

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

CÉSAR DAVID PAREDES CROVATO

**CLASSIFICAÇÃO DE SINAIS DE VOZ UTILIZANDO A
TRANSFORMADA WAVELET PACKET E REDES NEURAIS
ARTIFICIAIS**

Porto Alegre

2004

CÉSAR DAVID PAREDES CROVATO

**CLASSIFICAÇÃO DE SINAIS DE VOZ UTILIZANDO A
TRANSFORMADA WAVELET PACKET E REDES NEURAIAS
ARTIFICIAIS**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica.

Área de concentração: Automação e Instrumentação Eletro-Eletrônica

ORIENTADOR: Prof. Dr. Adalberto Schuck Jr.

Porto Alegre

2004

CÉSAR DAVID PAREDES CROVATO

**CLASSIFICAÇÃO DE SINAIS DE VOZ UTILIZANDO A
TRANSFORMADA WAVELET PACKET E REDES NEURAIS
ARTIFICIAIS**

Esta dissertação foi julgada adequada para a
obtenção do título de Mestre em Engenharia Elétrica e
aprovada em sua forma final pelo Orientador e pela
Banca Examinadora.

Orientador: _____

Prof. Adalberto Schuck Jr., UFRGS

Doutor pela Universidade Federal do Rio Grande do Sul

Banca Examinadora:

Prof. Paulo Roberto Girardello Franco, PUCRS

Dr. pela University of New Mexico - UNM – 1993

Prof. Paulo Martins Engel, UFRGS

Dr. pela Technical University of Munich, Alemanha - 1986

Prof. Leticia Vieira Guimarães, UFRGS

Dra. pela Muroran Institute of Technology , Japão – 2002

Coordenador do PPGE: _____

Prof. Dr. Carlos Eduardo Pereira

Porto Alegre, Dezembro 2004

DEDICATÓRIA

Dedico este trabalho a minha esposa, Anelise, pelo seu amor e pela sua compreensão e especialmente pelo apoio em todos os momentos difíceis.

RESUMO

Este trabalho apresenta um sistema de classificação de voz disfônica utilizando a Transformada Wavelet Packet (WPT) e o algoritmo *Best Basis* (BBA) como redutor de dimensionalidade e seis Redes Neurais Artificiais (ANN) atuando como um conjunto de sistemas denominados “especialistas”. O banco de vozes utilizado está separado em seis grupos de acordo com as similaridades patológicas (onde o 6^o grupo é o dos pacientes com voz normal). O conjunto de seis ANN foi treinado, com cada rede especializando-se em um determinado grupo. A base de decomposição utilizada na WPT foi a Symlet 5 e a função custo utilizada na *Best Basis Tree* (BBT) gerada com o BBA, foi a entropia de Shannon. Cada ANN é alimentada pelos valores de entropia dos nós da BBT. O sistema apresentou uma taxa de sucesso de 87,5%, 95,31%, 87,5%, 100%, 96,87% e 89,06% para os grupos 1 ao 6 respectivamente, utilizando o método de Validação Cruzada Múltipla (MCV). O poder de generalização foi medido utilizando o método de MCV com a variação *Leave-One-Out* (LOO), obtendo erros em média de 38.52%, apontando a necessidade de aumentar o banco de vozes disponível.

Palavras-Chave: Transformada Wavelet Packet, Rede Neural Artificial, Análise Acústica da Voz, Classificação da Voz Disfônica.

ABSTRACT

This work proposes a dysphonic voice classification system using Wavelet Packet Transform (WPT) and best basis algorithm (BBA) as dimensional reducer and a set of six Artificial Neural Network (ANN) as “specialist” systems. The data bank voices were separated in six different groups according to pathology similarities (where the 6th group is the normal voice patient group) and a set of six ANN was trained, one per group. The decomposition base used in WPT was Symlet 5 and the cost function in best basis tree (BBT) was the Shannon entropy. The ANN were fed by the BBT nodes entropy values. The system presented a success rate of 87,5%, 95,31%, 87,5%, 100%, 96,87% and 89,06% for groups 1 to 6 respectively, using the Multiple Cross Validation (MCV) method. The generalization power was evaluated by the use of MCV method with *Leave-One-Out* (LOO) variation, getting about 38,52% of mean error, showing the necessity to increment the voice database.

Keywords: Wavelet Packet Transform, Artificial Neural Networks, Acoustical Analysis of Voices, Dysphonic voice classification.

SUMÁRIO

1	INTRODUÇÃO.....	12
2	REVISÃO DE LITERATURA.....	15
2.1	INTRODUÇÃO	15
2.2	CLASSIFICAÇÃO DE SINAIS DE VOZES PATOLÓGICAS	15
2.3	APLICAÇÕES DA TRANSFORMADA WAVELET, WAVELET PACKET E REDES NEURAIS ARTIFICIAIS PARA SISTEMAS DE CLASSIFICAÇÃO E PROCESSAMENTO DE SINAIS DE VOZ.....	22
3	FUNDAMENTOS TEÓRICOS.....	39
3.1	INTRODUÇÃO	39
3.2	FISIOLOGIA DA VOZ E PATOLOGIAS	39
3.2.1	Fisiologia Da Voz.....	39
3.2.2	Modelos Acústicos da Produção da Fala e Alguns Aspectos Espectrais	43
3.2.3	Patologias	48
3.3	TRANSFORMADA WAVELET PACKET.....	58
3.3.1	Introdução	58
3.3.2	A Transformada Wavelet Discreta (DWT) e a Análise Multi-Resolução (MRA).....	64
3.3.3	Transformada Wavelet Packet (WPT)	70
3.3.4	Algoritmo <i>Best Basis</i> (BAA) e <i>Best Basis Tree</i> (BBT).....	72
3.4	REDES NEURAIS ARTIFICIAIS.....	76
3.4.1	Introdução	76
3.4.2	Reconhecimento de Padrões	79
3.4.3	Modelo do Neurônio dos Perceptrons Muticamadas (MLP)	80
3.4.4	Algoritmo <i>Backpropagation</i>	83
4	MATERIAIS E MÉTODOS.....	89
4.1	INTRODUÇÃO.....	89
4.2	BANCO DE VOZES PATOLÓGICAS	89
4.2.1	Equipamento utilizado.....	89
4.2.2	Organização do Banco de Vozes.....	91
4.3	APLICAÇÃO DA WPT NAS VOZES PATOLÓGICAS	92
4.4	TESTE DO ALGORITMO BACKPROPAGATION PARA MLP.....	93
4.5	SISTEMA DE REDES NEURAIS INDEPENDENTES POR GRUPO DE DOENÇAS	94
4.5.1	Cálculo das saídas do sistema	106
5	RESULTADOS	107
5.1	INTRODUÇÃO	107
5.2	RESULTADOS DA CLASSIFICAÇÃO UTILIZANDO O ALGORITMO BKP NOS MLP.....	107
5.2.1	Apresentação de Relatório.....	108
5.2.2	Curvas de Aprendizado	109
6	DISCUSSÕES E CONCLUSÕES	118
7	REVISÃO DE LITERATURA.....	121
8	APÊNDICES	128
8.1	APÊNDICE A	128
8.2	APÊNDICE B	130
8.3	APÊNDICE C	133
8.4	APÊNDICE D	138
8.5	APÊNDICE E	140

