são mostradas na Tabela 5.1.

Tabela 5.1: Desempenho dos classificadores para a técnica proposta com as classes: voz saudável, nódulo vocal e edema de Reinke.

Classificador	Características por sub-*frame*	Taxas de Reconhecimento(%)			
		mínima	*média*	*máxima*	*desvio padrão*
MLP	4	79,35	82,23	84,51	1,98
	5	88,38	90,01	91,95	1,30
	6	89,53	90,27	90,93	0,55
	7	**91,57**	**93,03**	**94,00**	**0,70**
	8	84,51	86,63	88,32	1,67
ELM	4	64,67	70,43	75,00	0,07
	5	63,86	71,92	76,09	0,10
	6	74,18	77,16	80,98	0,03
	7	74,73	77,09	81,25	0,03
	8	**73,37**	**78,53**	**82,88**	**0,05**
KNN	4	77,99	80,76	82,88	0,02
	5	76,63	79,27	82,88	0,02
	6	75,54	81,02	85,60	0,05
	7	80,71	84,28	86,96	0,03
	8	**80,43**	**84,71**	**87,77**	**0,04**
NN	4	77,45	80,11	83,42	0,03
	5	76,90	78,94	82,34	0,03
	6	79,62	82,88	87,50	0,05
	7	**81,79**	**85,11**	**88,86**	**0,03**
	8	81,52	84,66	86,68	0,01

A Tabela 5.1 mostra o desempenho dos classificadores variando a quantidade de

características extraídas em cada sub-*frame*. Como esperado existe um aumento no desempenho de reconhecimento quando o número de características é aumentada. Esse ganho de desempenho foi maior nas redes neurais MLP e ELM com variação de aproximadamente 11% e 8% respectivamente. Porém esse crescimento não é constante, estabilizando com 6 a 8 características. Isto ocorre porque o nível mais elevado de decomposição *wavelet* gera as características da banda de frequência muito baixa, que não contém nenhuma informação discriminatória.

Com o intuito de avaliar conjuntamente os desempenhos médios de todos os classificadores foi gerado o gráfico de barras da Figura 5.3. Cada conjunto de barras representa um algoritmo, de modo que a avaliação e a comparação do desempenho dos algoritmos para o método proposto é mais fácil de avaliar.

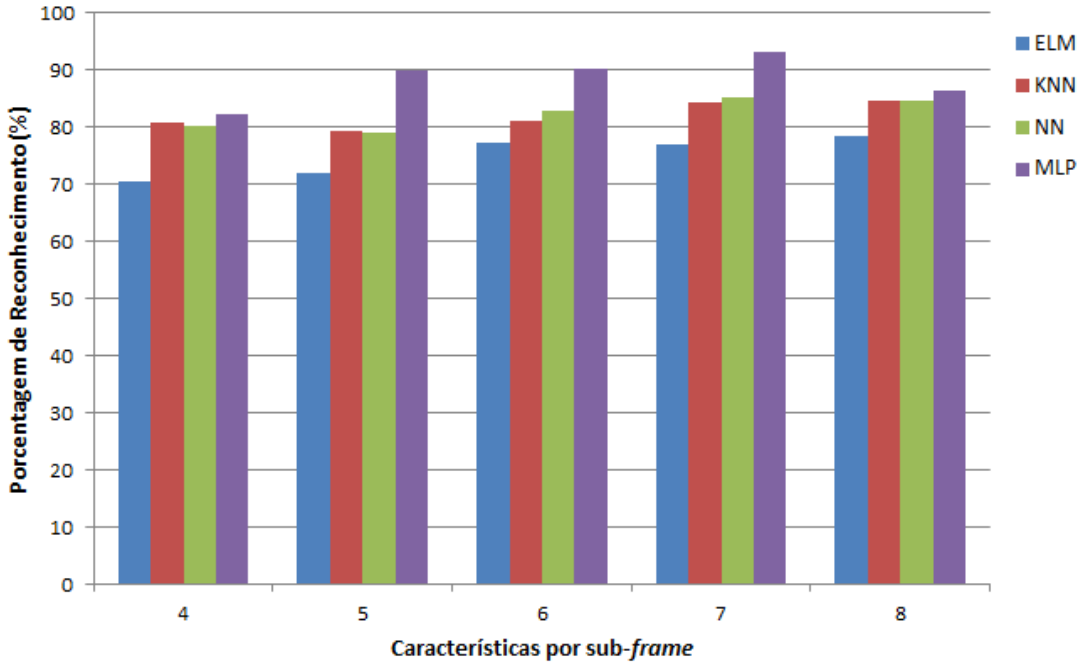


Figura 5.3: Variação do desempenho dos classificadores estudados para a técnica proposta sem a classe de Disfonia Neurológica.

Conforme pode ser visto na Figura 5.3, a rede ELM obteve desempenho inferior aos dos demais classificadores para todos as configurações. Já os classificadores KNN e NN obtiveram resultados similares e a MLP possuiu o melhor desempenho, com 93,03% em média, devido ao seu alto poder de generalização e especialização devido ao processo de aprendizagem.