LISTA DE ILUSTRAÇÕES

Figura 1 Diagrama de blocos do sistema proposto.....	13
Figura 2: Diagrama esquemático que mostra as partes do sistema respiratório e a laringe. (extraído de (ROCHA, 2004)).	40
Figura 3: Partes que compõem a laringe (extraído (ROCHA, 2004)).	41
Figura 4: Vista superior da laringe, mostrando as posições da glote.....	41
Figura 5: Representação das partes do sistema articulatório.....	43
Figura 6: Representação do modelo filtro-fonte para vogais (extraído de (KENT, 1993)).	45
Figura 7: Esquema simplificado do trato vocal. (extraído de (KENT, 1992)).	46
Figura 8: Modos de ressonância dos formantes.....	47
Figura 9: Fotos de nódulo. (Extraído de (GHORAYEB, 2004) e (UPMC, 2004)).	51
Figura 10: Diversas imagens mostrando tipos de pólipos. (Extraído de (GHORAYEB, 2004)).	52
Figura 11: Fotos de cisto. (Extraído de (GHORAYEB, 2004)).	52
Figura 12: Fotos de edema de Reinke. (Extraído de (GHORAYEB, 2004) e (UPMC, 2004))	53
Figura 13: Fotos de laringite. (Extraído de (KAVANAGH, 2004))	54
Figura 14: Foto de câncer. (Extraído de (GHORAYEB, 2004)).	55
Figura 15: Fotos de paralisia. (Extraído de (UPMC, 2004)).	56
Figura 16: Foto de granuloma. (Extraído de (KAVANAGH, 2004)).	57
Figura 17: Espectro das Funções Wavelets	63
Figura 18: Espectro das Funções Escalonamento.....	63
Figura 19: Codificação Subbanda	65
Figura 20: Simbologia da decimação.....	67
Figura 21: Análise com banco de filtros em duas bandas	68
Figura 22: Decomposição em sub-bandas , mostrando os espaços gerados.	68
Figura 23: Espaços de sinais gerados pelas Funções de Escalonamento.	69
Figura 24: Decomposição em sub-bandas WPT, mostrando os espaços gerados.	72
Figura 25: Diagrama de um nó pai e nós filhos.....	74
Figura 26: Ilustração do princípio de seleção <i>Best Basis</i>	75
Figura 27: Modelo de neurônio utilizado. (Adaptado de (WIDROW, 1990)).	80
Figura 28: Representação simplificada do modelo do neurônio	81
Figura 29: Rede neural do tipo Multi-Layer-Perceptron	82
Figura 30: Algoritmo <i>Backpropagation</i> exemplificado. (Adaptado de (WIDROW, 1990)).	84
Figura 31: Diagrama de blocos do sistema proposto.....	94
Figura 32: Bloco de aquisição do sinal de voz.	97
Figura 33: Bloco de extração de características.	97
Figura 34: Bloco de classificação, rede neural “especialista”.....	100
Figura 35: Função de ativação tangente hiperbólica, com valores alvo para evitar saturação.....	102
Figura 36 Conjuntos de Treinamento e Teste, para a configuração E.	104
Figura 37: Curvas de aprendizado. Especialista Grupo 1 - Configuração A e B.	109
Figura 38: Curvas de aprendizado. Especialista Grupo 1 - Configuração C e D.	109
Figura 39: Curvas de aprendizado. Especialista Grupo 1 - Configuração E.	110
Figura 40: Curvas de aprendizado. Especialista Grupo 2 - Configuração A e B.	110
Figura 41: Curvas de aprendizado. Especialista Grupo 2 - Configuração C e D.	110
Figura 42: Curvas de aprendizado. Especialista Grupo 2 - Configuração E.	111

Figura 43: Curvas de aprendizado. Especialista Grupo 3 - Configuração A e B.	111
Figura 44: Curvas de aprendizado. Especialista Grupo 3 - Configuração C e D.	111
Figura 45: Curvas de aprendizado. Especialista Grupo 3 - Configuração E.	112
Figura 46: Curvas de aprendizado. Especialista Grupo 4 - Configuração A e B.	112
Figura 47: Curvas de aprendizado. Especialista Grupo 4 - Configuração C e D.	112
Figura 48: Curvas de aprendizado. Especialista Grupo 4 - Configuração E.	113
Figura 49: Curvas de aprendizado. Especialista Grupo 5 - Configuração A e B.	113
Figura 50: Curvas de aprendizado. Especialista Grupo 5 - Configuração C e D.	113
Figura 51: Curvas de aprendizado. Especialista Grupo 5 - Configuração E.	114
Figura 52: Curvas de aprendizado. Especialista Grupo 6 - Configuração A e B.	114
Figura 53: Curvas de aprendizado. Especialista Grupo 6 - Configuração C e D.	114
Figura 54: Curvas de aprendizado. Especialista Grupo 6 - Configuração E.	115
Figura 55: Curvas de Treinamento. Especialista Grupo 1 e 2 - Final.	115
Figura 56: Curvas de Treinamento. Especialista Grupo 3 e 4 - Final.	115
Figura 57: Curvas de Treinamento. Especialista Grupo 5 e 6 - Final.	116

LISTA DE TABELAS

Tabela 1: Grupos de patologias, conforme (ROSA, 1998).	12
Tabela 2: Principais Sintomas (extraído de (COLTON, 1996)).....	49
Tabela 3: Sinais de distúrbios na voz.....	50
Tabela 4: Grupos de patologias utilizado neste trabalho.....	91
Tabela 5: Grupos de patologias para o Banco de Vozes descrito no Apêndice A.....	92
Tabela 6: Subconjuntos criados a partir do Banco de Vozes.	103
Tabela 7: Configurações de Estimação (Treinamento) e Validação.	104
Tabela 8 Exemplificação dos erros de validação LOO, Grupo 2.....	105
Tabela 9: Erros de Classificação e Taxas de Sucesso.	107
Tabela 10: Pacientes (PACs) classificados como Falsos Negativos (FN) e Falsos Positivos (FP).	107
Tabela 11: Erro devido a Falsos Negativos (FN) e Falsos Positivos (FP).....	108
Tabela 12: Configurações Finais.....	108
Tabela 13: Erros de validação (MAPE) Vs. Número de neurônios ocultos.	116
Tabela 14: Erros de generalização (MAPE) das ANN aplicando o LOO.....	116

LISTA DE ABREVIATURAS

	Significado em Inglês	Sugestão em Português
AC	Acoustic Emission	Emissão Acústica
AMDF	Average Magnitude Difference Function	
ANN	Artificial Neural Network	Redes Neurais Artificiais
AR	Autoregressive	Autoregressivo
ARS	Automatic Response System	Sistema de Resposta Automática
AS	Autocorrelation Sequence	Seqüência de Autocorrelação
BBA	Best Bassis Algorithm	Algoritmo da Melhor Base
BBT	Best Bassis Tree	Árvore da Melhor Base
BKP	Back Propagation Error	Retropropagação do Erro
CART	Classification And Regression Trees	
CEP	Cepstrum	
CVQPM	Cross-Validated Based On Posterior Probabilities	
CWT	Continuous Wavelet Transform	Transformada Wavelet Contínua
DAT	Digital Audio Tape	
DP	Discriminant Pursuit	
DyWT	Dyadic Wavelet Transform	Transformada Wavelet Diádica
DWT	Discrete Wavelet Transform	Transformada Wavelet Discreta
ECG	Eletrocardigraphy	Eletrocardiografia
EEG	Electroencephalography	Eletroencefalografia
EGG	Electroglottography	Eletroglotografia
FCM	Fuzzy Clustering Method	Método de Clusterização Fuzzy
FI	Inverse Filter	Filtro Inverso
FIR	Finite Impulse Response Filter	Filtro de Resposta Impulsiva Finita
FT	Fourier Transform	Transformada de Fourier
FWP	Fast Wavelet Packet	
HMM	Hidden Markov Model	Modelo Oculto de Markov
k-NN1'	k-Nearest Neighbour	

LDA	Linear Discriminant Analysis	Análise Discriminante Linear
LDB	Local Discriminant Basis	
LMS	Least Mean Square	
LPC	Linear Predictive Coding	
MCV	Multiple Cross Validation	Validação Cruzada Múltipla
MES	Myoelectric Signal	Sinais Eletromiográficos
MFCC	Mel-Frequency Cepstral Coefficients	
MLP	Multlayer Perceptron	Perceptron Multicamadas
MUAP	Motor Unit Action Potencial	Potencial de Ação da Unidade
NNE	Normalized Noise Energy	Energia Normalizada de Ruído
PCNN	Parallel Consensual Neural Network	
PDM	Prototype Distribution Map	Protótipos de Mapas de Distribuição
RMS	Root Mean Square	Raiz Média Quadrada
SAR	Synthetic Aperture Radar	
SBC	Sub-Band Base Cepstral	
SIFT	Simple Inverse Filter Tracking	
SNR	Signal Noise Ratio	Relação Sinal/Ruído
SOM	Self-Organizing Map	Mapa Auto Organizado
STFR	Short Time Fourier Transform	
WPT	Wavelet Packet Transform	Transformada Wavelet Packet

1 INTRODUÇÃO

Exames típicos como a Videolaringoscopia e a Videostroboscopia são procedimentos bem estabelecidos de diagnóstico de patologias na laringe. Estes exames são geralmente dispendiosos, podem causar mal estar ao paciente, os resultados podem ser de difícil quantificação, além de precisar perícia do avaliador.

Pelos motivos citados acima, para auxiliar o diagnóstico de doenças na laringe e acompanhar a evolução de tratamentos fonoterápicos, existem diversas técnicas alternativas que utilizam apenas a voz do paciente para extrair índices acústicos quantitativos usados para os fins propostos (BAKEN, 1987). Além destes, alguns métodos de diagnóstico automático de qualidade da voz também têm sido propostos, tais como LPC (filtragem inversa) (CHILDERS, 1987), (FRITZELL, 1992), redes neurais (MARTINEZ, 2000), (RITCHINGS, 2001), (JO, 2001) e medidas acústicas da voz (FROHLICH, 1998).