O desempenho médio da MLP para cada uma das três classes é mostrado na

Figura 5.4. Neste figura, verifica-se que os melhores resultados para as classes Edema de Reinke e Nódulo Vocal ocorrem com 7 características por sub-*frame* e com taxas acima de 90%. Enquanto para a classe de vozes saudáveis, o melhor resultado ocorreu com 8 características por sub-*frame*, obtendo aproximadamente uma taxa de acerto média de 100%. Os desempenhos médios das classes para 7 características extraídas por sub-*frame* acima de 90% demonstram que a técnica proposta permite discriminar essas duas doenças da laringe com boa probabilidade de acerto.

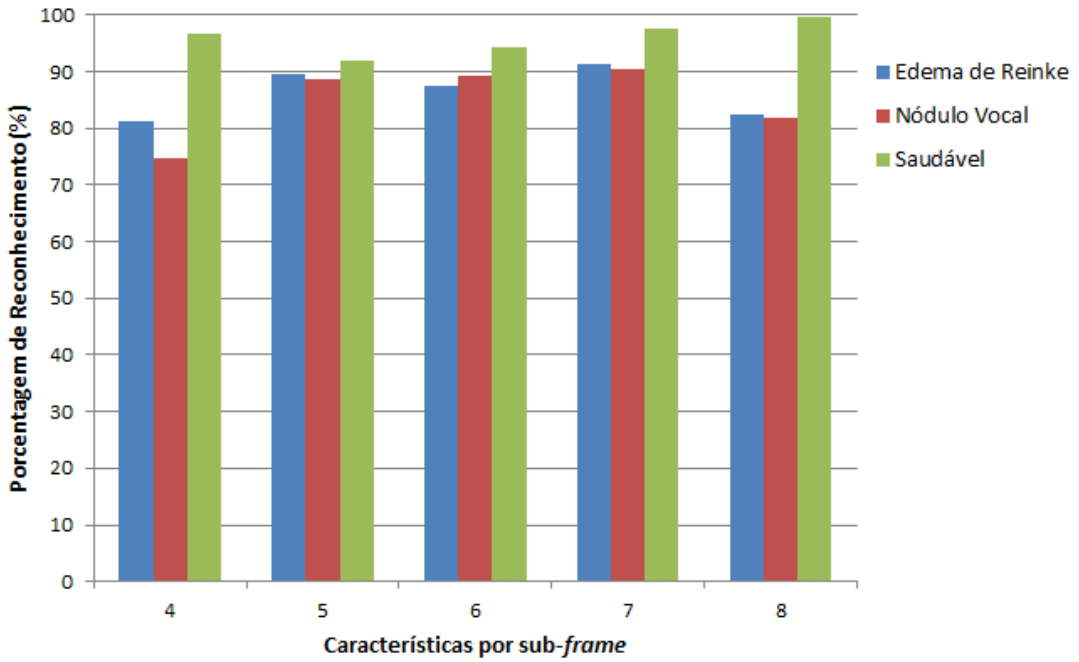


Figura 5.4: Desempenhos médios por classe (Voz Saudável, Nódulo Vocal e Edema de Reinke) usando MLP para a técnica proposta.

5.4.2 Resultados: Técnica Padrão

Os mesmos resultados obtidos para decomposição *wavelet* proposta também foram obtidos para a padrão e são mostrados na Tabela 5.2. Comparando esses resultados com os mostrados na Tabela 5.1, verifica-se que todos os classificadores, exceto a MLP, obtiveram desempenhos similares com as características extraídas pelas duas técnicas.

No classificador MLP verifica-se uma queda de aproximadamente 15% para o mesmo nível de decomposição *wavelet* aplicado, conforme mostrado na Figura 5.5. Como a MLP é um classificador que tem alto poder de generalização e especialização devido ao processo de treinamento, esse resultado mostra que a técnica proposta

consegue extrair um vetor de atributos que permite uma maior separabilidade entre as classes, apesar desse vetor ter uma dimensão 4 vezes maior que o da técnica padrão.

Tabela 5.2: Desempenho dos classificadores para a técnica padrão com as classes: voz saudável, nódulo vocal e edema de Reinke.

Classificador	Características por *frame*	Taxas de Reconhecimento(%)			
		mínima	*média*	*máxima*	*desvio padrão*
MLP	4	61,68	66,44	71,74	3,57
	5	66,03	69,13	73,64	2,78
	6	72,28	74,73	79,62	2,44
	7	**74,73**	**78,45**	**83,15**	**2,61**
	8	75,54	78,34	80,98	2,89
ELM	4	60,87	65,08	70,11	0,05
	5	60,05	66,07	69,84	0,07
	6	69,84	74,05	77,72	0,03
	7	70,65	76,28	80,43	0,06
	8	**75,00**	**78,56**	**81,52**	**0,03**
KNN	4	69,29	73,44	76,63	0,05
	5	69,84	73,82	77,72	0,04
	6	77,17	81,3	84,78	0,04
	7	79,89	84,25	88,32	0,03
	8	**82,61**	**85,19**	**89,13**	**0,04**
NN	4	65,76	69,62	73,37	0,05
	5	70,38	73,76	77,45	0,03
	6	74,73	79,42	84,24	0,05
	7	79,62	83,56	86,96	0,05
	8	**81,79**	**85,24**	**87,23**	**0,02**

A comparação do desempenho de reconhecimento por classe entre as duas técnicas para os melhores resultados da rede MLP é mostrada na Figura 5.6. Este resultado mostra que as características obtidas pela técnica proposta permitem melhorar a desempenho de reconhecimento das classes de vozes patológicas.

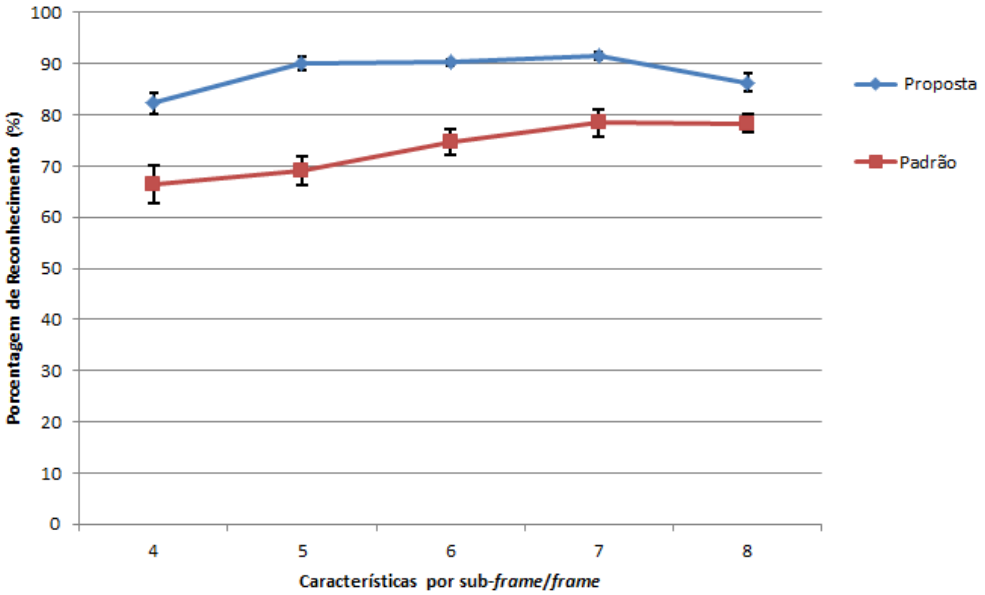


Figura 5.5: Comparação de desempenho das características propostas para a rede MLP.

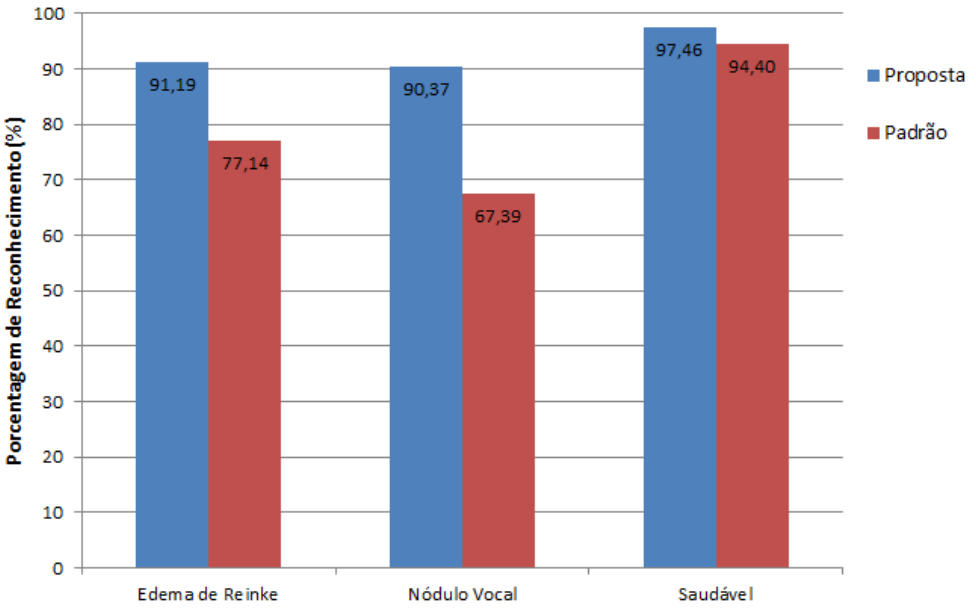


Figura 5.6: Comparação de desempenho de reconhecimento de cada classe para as características propostas usando a rede MLP.

5.5 Desempenho com classe de Disfonia Neurológica

Nesta seção, todas as classes foram utilizadas nas simulações e os resultados foram obtidos para a decomposição *wavelet* padrão e proposta em todos os

classificadores.