Devido ao número insuficiente de amostras de voz de algumas patologias para uma classificação, as vozes da base de dados em (ROSA, 1998) foram agrupadas em 6 diferentes grupos baseados na acústica e nas similaridades patológicas da voz. Estes grupos são mostrados na Tabela 1.

Tabela 1: Grupos de patologias, conforme (ROSA, 1998).

Nº	Grupo	Patologias
1	Laringite crônica	Laringite
2	Degenerativa	Carcinoma, granuloma, leucoplasia,
3	Mobilidade Incorreta	Paralisia unilateral, disfonia unilateral e disfonia hipercinética
4	Alterações orgânicas	Edemas de Reinke e edemas generativos
5	Crescimentos orgânicos	Pólipos, nódulos e cistos
6	Normal	Normais

Em (PARRAGA, 2002) é mostrado que um classificador linear entre vozes normais e vozes disfônicas (ou vozes de indivíduos cujas pregas vocais apresentam alguma patologia diagnosticada, daqui por diante também podendo ser denominadas de “vozes patológicas”) pode ser obtido baseado na Transformada Wavelet Packet (WPT) e no algoritmo *Best Basis* (BBA). Este classificador é baseado na presença ou não de certos nós de uma árvore binária

gerada pela WPT e BBA, para cada trecho de sinal de voz. A escolha da função base e função custo que otimizaram o desempenho do classificador foi a Symlet 5 e entropia de Shannon, respectivamente. A taxa de acerto do classificador proposto, para as condições do trabalho, foi de 67,2%.

Em (SCHUCK, 2003) os autores propuseram uma extensão de (PARRAGA, 2002), usando agora uma Rede Neural Artificial (ANN) para a obtenção um classificador não linear para voz normal ou voz disfônica. A rede neural é alimentada pelos valores das entropias dos nós das árvores binárias geradas pela WPT e BBA. Com este novo classificador, a taxa de acerto na classificação aumentou para 84,3%, com o mesmo banco de voz utilizado em (PARRAGA, 2002), usando o método da Validação Cruzada (CV).

Este trabalho propõe o uso da WPT, BBA e ANN para se classificar as vozes do banco de vozes usadas em (PARRAGA, 2002) e (SCHUCK, 2003) em algum dos agrupamentos de patologias sugeridos em (ROSA, 1998). Foi utilizado um sistema de ANN (chamado de “sistema de especialistas”, no contexto deste trabalho) do tipo *Multilayer Perceptron* (MLP) treinado com o algoritmo supervisionado *Backpropagation* (BKP), tendo como entradas os valores de entropia dos nós da BBT. O sistema proposto pode ser sumarizado na Figura 1, onde o primeiro “Bloco 1” e o “Bloco 2” foram implementados nos trabalhos de (PARRAGA, 2002) e (SCHUCK, 2003). O terceiro bloco (“Classificação”) foi desenvolvido e discutido neste trabalho.

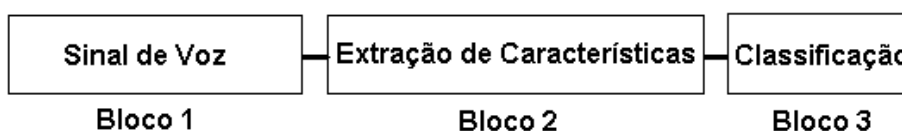


Figura 1 Diagrama de blocos do sistema proposto

O resultado na saída deste sistema indica as “chances” do paciente pertencer a um determinado grupo de patologias (isto, é claro, na “opinião” da rede neural treinada), podendo assim auxiliar ao diagnóstico. O processo de classificação se deu em duas abordagens, a primeira apresentada em (CROVATO, 2004), utiliza o método da Validação Cruzada (CV), visa classificar o banco de vozes e avaliar a taxa de acerto obtida. Neste trabalho, utilizou-se a segunda abordagem, a Validação Cruzada Múltipla (MCV), que propõe avaliar estatisticamente o modelo mais apropriado de cada ANN, já que o número de indivíduos é escasso. Os resultados de cada abordagem foram comparados e comentados. Para avaliar o

poder de generalização do sistema, após a determinação do número de neurônios ótimo, aplicou-se se o método de treinamento conhecido como *leave-one-out* (LOO) nas configurações finais encontradas.

O trabalho está organizado como segue.

No capítulo 2 são apresentadas as diferentes técnicas de classificação e análise de patologias na laringe. Também são apresentadas diversas aplicações da WT e WPT assim como de ANN em sistemas de classificação e processamento de sinais de voz.

No capítulo 3 encontram-se os fundamentos teóricos deste trabalho. É apresentada a fisiologia da voz, o modelo filtro-fonte e um breve resumo das principais patologias na laringe. Também é abordada a teoria da Transformada Wavelet Packet e algoritmo *Best Basis*. E finalmente uma abordagem sobre as Redes Neurais Artificiais, dando ênfase as redes MLP, utilizadas neste trabalho.

No capítulo 4 encontram-se descritos os materiais e métodos que foram utilizados neste trabalho. Que incluem validação de software, determinação de número de neurônios ocultos ótimo de cada “especialista”, determinação do poder de generalização do sistema.

No capítulo 5 são mostrados os resultados obtidos aplicando a WPT e o sistema de ANN nos sinais do banco de vozes. Os erros de classificação, falsos negativos e falsos positivos. Também são mostrados os erros de generalização com o método de validação cruzada múltipla com a variação *Leave-One-Out* (LOO).

Finalmente, no capítulo 6 são discutidos os resultados obtidos. Tem-se como principal conclusão que o sistema classificador tem performance maior que os classificadores dos trabalhos anteriores desta linha de pesquisa, no que se refere à taxa de sucesso; porém apresenta ainda um erro de generalização elevado, mostrando que o sistema tem potencial de ser utilizado em forma de produto, mas são necessários mais pacientes para um melhor treinamento e generalização das ANN que o compõem. São também apresentadas outras sugestões para melhorar o desempenho nos trabalhos futuros desta linha de pesquisa.

2 REVISÃO DE LITERATURA

2.1 INTRODUÇÃO

A revisão de literatura está dividida em duas partes. Na Seção 2.2, serão apresentados os vários métodos para classificação dos sinais de vozes “patológicas” utilizados ao longo dos anos. Na Seção 2.3 serão apresentadas diversas aplicações da Transformada Wavelet, incluindo aplicações onde operam em conjunto com Redes Neurais Artificiais, inclusive na classificação de vozes “patológicas”.

2.2 CLASSIFICAÇÃO DE SINAIS DE VOZES PATOLÓGICAS

No artigo (RONTAL, 1975), os autores apresentaram o uso do espectrograma para a avaliação da situação clínica da reabilitação vocal e do tratamento cirúrgico e médico de uma variedade de doenças das cordas vocais. Artigos anteriores discutiram o uso do espectrograma no estudo de doenças da laringe, concluindo que os padrões produzidos podem ser aplicados a diversas situações clínicas. Este artigo apresenta diferentes espectrogramas de vozes normais e com patologias, analisando-as e apontando diferenças. Uma das conclusões mais importantes deste artigo é a respeito do espectrograma da voz. Primeiramente, segundo os autores, não é possível se obter uma medida exata ou quantificar um espectrograma da voz. Mudanças nos padrões do espectrograma da voz são percebidas qualitativamente. Os espectrogramas não apresentam diferenças entre tipos de lesões de cordas vocais, pois muitas patologias apresentam características acústicas similares. Os autores concluem que uma avaliação final das condições das cordas vocais deve ser feita por todos os meios possíveis, e não baseado apenas em um único método.

Os autores do artigo (DELLER, 1980) apresentaram uma técnica de classificação automática de disfunções da laringe, utilizando os pólos e zeros de um filtro inverso (FI) digital obtido a partir da modelagem autoregressiva do sinal de voz, sendo a informação para a classificação está contida na posição destes zeros. Um vetor padrão de características é processado por um procedimento de agrupamento (*clustering*) automático para classificar

casos. Segundo os autores, dois procedimentos de análise são possíveis na população de sinais de voz. Primeiro a classificação entre casos, auxiliando no diagnóstico, e segundo, a avaliação dentro de um mesmo caso, onde se tem um auxílio na avaliação da evolução da terapia. Nas simulações, foram utilizados dois tipos de sinais sintéticos. Os primeiros foram subdivididos em duas categorias de anomalia:

- a) Casos de performance de filtro (como *Jitter*, *Shimmer*, entre outros);
- b) Casos de inserção de ruído nos sinais.

Em geral, os resultados mostraram que a análise automática dos padrões pode ser um discriminante efetivo de diferentes aberrações acústicas. Os autores concluíram que este método demonstrou significativa habilidade para identificar simulações de anomalias na laringe. Porém, eles ressaltaram que, como foram utilizados sinais sintéticos e que devido à enorme variabilidade inerente a sinais patológicos reais, pesquisas adicionais são necessárias para aplicar esta técnica clinicamente.