5.5.1 Resultados: Técnica Proposta

Incluindo a classe de pessoas com disfonia neurológica, ocorre uma diminuição no desempenho de reconhecimento em todos os classificadores, conforme mostrado na Figura 5.7. Porém a MLP conseguiu manter uma desempenho próximo a 90% com 8 características extraídas em cada sub-*frame*.

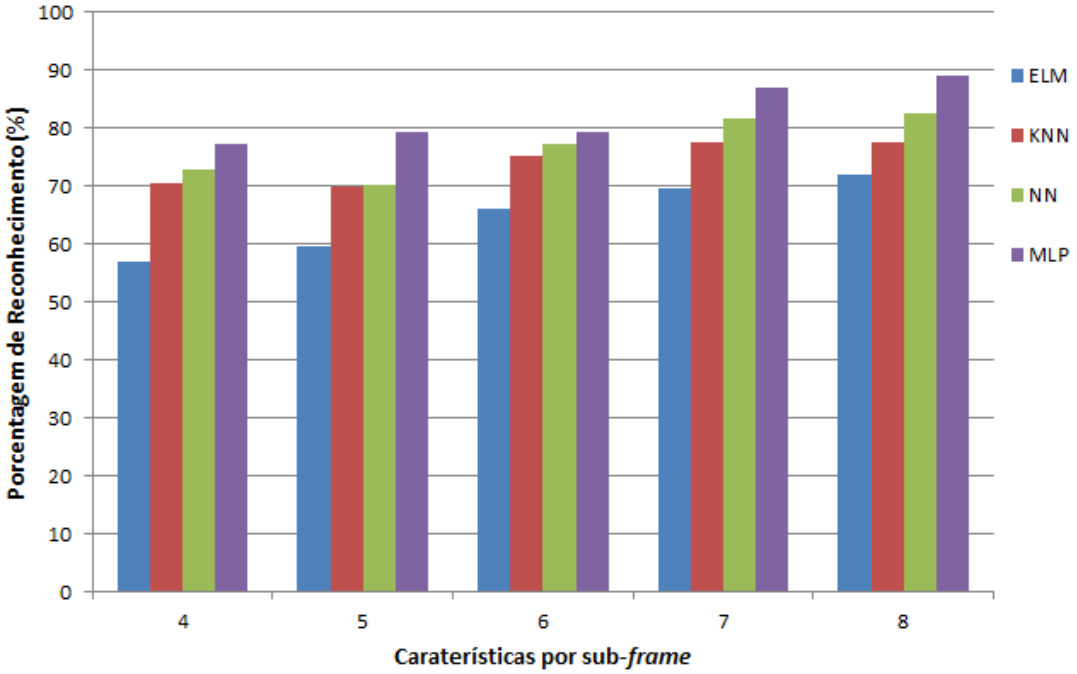


Figura 5.7: Desempenhos médios dos classificadores estudados para a técnica proposta com todas as classes.

Esta queda no desempenho de classificação com a adição da classe Disfonia Neurológica era esperada, em função do resultado mostrado na Seção 5.3, na qual mostrou-se que essa classe é muito parecida com a classe de voz saudável. Além disso, outro motivo que contribuição para o baixo de desempenho de classificação é o fato das amostras dessas classes serem de pacientes com diferentes doenças de origem neurológica, mostrando que essas características extraídas não foram suficientes para separar essa classe das demais.

O desempenho médio de reconhecimento por classe em função da variação na quantidade de características extraídas é mostrado na Figura 5.8. Nesta figura verifica-se que ocorreu um queda no desempenho de classificação de todas as classes

com a inclusão da classe de disfonia neurológica. Apesar disso, todas as classes obtiveram desempenho maior que 86%, o que demonstra que a técnica proposta permite extrair características que permitem classificar as amostras de voz com boa taxa de acerto.

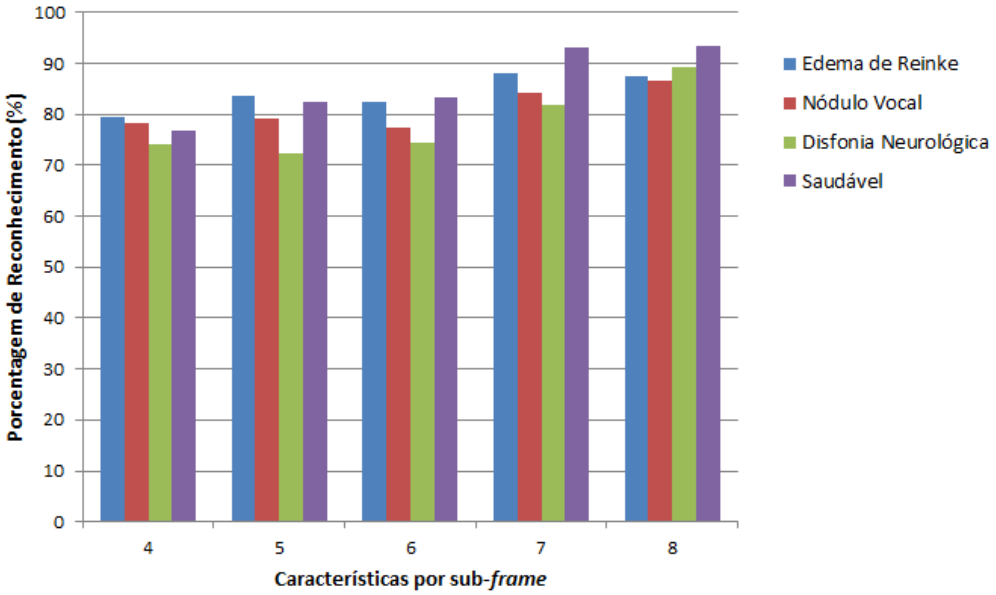


Figura 5.8: Variação do desempenhos médios por classe usando MLP para a técnica proposta.

5.5.2 Resultados: Técnica Padrão

Assim como para a técnica proposta, ocorreu uma queda no desempenho de reconhecimento com a inclusão da classe de disfonia neurológica conforme mostrado na Figura 5.9.

Analisando os resultados da Figura 5.9 e comparando-os aos mostrados na Figura 5.7, verifica-se que todos os classificadores, exceto o ELM, precisaram aumentar a quantidade de características extraídas para obter resultados próximos aos obtidos pela técnica proposta. Assim, as características propostas permitiram um melhoria do desempenho geral de aproximadamente 7%.

A rede MLP também obteve o melhor desempenho de reconhecimento. Para as duas técnicas, o desempenho de reconhecimento por classe utilizando a MLP é mostrada na Figura 5.10. Conforme mostrado na Figura 5.10, as características propostas permitem uma melhora de reconhecimento de aproximadamente 16% e

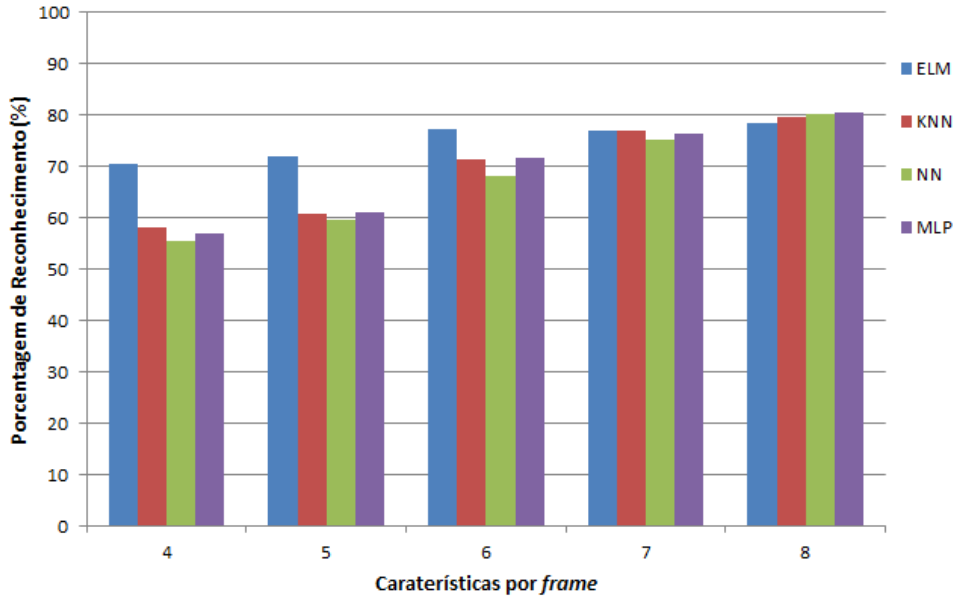


Figura 5.9: Desempenhos médios dos classificadores estudados para a técnica padrão com todas as classes.

13% para as classes de Edema de Reinke e Nódulo Vocal respectivamente, mantendo praticamente estável o desempenho para as classes saudável e disfonia neurológica.