Em (GUOXIN, 1989), os autores utilizaram um sistema de análise espectral de sinais de voz de pacientes com doenças na laringe. Estudos mostraram que quando alguma desordem ocorre na laringe, a voz do paciente muda sua forma de onda e seu espectro de potência. Os resultados deste artigo mostraram que este método é capaz de reconhecer características acústicas de vozes disfônicas. Os autores concluíram que este método de análise espectral é um método efetivo para auxiliar médicos na pesquisa da fisiologia e da patologia da voz, e também para monitorar o processo do tratamento. Os resultados foram qualitativos e aparecem em gráficos comparativos.

Em (CHILDERS, 1987) os autores desenvolveram uma medida quantitativa para a avaliação da função da laringe, usando a fala e dados de eletroglotografias (EGG). Foram desenvolvidos dois procedimentos para a detecção de patologias da laringe:

- a) Medida de distorção espectral usando métodos de *Pitch* com *Linear Predicting Code* (LPC) e quantização vetorial;

b) Análise do sinal de EGG, medindo a diferença do intervalo de tempo e amplitude.

Os autores utilizaram vinte e três pacientes com patologias e cinquenta e duas vozes normais. A vogal utilizada foi o /i/ sustentado. O melhor resultado adquirido com o primeiro método foi de 75,9% (22 corretos/29 sujeitos) de detecção correta de patologias. Já o segundo método apresentou 69% de detecção correta (20 corretos/29 sujeitos). A probabilidade de falsos positivos foi de 9,6% para sujeitos normais.

Os autores em (ACCARDO, 1998), desenvolveram um algoritmo baseado nas técnicas de combinação de padrões acústicos, para obter um discriminante automático de vozes normais e com a patologia ataxia de Friedreiche. Ataxia de Friedreiche é uma doença degenerativa do sistema nervoso central, que é caracterizada pela desordem dos movimentos da articulação, hipotonia e tremor. A técnica proposta utiliza parâmetros de dimensão fractal de tempo-curto. Para comparação, os autores utilizaram os parâmetros de cruzamentos por zero (*zero-crossing*) e taxa de energia (*energy rate*), que são estimados na tarefa de classificação. Foram utilizados oito pacientes com ataxia de Friedreiche e nove pessoas sem patologias na laringe ou disfunções neurológicas. Os sujeitos falaram três frases diferentes. Os resultados mostraram que todos os parâmetros são capazes de diferenciar os dois grupos, porém o parâmetro de dimensão fractal fornece uma classificação com maior índice de acerto. O valor médio de classificação para os parâmetros fractal, taxa de energia e cruzamento por zero são, respectivamente, 96,1%, 92,1% e 94,1%. Com isso, os autores concluem que o método pode ser aplicado para classificação automática de vozes, normais *versus* patológicos.

Em (FROHLICH, 1998), os autores propuseram uma medida acústica para soprosideade com o objetivo de fazer a distinção entre diferentes condições de fonações fisiológicas em casos de vozes patológicas. O índice proposto é o *glotal to noise excitation ratio* (GNE), e este é comparado a outros sete índices descritos na literatura. A avaliação deste método é feita com base na capacidade de distinguir entre diferentes mecanismos bem definidos de fonação patológica. Para os testes, foi utilizado um banco de 454 vozes, entre patológicas e normais, como grupo de referência, e seis vogais diferentes isoladas. Os

resultados mostraram que o GNE, entre todas as medidas feitas, apresentou a mais significativa distinção entre os diferentes grupos de mecanismos de fonação patológica.

Os autores em (ROSA, 1998) propuseram um método quantitativo de classificação de patologias da voz, de forma a padronizar o diagnóstico entre especialistas da fala. O objetivo deste trabalho é utilizar redes neurais para classificar padrões de patologias através de medidas acústicas obtidas do resíduo do sinal (filtragem inversa). Foram utilizados 119 falantes entre normais e patológicos. Os vinte e dois tipos de patologias foram reagrupados para produzir apenas seis grupos baseados na acústica e nas similaridades patológicas da voz. Este procedimento foi justificado em experimentos prévios pelos autores, onde utilizaram todas as disfonias e obtiveram uma grande variabilidade nos resultados através da rede neural. Foram aplicados métodos estatísticos nas medidas acústicas, e estes foram os vetores de entrada da rede neural para identificar doenças da laringe. Os autores testaram diversas combinações entre as medidas acústicas e os métodos estatísticos. A média de acerto de classificação dos grupos de patologias foi de 62,33%.

Os autores do artigo (ROSA, 1999) utilizaram técnicas de processamento de sinais, como Cepstrum (CEP) e *Least Mean Square* (LMS), para extrair características acústicas da voz, e estas foram utilizadas em redes neurais para determinar as condições globais relacionadas à patologia da laringe. Diversos índices acústicos foram calculados, como *Jitter* e *Shimmer*. Foram combinadas vinte e duas patologias duas a duas, e um segundo indicador de habilidade de identificação foi o nível de separação entre os grupos (*clusters*) formados entre os pares de patologias.

As medidas acústicas foram fornecidas à entrada da ANN. Cada neurônio de saída representa uma condição das cordas vocais (agrupadas em 6 grupos diferentes). A ANN foi treinada minimizando a entropia relativa do conjunto de treinamento. Nos resultados experimentais, diversas combinações de índices produziram diferentes percentuais de acerto ao processo de classificação.

Algumas conclusões dos autores são: medidas de *Jitter* e *Shimmer* para calcular movimentos anormais das cordas vocais produzem identificações mais confiáveis nas redes

neurais; O melhor resultado obtido foi de 82,22% de identificação das condições da glote; e por fim, a discriminação entre vozes patológicas e normais foi alcançada, conforme as próprias palavras dos autores “perto de 100%”.

Em (HADJITODOROV, 2000), os autores propuseram um método de classificação de vozes normais e de vozes “patológicas”. O método proposto para aumentar a precisão de detecção de patologia da laringe é baseado na modelagem da função densidade de probabilidade (fdp) dos falantes normais e patológicos por meio de protótipos de mapas de distribuição (PDM – *Prototype Distribution Map*), desenvolvidos pelos autores. Para os testes foram utilizadas 400 vozes, 100 normais e 300 pacientes com patologias na laringe. Os falantes pronunciaram a vogal sustentada /a/. Para comparação, os autores utilizaram métodos clássicos de reconhecimento, como análise discriminante linear (LDA), o método *k-nearest neighbour* (k-NN) e o *self-organizing map* (SOM). A precisão na classificação do método proposto foi de 95,1%, enquanto que para LDA foi de 69,95%, para o k-NN foi de 90,95% e para o SOM 90,46%.

Os autores em (MARTINEZ, 2000), implementaram um sistema automático de detecção de patologias da laringe utilizando a análise acústica da fala no domínio frequência. Diferentes técnicas de processamento são aplicadas: cepstrum, mel-cepstrum, delta cepstrum, delta mel-cepstrum e FFT. Os sinais processados foram utilizados como parâmetros de entrada de uma rede neural. Eles afirmaram que a rede neural é um excelente sistema de classificação e especialista em dados com ruído, incompletos, sobrepostos, etc...Neste trabalho, o sistema é treinado para distinguir entre vozes normais e patológicas (duas classes), e para classificar vozes normais, diplofonia e rouca (três classes). Foi utilizada a vogal sustentada /a/ e as vozes patológicas foram diagnosticadas previamente (vinte e quatro amostras de patologias). Para a extração de características, uma janela hamming foi aplicada a cada segmento e então os padrões foram extraídos dos dezesseis primeiros coeficientes cepstrais. Os outros processamentos foram cepstrum em escala mel com dezesseis coeficientes, delta cepstrum e delta mel-cepstrum com trinta e dois coeficientes, e a FFT com cento e vinte oito coeficientes. Os resultados experimentais mostraram que a análise cepstral obteve a melhor performance (91%) para separar as duas classes e 81,91% para as três classes. Segundo os autores, a técnica que teve o pior desempenho foi a FFT devido à

dimensão dos padrões que dificultou o treinamento da rede. Os autores concluem que é possível utilizar esta ferramenta para a análise quantitativa no processo de diagnóstico, assim como acompanhar a evolução da terapia de reabilitação.

No artigo (ROSA, 2000) os autores propuseram uma técnica não invasiva para auxiliar especialistas da voz a fornecer um diagnóstico preciso de doenças da laringe e das cordas vocais. A técnica utilizada é a filtragem inversa do sinal de voz, utilizando filtros Kalman e Wiener. Os autores utilizaram três vogais /a/, /e/ e /i/. A frequência de amostragem do sinal foi de 22050 Hz e o nível de quantização de 16 bits. A técnica de filtragem inversa descrita na literatura, é desenvolvida no segmento mais estacionário do sinal de fala. Porém, muitas patologias não possuem características de ‘quasi-estacionariedade’. Para superar este problema, os autores propuseram três algoritmos adaptativos para estimação do modelo autoregressivo (AR). Foram utilizadas sete medidas acústicas para diferenciar entre falantes normais e patológicos. Nos resultados, as vogais /a/ e /e/ foram as que obtiveram os melhores resultados. Das diversas conclusões relatadas pelos autores, uma delas afirma que a combinação destes filtros permite uma ótima discriminação de doenças com características similares e que a melhor característica (medida acústica) para a discriminação patológica foi *Jitter* com 54,79%.

No artigo (JO, 2001), os autores descreveram um método de classificação de vozes patológicas ou normais baseados em um sistema de resposta automática (ARS –*Automatic Response System*). As vozes são classificadas utilizando uma rede neural cujos parâmetros de entrada são características acústicas como *Jitter*, *Shimmer*, razão harmônicas/ruído (NHR), entre outros. O objetivo principal deste trabalho é poder classificar vozes, patológicas ou não, utilizando a rede de telefonia pública, de forma que mais doutores não especialistas no assunto e pacientes possam diagnosticar suas vozes sem ir ao hospital (a idéia não é emitir o diagnóstico exato nem o tratamento). A taxa de classificação é comparada à taxa de classificação obtida utilizando vozes que foram gravadas diretamente em *Digital Audio Tape* (DAT). Foram utilizados 76 casos de patologias e 35 casos de vozes normais para treinar a rede neural. As taxas de classificação do ARS ficaram entre 88% e 91%. Os autores concluem que não foi observada diferença significativa entre ARS e DAT em termos de taxa de classificação. Baseado nisso, é possível separar patológicos de normais com o sistema ARS e

pode ser aplicado a um sistema automático de diagnóstico. Porém, o número total de vozes não é ainda suficiente para generalizar o desempenho, necessitando mais dados para isso.

Em (NIEDZIELSKA, 2001), os autores utilizaram testes acústicos em sinais de voz de crianças (4-14 anos) com nódulos vocálicos. Estes testes foram feitos antes e depois do tratamento. Os parâmetros acústicos da voz foram comparados com um grupo controle de crianças sem patologia. Os resultados da investigação foram analisados acusticamente. Foi mostrado que os valores de *Jitter*, *Shimmer*, frequência fundamental e a relação sinal-ruído mudaram significativamente entre as crianças com nódulos vocálicos e sem patologias na laringe. Estes parâmetros tendem a se aproximar dos valores normais durante o tratamento. A conclusão dos autores é que o método acima permite a detecção de desordens do funcionamento da laringe, podendo ser útil no processo de avaliação do tratamento. No artigo são mostradas diversas tabelas comparativas destes índices ao longo do tratamento.

Em (RITCHINGS, 2001) os autores descreveram o protótipo de um sistema para a avaliação objetiva da qualidade da voz em pacientes em recuperação de vários estágios de câncer na laringe. Foi utilizada a vogal sustentada /i/, e esta foi avaliada subjetivamente por especialistas da fala e da língua de acordo com sete níveis de qualidade (0 = menos disfônica e 6 = mais disfônica). O sistema extrai os principais parâmetros, através dos sinais da EGG, e estes são usados para treinar e testar ANN. O sistema utilizado para capturar os sinais da impedância elétrica consiste de dois eletrodos, um em cada lado do pescoço. Setenta e sete vozes anormais foram utilizadas para treinar e testar uma rede neural do tipo MLP com sete saídas. O algoritmo de treinamento utilizado foi o *backpropagation*. A função de erro utilizada foi a entropia cruzada. Os melhores resultados foram obtidos mediante uma combinação de parâmetros do sinal, alcançando um índice de acerto de 92%. Os autores concluem que um sistema de rede neural pode ser treinado para fornecer uma classificação objetiva da qualidade da voz.

2.3 APLICAÇÕES DA TRANSFORMADA WAVELET, WAVELET PACKET E REDES NEURAIS ARTIFICIAIS PARA SISTEMAS DE CLASSIFICAÇÃO E PROCESSAMENTO DE SINAIS DE VOZ

Mallat, em (MALLAT, 1989), descreveu o modelo matemático para o cálculo e interpretação do conceito de representação multi-resolução em imagens. Esta representação, chamada representação Wavelet, decompõe o sinal original usando bases Wavelets ortogonais, o que pode ser interpretado como uma decomposição do sinal utilizando um conjunto de filtros passa-banda independentes. O cálculo é eficiente, pois o algoritmo é baseado na convolução do sinal com filtros digitais *Quadrature Mirror Filters* (QMF), e o método é também designado de *Fast Wavelet Transform* (FWT).

Kadambe e Boudreaux-Bartels em (KADAMBE, 1992) descreveram a utilização da Transformada Wavelet Diádica (DyWT) como um detector de *Pitch*. Os autores compararam a utilização da DyWT com dois métodos clássicos de detecção de *Pitch*. O primeiro é o método de autocovariância. Este estima precisamente o *Pitch* para algumas vogais, porém tem a desvantagem de não ser adequado para todas as vogais e para períodos de *Pitch* não-estacionários, além de ser computacionalmente complexo. O outro método é o uso do cepstrum como detector de *Pitch*. Neste caso, é assumido que o sinal é estacionário e que no segmento analisado existe no mínimo dois períodos de *Pitch* completos. Este método não é apropriado para falantes que possuam o período de *Pitch* muito baixo ou muito elevado. Nos testes foram utilizados sinais sintéticos com períodos de *Pitch* estacionários e não-estacionários, sendo comparada a performance dos dois métodos clássicos mencionados anteriormente e a DyWT. Para comparar a robustez ao ruído, os autores adicionaram ruído branco ao sinal sintético vogal /i/. Por fim, os autores utilizam dados reais da fala humana. A partir dos resultados, os autores concluíram que o detector de *Pitch*, usando DyWT, comparado aos métodos clássicos, exibiu uma performance superior. As principais vantagens deste método, segundo os autores, são:

- a) Não assume estacionariedade dentro da janela de análise;
- b) Estima o período *Pitch* precisamente, com erro relativo máximo de 2% para $SNR \geq -18dB$;

- c) É apropriado para uma maior variação de períodos de *Pitch*;
- d) É computacionalmente simples;
- e) Exibe resultados superiores aos métodos clássicos de detecção de *Pitch*.

Em (KADAMBE,1999) é apresentado um detector de complexo QRS baseado na Transformada Wavelet Dyadica (DyWT). Este detector é robusto à morfologia do sinal e ao ruído. Foi projetada uma *Spline Wavelet* apropriada para a detecção do QRS. As escalas desta wavelet são escolhidas em função das características espectrais do eletrocardiograma (ECG). Para ilustrar a performance do detector, foram utilizados os ECG da base de dados do *American Heart Association* (AHA). A performance obtida foi comparada com a de outros algoritmos conhecidos: Okada, Hamilton-Tompkins e *Multiplication of the Backward Difference* (MOBD). Todos estes algoritmos foram aplicados em ECG de indivíduos com diferentes características: contrações ventriculares prematuras (PVC), PVC simples e uniforme, PVC multiforme isolada, batimentos R e T, acoplamentos, ritmos ventriculares. Todos os algoritmos se mostraram deficientes na identificação do complexo QRS em pelo menos um caso, exceto o algoritmo de Kadambe, que mostrou uma taxa de acerto maior que 40% em todos os casos podendo atingir até 80% de acerto.

Coifman e Wickerhauser em (COIFMAN, 1992), propuseram um método eficiente de compressão de sinais de som e imagem. Este método utiliza uma biblioteca de bases ortogonais pré-definidas, denominadas Wavelet Packets. O propósito dos autores é construir uma biblioteca de funções na qual a representação de um determinado sinal nesta base tenha o menor custo de informação, através do cálculo de uma função custo. Este método está baseado na propriedade de ortogonalidade das novas bibliotecas construídas. A decomposição de sinais nestas bases forma uma estrutura de árvore binária, e a escolha da melhor representação do sinal é chamado algoritmo *Best Basis*. Diversas funções custo são possíveis no algoritmo *Best Basis*, sendo que a mais atraente, segundo os autores, é a entropia de Shannon, uma medida de concentração ou eficiência da expansão do sinal na biblioteca de funções.