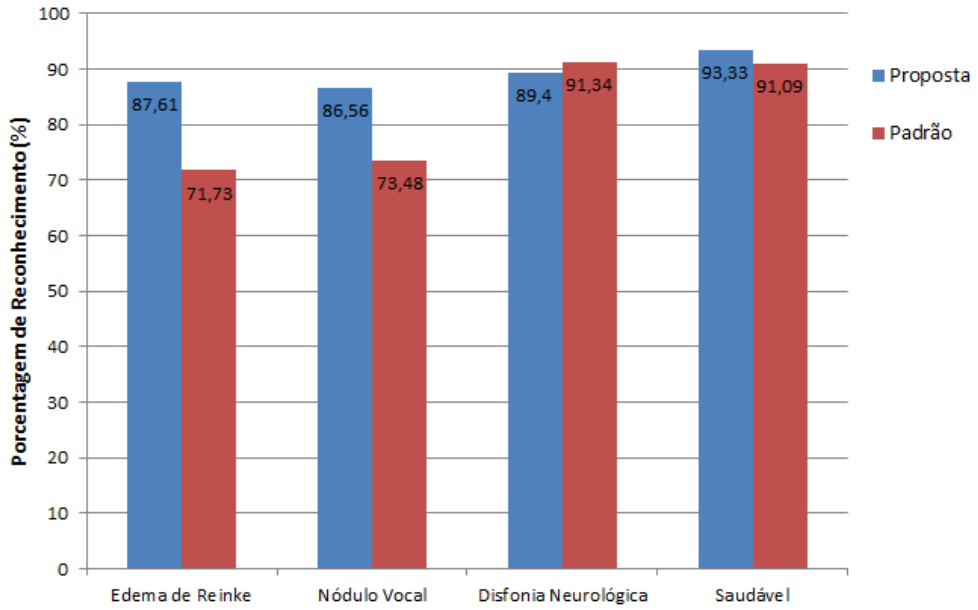


Figura 5.10: Comparação de desempenho de reconhecimento de cada classe para as características propostas usando a rede MLP.

5.6 Resumo do Capítulo

Neste capítulo, foram apresentados os resultados obtidos pela extração de características utilizando a Transformada *Wavelet* (WT). Conforme mostrado na Seção 6.3, as características obtidas dos coeficientes da decomposição *wavelet* permitem classificar entre saudável e não saudável apenas usando a métrica de entropia *wavelet* relativa, sem conseguir determinar um limiar para separar as classes de vozes não saudáveis. Além disso, observou-se que as características extraídas pela técnica proposta obtiveram um desempenho de reconhecimento igual ou maior que a técnica padrão apesar do aumento na quantidade total de características a serem classificadas. No capítulo seguinte são apresentados as considerações finais e as perspectivas de trabalhos futuros.

Capítulo 6

Conclusões e Perspectivas

Este trabalho apresentou um estudo e avaliação da utilização da Transformada *Wavelet* para extrair características de sinais de voz, que permitissem classificá-los em quatro possíveis classes: saudáveis, nódulos, edema de Reinke e disfonia neurológica.

Inicialmente, foram estudados os aspectos relacionados à fisiologia da voz e às patologias da laringe que foram analisadas neste trabalho. Conforme visto, a presença de patologias na laringe altera o sinal de voz e, portanto, pode ser detectada através de uma análise acústica. Em função disso, foram estudados os conceitos básicos da decomposição *wavelet* com o objetivo de extrair características dos sinais de voz saudável e patológico e foi proposto uma nova técnica para extração para sinais de voz, eliminando alguns problemas inerentes a esse tipo de sinal. Além disso, foram discutidas algumas abordagens existentes para o projeto do classificador de padrões, tais como as redes Perceptron Multicamadas (MLP) e *Extreme Learning Machine* (ELM), o Vizinho mais próximo (NN), o k-Vizinhos mais próximo (KNN) e o Naive Bayes.

Os resultados apresentados utilizando a entropia *wavelet* relativa mostraram que é possível definir um limiar de separação entre as classes de edema de Reinke e nódulo vocal e a classe saudável, mas não é suficiente para classificar cada tipo de patologia, permitindo classificar as amostras de voz em saudável e não saudável. Além disso, observou-se que a classe de disfonia neurológica é bastante próxima da classe saudável, sendo, portanto, necessária utilizar classificadores de padrões para separar as classes.

Sem a classe de disfonia neurológica, os resultados das taxas de reconhecimento obtidas com o classificador MLP (82,23% - 93,03%) mostraram que a técnica proposta é bastante eficaz, pois permitiu um ganho de aproximadamente 15% comparada a técnica clássica. Para os demais classificadores, o desempenho de ambas as técnicas foi semelhante, mostrando que para um classificador com alto poder de generalização e especialização como a MLP, as características propostas foram melhores. Além disso, esses resultados obtidos foram equivalentes ou melhores que os apresentados em outros trabalhos (SCALASSARA *et al.*, 2009; SCALASSARA *et al.*, 2007) com o mesmo banco de amostras, mas usando diferentes métodos para extração de características.

Incluindo a classe de disfonia neurológica, ocorre uma diminuição no desempenho de reconhecimento em todos os classificadores em ambas as técnicas. Esse resultado foi esperado em função do resultado mostrado na Seção 5.3, na qual mostrou-se que essa classe é muito próxima da classe de voz saudável. Além disso, os resultados obtidos pelo melhor classificador, a rede MLP, indicaram que as características extraídas pela técnica proposta permitem uma melhora de reconhecimento de aproximadamente 16% e 13% para as classes de Edema de Reinke e Nódulo Vocal, respectivamente, comparando com os resultados da técnica padrão. Com as características propostas, todas as classes obtiveram resultados maiores que 86% usando a rede MLP.

Estes resultados demonstram que a técnica proposta não só permite discriminar vozes saudáveis e patológicas com taxas de acerto acima de 90%, mas também permite classificar as amostras de voz em um tipo de patologia da laringe, incluido a classe de disfonia neurológica, com desempenho superior a 86%. Com isso, conclui-se que técnica proposta utilizando a Transformada *Wavelet* para extrair características é uma ferramenta eficaz e poderosa para ser aplicada no problema estudado.

6.1 Proposta de Trabalhos Futuros

Como proposta de trabalhos futuros sugere-se:

- melhoria na separação das classes usando a entropia *wavelet*;
- avaliar outras características extraídas, como entropia *wavelet*, que é uma medida do grau de ordem ou desordem do sinal, em conjunto com as de energia dos coeficientes *wavelet*;

- avaliar métodos de redução do espaço de características como o Análise por Componentes Principais (PCA) e algoritmo genético, a fim de selecionar as características principais e melhorar o tempo para treinamento das redes neurais;
- avaliar o desempenho de outros classificadores, especialmente as redes neurais não supervisionadas como a SOM (*self-organizing map*), e de um comitê dos melhores classificadores;
- avaliar o desempenho de reconhecimento incluindo amostras de vozes com outras patologias e, com isso, desenvolver uma aplicação para ser testado por médicos, validando a técnica proposta.

Referências Bibliográficas

ABREU, M. H. L. de. Edema de Reinke: aspectos gerais e tratamento. *Monografia de Final de Curso*, CEFAC Pós-Graduação em Saúde e Educação, Rio de Janeiro, Brasil, Dezembro 1999.

ALONSO, J. B. *et al.* Using nonlinear features for voice disorder detection. *ITRW on Non-Linear Speech Processing*, p. 94–106, 2005.

ARAÚJO, R. T. S. *Detecção de Manchas de Óleo na Superfície do Mar em Imagens de Radar de Abertura Sintética*. Dissertação (Mestrado) — Universidade Federal do Ceará, 2004.

BOONE, D.; MCFARLANE, S. *A voz e a terapia vocal*. Porto Alegre, Brasil: Artmed Editora, 2003.

BROOMHEAD, D. S.; LOWE, D. Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems 2*, p. 321–355, 1988.

CARVALHO, R. T. S. Estudo comparativo de técnicas de extração de características para reconhecimento de fonemas. *Monografia de Final de Curso, Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará*, Dezembro 2009.

COSTA, R. C. S. Inspeção Automática de Laranjas Destinadas à Produção de Suco, Utilizando Técnicas de Processamento Digital de Imagens. *Monografia de Final de Curso, Centro Federal de Educação Tecnológica do Ceará*, 2006.

COUREY, M. S. *et al.* Endoscopic vocal fold microflap: a three-year experience. *Ann Otol Rhinol Laryngol*, v. 104, n. 4 Pt 1, p. 267–73, 1995.

COVER, T. M.; THOMAS, J. A. *Elements of Information Theory*. 2nd. ed. [S.l.]: Wiley-Interscience, 2006.

DAUBECHIES, I. *Ten lectures on wavelets*. 1st. ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992. ISBN 0-89871-274-2.

DELLER, J. R. *et al. Discrete-Time Processing of Speech Signals*. [S.l.]: IEEE, 2000.

DUDA, R. O. *et al. Pattern Classification (2nd Edition)*. 2. ed. [S.l.]: Wiley-Interscience, 2001. Hardcover.

FALCÃO, H. H. *et al.* O uso da entropia na discriminação de vozes patológicas. *Salvador, Brasil*, Congresso Brasileiro de Engenharia Biomédica, 2008.

FANT, G. *Speech Sounds and Features*. [S.l.]: Cambridge, MA: MIT Press, 1973.

FAROOQ, O.; DATTA, S. Phoneme recognition using wavelet based features. *Information Sciences*, ELSEVIER SCIENCE INC, v. 150, n. 1-2, p. 5–15, Mar 2003.

FONSECA, E.; PEREIRA, J. Normal versus pathological voice signals. *IEEE Eng. in Medicine and Biology Magazine*, v. 28, n. 5, p. 44–48, september-october 2009.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 3rd. ed. New York: Prentice Hall, 2008.

GREENE, M. *Distúrbios da Voz*. São Paulo, Brasil: Manole, 1989. 134 - 135 p.

HAN, J. *et al. Data Mining: Concepts and Techniques*. 2. ed. [S.l.]: Morgan Kaufmann, 2006. Hardcover.

HAYKIN, S. S. *Redes Neurais: Princípios e Prática*. 2nd. ed. Porto Alegre: Bookman, 2001.

HILLMAN, R. E. *et al.* Phonatory function associated with hyperfunctionally related vocal fold lesions. *Journal of Voice*, v. 4, n. 1, p. 52 – 63, 1990.

HIRANO, M. Structure of the vocal folds in normal and disease states: anatomical and physical studies. *Proc. of the Conf. on the Assessment of Vocal Pathology*, p. 11–30, 1981.

HIRANO, M. Vocal mechanisms in singing: Laryngological and phoniatric aspects,. *Journal of Voice*, v. 2, n. 1, p. 51 – 69, 1988.

HIRANO, M.; BLESS, D. *Exame Videoestroboscópico da Laringe*. Porto Alegre, Brasil: Editora Artes Médicas, 1997.