No artigo (RAMCHANDRAN, 1993), Ramchandran e Vetterli propuseram um algoritmo para encontrar a melhor base (*Best Basis*) de forma mais rápida que o algoritmo de (COIFMAN, 1992). Este algoritmo utiliza a WPT para compressão de sinais, empregando uma estrutura que inclui taxa de distorção. O algoritmo é desenvolvido para “podar” as árvores completas da WPT, na qual uma sub-árvore é a melhor base segundo a medida da distorção global dado um orçamento de bits para codificação, ou então a codificação da taxa de bits para uma determinada qualidade.

Os autores descreveram a compressão em 2 imagens, utilizando o ambiente de codificação JPEG e a WPT, esta usando o conjunto de filtros Wavelet Daubechies D4. As aplicações usando a WPT apresentaram resultados favoráveis em relação ao JPEG.

Em (SAITO, 1994), Saito e Coifman descreveram uma extensão do método *Best Basis* (COIFMAN, 1992). Este método seleciona uma base ortogonal que maximiza a separação entre classes para o problema de classificação de sinais. O algoritmo reduz a dimensão do sinal, utilizando como extrator de características funções de base que são bem localizadas no plano tempo-frequência. Este método foi testado em dois conjuntos de sinais de voz sintéticos onde são extraídas características (coeficientes da expansão dos sinais nas funções bases). Estas características são fornecidas aos classificadores, sendo então calculada a taxa de erro de classificação. Os classificadores utilizados foram o *Linear Discriminant Analysis* (LDA) e o *Classification And Regression Trees* (CART). Esta extensão foi chamada de *Local Discriminant Basis* (LDB). Os exemplos mostraram que o método proposto é superior a aplicação direta dos sinais como entradas de tais classificadores.

Carl Taswell em (TASWELL, 1996) propôs uma técnica para a procura da *Best Basis* da WPT diferente do algoritmo apresentado por Coifman e Wickerhauser (COIFMAN, 1992). Esta técnica faz uma procura na estrutura de árvore da WPT de cima para baixo (ao contrário do algoritmo original), utilizando funções de custo aditivas e também não-aditivas. O algoritmo proposto foi chamado de “quase” *Best Basis* (*near Best Basis*), pois a performance é parecida com o *Best Basis*, porém com menos esforço computacional. Os testes experimentais consistiram da comparação de várias funções custo e métodos de seleção de

base para compressão de dados de fala e análise tempo-frequência de transientes artificiais. Os resultados apresentaram uma melhora em relação ao algoritmo *Best Basis* original.

Em (WU, 1996), os autores descreveram o uso da Wavelet Packet (WP) para o monitoramento de máquinas. Foi proposto um método de extração de características baseado na energia de cada nó da árvore da decomposição Wavelet. Devido ao fato das Wavelets representarem sinais filtrados em diferentes faixas, cada decomposição possui determinada informação do sinal em diferentes células tempo-frequência. Aparentemente, segundo os autores, nem todas as decomposições Wavelets contém informação relevante. Desta forma, o sinal é reconstruído eliminando ou tornando zero as decomposições em Wavelets Packets que contenham pouca informação, assim, o processo de extração de característica é chamado de compressão já que o sinal selecionado contém menos informação do que o sinal original, porém as principais características do sinal são preservadas.

Para medir a efetividade das WP, quatro critérios foram propostos. O método proposto é testado em dois exemplos práticos:

- a) Monitoramento da trepidação em torneamento;
- b) Desgaste da ferramenta ao perfurar.

Nas aplicações práticas deste método, o limiar de monitoramento foi obtido experimentalmente resultando em uma taxa de 100% de acerto para os dois exemplos práticos apresentados.

No artigo (BENEDIKTSSON, 1997), os autores utilizaram uma nova arquitetura de rede neural, chamada de *parallel consensual neural network* (PCNN) para a classificação de sinais de múltiplas fontes de sensoramento remoto e dados geográficos. Os sinais de entrada da rede neural sofrem diversas transformações utilizando a Transformada Wavelet Packet, e cada transformação é utilizada como um vetor de entrada independente na rede neural. Os experimentos relatados afirmam que o sistema proposto supera os existentes (métodos estatísticos), aumentando a exatidão na classificação dos dados de teste.

Em (INTRATOR, 1997) são discutidos métodos de extração de características baseados na decomposição em Wavelets, com o propósito discriminação de sinais. Os autores fazem referência ao uso dos coeficientes da DWT e da WPT como entradas de classificadores lineares e não lineares. No artigo é mostrada a superioridade dos classificadores não lineares na classificação de sinais de som de porcos do mar e cachalotes. Uma rede de neurônios BCM (Bienen-stock, Cooper e Munro) foi utilizada como classificador não linear, tendo como entradas os coeficientes da DWT, obtendo uma taxa de acerto de 99% e 76% para porcos do mar e cachalotes respectivamente. Os neurônios BCM são utilizados para extrair características de espaços de alta dimensionalidade (como o caso do espaço definido pelos coeficientes da DWT).

Outra aplicação dos autores é na modelagem de sistemas análise de vibração de aviões que requer medidas de frequência e amortecimento de modos de vibração crítica como uma função da velocidade do voo. Para reduzir a complexidade da ANN, apenas os maiores coeficientes da DWT são selecionados como entradas da Rede Neural. Desta forma, o modelo da Rede se torna mais eficiente e robusto. Os erros relativos para a predição dos coeficientes de amortecimento e para a predição de frequência foram, respectivamente, 5% e 3%.

Em (PAYA, 1997) os autores utilizaram a combinação Transformada Wavelet e Rede Neural para classificar seis diferentes falhas que ocorrem em máquinas de rotação. A Transformada Wavelet é usada como um pré-processador para a rede neural, tendo como função Wavelet a Daubechies 4. Esta combinação teve sucesso em distinguir os diferentes tipos de falhas que ocorreram no modelo mecânico (*drive-line*) descrito no artigo. Este modelo foi testado em condições normais, depois com falhas inseridas individualmente na caixa de câmbio do motor, e por fim, com todas as falhas em conjunto. A taxa de classificação foi de 96% no conjunto de teste. E, segundo os autores, o uso desta técnica para o monitoramento contínuo de máquinas é eficiente e confiável, sem a necessidade de desmontar a máquina.

No artigo (SOSTARIC, 1997), os autores utilizaram a Transformada Wavelet Packet como uma ferramenta para extração de características para fins de classificação de sinais

eletromiográficos. Para a seleção da *Best Basis* foi usado o algoritmo desenvolvido por Saito e Coifman (SAITO, 1994), chamado de *local discriminant basis* (LDB), que utiliza a entropia relativa, ou seja, a entropia de duas classes conhecidas. Os autores aplicaram a técnica em sinais eletromiográficos. Estes sinais representam a atividade elétrica dos músculos. O objetivo é extrair características importantes de diferentes tipos de potencial de ação da unidade motora (MUAP – *Motor Unit Action Potential*). Para classificar sete tipos (classes) de MUAP, foram utilizados os seguintes classificadores:

- a) Linear Discriminant Analysis (LDA);
- b) Classification and Regression Trees (CART);
- c) k-Nearest Neighbour (k-NN);
- d) Artificial Neural Network (ANN).

Os autores não fizeram comparações quantitativas com diferentes extrações de características, mas concluíram através da interpretação visual que é possível separar as classes utilizando este algoritmo.

Os autores do artigo (MALLET, 1997) propuseram um algoritmo para a classificação de dados espectrais minerais, através da Transformada Wavelet Discreta (DWT). A DWT fornece informação localizada de frequência importante para a classificação, conservando informação discriminante e reduzindo a dimensão do sinal. O problema, segundo os autores, é escolher qual função Wavelet usar para uma aplicação em particular e qual apresenta o melhor resultado. Os autores propuseram um algoritmo que escolhe a Wavelet de forma adaptativa; e o critério de escolha é baseado na probabilidade *a posteriori* chamado *cross-validated based on posterior probabilities*. Os autores aplicaram o algoritmo em um conjunto de dados simulados e em sinais mineralógicos. Nos dois exemplos apresentados, o classificador utilizado foi o Bayesiano linear. Os resultados da classificação utilizando o método proposto foram comparados as Wavelets Daubechies e Coiflet, e os índices da classificação do grupo de teste foram 97%, 99% e 93%, respectivamente, mas é mostrado que probabilisticamente é mais fácil classificar sinais submetidos ao método proposto.

Em (VAN De WOUWER, 1997) é descrito um sistema de classificação de sinais de voz disfônicas, baseado em Transformada Wavelet 2D de espectrogramas da voz (o espectrograma é uma imagem que é a entrada da WT-2D). Uma rede do tipo LVQ é utilizada para extrair as características do espaço de entrada. A saída da rede é interpretada por um outro sistema fuzzy. Os sinais de voz são classificados pela escala GRABS (General evaluation (G), roughness (R), asthenicity (A), breathiness (B), Strain (S)). Foi utilizada uma base de dados composta por 185 mulheres e 160 homens diagnosticados na escala GRABS. Os resultados da classificação (taxa de acerto) são da ordem de 60% para mulheres e 64% para homens.