HOSSEINI, P. *et al.* Pathological voice classification using local discriminant basis and genetic algorithm. In: *Control and Automation, 2008 16th Mediterranean Conference on*. [S.l.: s.n.], 2008. p. 872 –876.

HOTELLING, H. Relations Between Two Sets of Variates. *Biometrika*, Biometrika Trust, v. 28, n. 3/4, p. 321–377, 1936.

HUANG, G. *et al.* Extreme learning machine: Theory and applications. *Neurocomputing*, Elsevier, v. 70, n. 1-3, p. 489–501, 2006.

JAIN, A.; ZONGKER, D. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, p. 153–158, 1997.

KENT, R. D. Vocal tract acoustics. *Journal of Voice*, v. 7, n. 2, p. 97–117, jun. 1993.

KLEINSASSER, O. *Microlaringoscopia e Microcirurgia da Laringe*. São Paulo, Brasil: Manole, 1997.

LASS, N. *Speech and Language: Advances in Basic Research and Practice*. New York: Academic Press, 1979.

MAKHOUL, J. Linear Prediction: A Tutorial Review. *Proceedings of the IEEE*, IEEE-Institute Electrical Electronics Engineers Inc., v. 63, n. 4, p. 561–580, 1975.

MALLAT, S. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 11, n. 7, p. 674 –693, jul 1989.

MARQUES, J. S. *Reconhecimento de padrões: Métodos estatísticos e neuronais*. 2. ed. Lisboa, PT: IST Press, 2005.

MARTINS, R. H. G. *et al.* Vocal fold nodules: Morphological and immunohistochemical investigations. *Journal of Voice*, v. 24, n. 5, p. 531 – 539, 2010.

MELO, D. B. de. Um Sistema de Reconhecimento de Comandos de Voz Utilizando a Rede Neural ELM. *Monografia de Final de Curso, Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará*, Maio 2011.

NEVES, B. M. J. *et al.* Diferenciação histopatológica e imunoistoquímica das alterações epiteliais no nódulo vocal em relação aos pólipos e ao edema de laringe. *Revista Brasileira de Otorrinolaringologia*, Scielo, v. 70, p. 439 – 448, 08 2004. ISSN 0034-7299.

OPPENHEIM, A. V. *et al. Signals and Systems (2nd ed.)*. New Jersey, USA: Prentice-Hall, Inc., 1996.

ORTIZ, K. Z.; CARRILLO, L. Comparação entre as análises auditiva e acústica nas disartrias. *Revista da Sociedade Brasileira de Fonoaudiologia*, Scielo, v. 13, p. 325 – 331, 00 2008. ISSN 1516-8034.

PARRAGA, A. *Aplicação de Transformada Wavelet Packet na análise e classificação de sinais e vozes patológicas*. Dissertação — Universidade Federal do Rio Grande do Sul, Escola de Engenharia, 2002.

RIOUL, O.; VETTERLI, M. Wavelets and signal processing. *Signal Processing Magazine, IEEE*, 1991.

ROSA, M. de O. *et al.* Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Trans. on Biom. Eng.*, v. 47, n. 1, p. 96–104, jan 2000.

ROSSO, O. A. *et al.* Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, Elsevier, v. 105, n. 1, p. 65–75, 2001.

RUSSELL, S. J. *et al. Artificial intelligence: a modern approach*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

SCALASSARA, P. R. *et al.* Relative entropy measures applied to healthy and pathological voice characterization. *Applied Mathematics and Computation*, Elsevier Science Inc, v. 270, n. 1, p. 95 – 108, 2009.

SCALASSARA, P. R. *et al.* Autoregressive decomposition and pole tracking applied to vocal fold nodule signals. *Pattern Recognition Letters*, v. 28, n. 11, p. 1360 – 1367, 2007.

SHANNON, C. E. A mathematical theory of communication. *Bell system technical journal*, v. 27, 1948.

SILVA, I. N. da *et al. Redes Neurais Artificiais para Engenharia e Ciências Aplicadas*. São Paulo, Brasil: Editora Artliber, 2010.

STORY, B. An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology*, J-STAGE, v. 23, n. 4, p. 195–206, 2002.

SULICA, L. *Voice Disorders*. 2009. Disponível em: <http://www.voicemedicine.com/>. Acesso em: 30 de agosto de 2011.

TANG, E. K. *et al.* Linear dimensionality reduction using relevance weighted LDA. *Pattern Recognition*, v. 38, n. 4, p. 485–493, 2005.

ULOZA, V. *et al.* Perceptual and acoustic assessment of voice pathology and the efficacy of endolaryngeal phonomicrosurgery. *Journal of Voice*, v. 19, n. 1, p. 138 – 145, 2005.

ZUNINO, L. *et al.* Wavelet entropy of stochastic processes. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 379, n. 2, p. 12, 2006.

ZWETSCH, I. C. *et al.* Processamento digital de sinais no diagnóstico diferencial de doenças laríngeas benignas. *Scientia Medica*, PUCRS, v. 16, n. 3, Set 2006.