Um exemplo de aplicações da WPT em sistemas flexíveis de manufatura encontra-se no artigo (XIAOLI, 1998). Os autores desenvolveram um dispositivo para detectar sinais de emissão acústica (AE - *Acoustic Emission*) de ferramentas de rotação. É apresentado um método de monitoramento das condições de desgaste das ferramentas utilizando a WPT como um pré-processador dos sinais AE, seguido do método de clusterização Fuzzy (FCM - *Fuzzy Clustering Method*), com o objetivo de associar as características geradas na transformada à decisões apropriadas. Segundo os autores, dos métodos que monitoram processos de corte indiretamente, AE é a maneira mais efetiva de perceber o desgaste da ferramenta. Porém, estes sinais devem sofrer um tratamento de sinais para extrair o máximo de informação útil. A WPT é usada para decompor o sinal em diferentes bandas de frequência no domínio tempo, e a raiz média quadrada (RMS - *Root Mean Square*) do sinal decomposto em cada banda é usada como característica para a classificação das condições do sistema. A WPT pode capturar importantes características do sinal do sensor, sensível às mudanças das condições do processo, mas insensível às variações das condições de trabalho do processo e a diversos ruídos.

Nos experimentos realizados, o estado das ferramentas que estão sendo monitoradas foi classificado em 'A', 'B', 'C' e 'D', sendo que na classificação 'D' é sugerida a troca da peça. A média de classificação correta utilizando o método integrado com WPT e FCM foi de 90,25% para o conjunto de teste.

Os autores do artigo (CHANG, 1998) descreveram uma nova técnica de extração de características de sinais de voz, denominada por eles de Wavelet Adaptativa Para Reconhecimento de Fala. A energia de cada sub-banda do sinal decomposto é usada como característica da fala. Esta característica é aplicada a quantização de vetor e ao modelo oculto de Markov (HMM – *Hidden Markov Model*). Para extrair as melhores decomposições, a função custo utilizada foi a entropia de Shannon. Os resultados experimentais foram comparados com a classificação dos coeficientes cepstrum. Os resultados mostraram uma taxa de acerto de 80,35% usando o cepstrum, e 81,777% utilizando características obtidas com a Wavelet adaptativa. O desvio padrão para o cepstrum foi de 9,55% e para Wavelet adaptativa foi de 5,26%, mostrando que esta nova técnica, segundo os autores, é mais apropriada para o reconhecimento de fala independente do locutor.

Em (LIANG, 1998) os autores propõem um sistema para classificar os sons do coração em dois grupos, murmúrio fisiológico e murmúrio patológico. Para tal, utilizaram a WPT como extrator de características. A wavelet mãe utilizada foi a Symlet8 com as funções custo entropia de Shannon e o critério 1^P , este último com melhores resultados.

A classificação é realizada através de uma rede neural do tipo *Multi Layer Perceptron* (MLP) treinada com o algoritmo *Backpropagation* (BKP), obtendo 85% de exatidão, com um conjunto de treinamento de 65 pacientes e um conjunto de teste de 20 pacientes.

O autor em (SCHUCK, 1998) estudou a aplicação da Transformada Wavelet Discreta (DWT) como ferramenta de análise de vozes patológicas. É mostrado que os níveis básicos de resolução, cuja faixa espectral é onde se localiza a componente de frequência fundamental (*Pitch*), são distorcidos por “aliasing”, invalidando seu uso para análise de *Jitter* e *Shimmer*. É também proposto um índice para avaliar a soproisidade, usando a função Haar como base de decomposição. Foram realizadas duas aquisições por dois procedimentos diferentes, utilizando sinais de 64 pacientes previamente diagnosticados.

Foi mostrado que o índice proposto é estatisticamente correlacionado com o índice Br para o cálculo da soproisidade. Segundo o autor, este novo método serve tanto para o auxílio ao diagnóstico quanto para o acompanhamento dos resultados obtidos por um tratamento.

Em (OBAIDAT, 1998) e (OBAIDAT, 1999), os autores propuseram um algoritmo baseado na Transformada Wavelet Diádica (DyTW) para detectar o período de *Pitch*. Segundo os autores, as técnicas tradicionais como autocorrelação e o método baseado em cepstrum, falham em fornecer uma estimativa apurada. Os autores desenvolveram e testaram um algoritmo para detectar *Pitch* em sinais sintéticos. Diversos experimentos foram realizados utilizando sinais com ruído e em condições de ambiente ideal, de forma a avaliar a precisão e robustez do método proposto. O objetivo é encontrar o período do *Pitch* de um sinal ruidoso que seja o mais perto ou parecido de um sinal limpo (sem ruído). Para isso, o algoritmo calcula a Transformada Wavelet do sinal e seleciona uma escala L que minimiza o erro quadrático médio da decomposição com o próprio sinal. A função de escalonamento utilizada não só seleciona a escala L, mas também melhora a relação sinal/ruído. Os resultados apresentados utilizando o algoritmo proposto tiveram excelentes resultados em estimar o período *Pitch* (aproximadamente 100%) em sinais corrompidos com ruído branco.

Os autores em (ENGLEHART, 1999) propuseram as representações tempo-freqüência como extrator de características para a classificação de padrões de sinais eletromiográficos (MES – *Myoelectric Signal*). O artigo objetiva comparar a performance de três etapas de classificação definidas pelos autores (extração de características, redução de dimensão e classificação).

Extração de características é a primeira etapa usada para representar os sinais. Os métodos foram às representações nos domínios tempo e tempo-freqüência lineares:

- a) Características no domínio tempo (TD);
- b) *Short Time Fourier Transform* (STFR);
- c) Transformada Wavelet (WT);
- d) Transformada Wavelet Packet (WPT).

Os classificadores:

- a) Análise discriminante linear (LDA);
- b) Perceptron multicamadas (MLP).

Após os autores terem testado todas as combinações das etapas de classificação citadas acima, as técnicas que obtiveram melhor performance (classificadas em ordem decrescente) foram WPT, WT, STFT e TD. A WPT teve um erro médio de classificação de 6,25%.

Em (KESHAVA. 1999) é tratado o problema de aproximar um conjunto de processos aleatórios gaussianos por uma única representação de funções Wavelets. É medida a similaridade entre o processo original com as funções Wavelets, utilizando os coeficientes de Bhattacharyya como um critério de função custo para a seleção da *Best Basis*. O coeficiente de Bhattacharyya fornece um limite superior na probabilidade de erro quando detecta a presença de um dos dois processos. São usadas duas classes de processos para ilustrar o método:

- a) Markov de 1ª ordem e;
- b) Banda limitada.

Os autores aplicaram o algoritmo para uma classificação mais robusta de terrenos em imagens polarimétricas SAR (*Synthetic Aperture Radar*).

No artigo (JIANG, 1999) os autores propuseram três métodos de extração de características para fins de classificação de sinais, selecionando células no plano tempo-frequência (ou também denominados de átomos). Os átomos selecionados contêm características discriminantes entre diferentes classes de sinais. Estes métodos utilizam os átomos da biblioteca de funções ortogonais Wavelet Packet. Os autores comparam a classificação de dois sinais experimentais:

- a) Formas de onda triangular e;
- b) Sinais de fala.

O primeiro método proposto desempenha uma procura discriminante chamada *Discriminant Pursuit* (DP). O segundo método utiliza o *Local Discriminat Basis* (LDB), proposto por Saito e Coifman (SAITO, 1994), e o terceiro é uma combinação dos dois métodos. As menores taxas de erro de classificação para os dois sinais foram obtidas utilizando-se o método LDB; porém, os autores relataram que em um trabalho prévio

utilizando sons acústicos emitidos de motores de carro e caminhão, o método DP apresentou o melhor resultado. Com isso, eles afirmam que a escolha do melhor método é dependente dos tipos de sinais a serem classificados.

Em (BALMELLI, 1999) os autores desenvolveram um algoritmo de classificação de texturas de imagem utilizando técnicas no domínio Wavelet. O primeiro passo do algoritmo como extrator de características é a decomposição da imagem utilizando Wavelet Packet. O filtro utilizado foi o *quadratic spline*. A extração de características é desempenhada nas sub-bandas selecionadas. Para testes, foi utilizado um banco de dados com 340 imagens com características visuais de texturas definidas pela percepção humana como direção, simetria, regularidade e tipo de regularidade. Para as direções vertical e horizontal da imagem, são calculados os coeficientes que representam medidas de direção, regularidade e simetria. Por fim, a classificação é efetuada utilizando os coeficientes calculados, que são comparados com os coeficientes considerados como referência. Os valores limites estimados foram pré-selecionados do conjunto de diferentes texturas do banco de dados. Os resultados apresentados demonstraram a robustez do algoritmo, com um erro médio de classificação menor que 1%. Segundo os autores, esta técnica pode ser usada para diversas aplicações de processamento de texturas, como análise, segmentação e recuperação.

No artigo (MANFREDI, 2000a), alguns métodos de estimação de *Pitch* são comparados e suas vantagens e desvantagens são apontadas. As técnicas utilizadas e descritas neste artigo são:

- a) Modelagem Autoregressiva (AR);
- b) *Average Magnitude Difference Function* (AMDF);
- c) *Simple Inverse Filter Tracking* (SIFT);
- d) *Cepstrum* (CEP);
- e) Transformada Wavelet (WT).

A comparação entre as técnicas é feita em termos de robustez a ruído, resolução tempo-freqüência e capacidade de procura. Os autores utilizaram sinais de fala reais e sinais

sintéticos. Os sinais sintéticos gerados são sons vozeados, através da modelagem Autoregressiva (AR, *all-pole*), onde o efeito da cavidade nasal é desprezado. Dos sinais reais foram utilizados vozes patológicas e vozes normais. As comparações mostraram que para sinais de voz altamente degradados por ruído, os métodos SITF e WT apresentaram os melhores resultados.

Já em (MANFREDI, 2000b), a autora propõe a estimação da frequência fundamental filtrando o sinal com a escala de Transformada Wavelet Contínua (CWT) apropriada, e então extraindo seu período fundamental através do *Average Magnitude Difference Function* (AMDF). É também proposto no artigo o cálculo adaptativo da Energia do Ruído Normalizada (*Normalized Noise Energy* -NNE) para estimação do ruído. Este parâmetro é a medida da energia de componentes de vozes disfônicas, em relação a energia total do sinal. Dois métodos de estimação da frequência fundamental foram utilizados:

- a) CWT + AMDF
- b) CWT + AS (autocorrelation sequence)

Os métodos foram aplicados a sinais simulados, com diferentes relações sinal/ruído (SNR) e valores de *Jitter*, e em sinais patológicos reais. Em geral, os resultados em vozes normais para os métodos um e dois foram similares, porém, apresentaram grandes diferenças para vozes patológicas. Em particular, a autora afirma que o novo método oscila menos que os outros, oferecendo resultados mais confiáveis. A autora conclui que o procedimento proposto consegue seguir rápidas variações do *Pitch* de forma robusta a ruído e a *Jitter*. Assim, a funcionalidade da glote pode ser analisada por meio de índices objetivos ao invés de por inspeção visual do espectrograma.

Em (SARIKAYA, 2000), os autores investigaram a utilização de Parâmetros Wavelet Packet (WPP) e parâmetros da *Sub-Banda Base Cepstral* (SBC) ambos definidos no artigo, em comparação ao tradicional *Mel-Frequency Cepstral Coefficients* (MFCC) para avaliar o impacto da “fala estressada” no reconhecimento da fala. A fala estressada (*stressed speech*) envolve estilos na fala da língua inglesa, como:

- a) *Areneutral* (natural);

b) *Angry* (irritado);

c) *Loud* (alto);

d) *Lombard effect* (situação onde o falante força uma modificação na sua fonação para aumentar ou melhorar a qualidade de comunicação, típica em locutores de rádio e televisão).

Os autores mostraram que o uso dos parâmetros WPP e dos SBC implicaram em um aumento na taxa de reconhecimento se comparado ao método tradicional MFCC, e que esta melhora foi mais significativa e consistente para classes de consoantes.

Em (SHIRAZI, 2000) é apresentada uma metodologia de classificação de texturas de imagem baseado no modelamento estocástico de texturas no domínio frequência. Os autores mostram que diferentes texturas possuem respostas em frequência em bandas intermediárias, o que justifica o uso da WPT na tentativa de caracterizar e classificar texturas. Os autores propuseram um esquema de classificação de texturas baseado em modelos de Markov no domínio Wavelet. Em relação à Transformada Wavelet padrão (DWT), a Transformada Wavelet Packet (WPT) fornece a possibilidade de dar *zoom* em qualquer faixa de frequência desejada. Os resultados experimentais relatados mostraram alta taxa de classificação sobre uma variedade de 30 tipos de texturas diferentes.

Em (HSIN, 2000), os autores propuseram a implementação da WPT para análise e classificação de texturas. Porém, as características de texturas similares e com orientação simétrica não são de simples separação. Por isso, os autores sugeriram a implementação da Transformada Wavelet Packet modulada, que pode ser implementada pela estrutura de árvore convencional, desenvolvida por Coifman e Wickerhauser (COIFMAN, 1992). A performance desta nova transformada proposta pelos autores foi demonstrada na segmentação de texturas de Brodatz e em imagens aéreas de São Francisco. Por fim, eles relataram que os resultados da segmentação de texturas são melhores com a WPT modulada, pois estas preservam as propriedades desejadas.

Em (GARCIA, 2000) é apresentado um novo método de reconhecimento de faces humanas com iluminação constante. Este sistema é baseado na análise da WPT de imagens de faces. Cada face é localizada e descrita por um subconjunto de imagens filtradas em bandas, contendo os coeficientes Wavelets. A partir destes coeficientes são construídos vetores de características as quais caracterizam a textura da face através de medidas estatísticas simples. É utilizada a distância de Bhattacharyya para classificar os vetores de características de faces em classes de pessoas. A Transformada Wavelet é aplicada nas imagens e o critério de seleção do nível mais profundo da árvore de decomposição foi baseado em entropia. As imagens foram decompostas até o nível dois, formando 15 imagens de detalhes e 1 imagem de aproximação. Cada uma das 16 matrizes de coeficientes contém informação da textura da face. Em cada imagem foram calculados os valores médios e seus desvios. É assumido que as distribuições de probabilidade são gaussianas.

Os autores concluem que a análise da imagem com a WPT fornece excelente descrição de textura e sua implementação é muito rápida. A distância de Bhattacharyya provou ser muito eficiente para a classificação, obtendo uma taxa de classificação de 97% para 600 imagens testadas. Comparando os métodos de reconhecimento, outra técnica descrita no artigo requer uma quantidade extra de espaço para armazenar os dados e as imagens têm que ser normalizadas para uma melhor taxa de classificação. E por fim, o tempo de treinamento utilizado pela WPT é consideravelmente menor.

No artigo (MEYER, 2000), é desenvolvido um algoritmo numérico rápido para WPT, chamado *Fast Wavelet Packet* (FWP), com o objetivo de demonstrar a vantagem em se construir uma base adaptada a uma determinada imagem. Os autores afirmam que as Wavelets não são as mais apropriadas para representar padrões oscilatórios, pois rápidas variações de intensidade podem apenas ser descritas por coeficientes de pequena escala na Wavelet, na qual são freqüentemente quantizados para zero.

Este algoritmo então foi aplicado para compressão de imagens e testado em quatro imagens com variações de texturas escolhidas por serem de difícil compressão, pois elas contêm uma mistura de longas regiões suaves e longos padrões oscilatórios. A performance do algoritmo foi comparada com outro codificador Wavelet. Os resultados demonstraram que

o codificador FWP tende a criar artefatos na mesma localização (nas bordas) que o codificador Wavelet, com intensidade similar. As vantagens da FWP, segundo os autores, são:

- a) O algoritmo seleciona uma base que preserva as oscilações de textura,
- b) O algoritmo é quatro vezes mais rápido.

No artigo (YEN, 2000) é introduzida uma alternativa para extrair informações tempo-frequência das assinaturas de vibrações de máquinas usando a WPT. O objetivo do trabalho desenvolvido pelos autores é verificar a viabilidade da WPT como uma ferramenta de procura de características que pode ser aplicável na detecção e classificação de sinais de vibrações mecânicas. Pode ser monitorado, com a aplicação de um algoritmo para análise de dados, capaz de detectar mudanças na assinatura de vibrações. Isso possibilita a manutenção preditiva.

Uma deficiência inerente das bases Wavelets é a ausência da propriedade de invariância à translação. Para resolver este problema, os autores utilizaram a energia do sinal contido em cada nó da árvore de decomposição Wavelet Packet. Ao invés de tentar encontrar uma transformação linear para reduzir a dimensionalidade do sinal, os autores calcularam a potência discriminante de cada característica individual, descartando aquelas componentes que continham pouca informação de separabilidade da classe, medida por um critério selecionado. Dois critérios foram propostos. O classificador utilizado foi uma ANN. Os resultados dos testes com WPT foram comparados com a transformada de Fourier, ambos alcançaram excelentes resultados na classificação. Mais ainda, quando os sinais foram corrompidos artificialmente por diferentes ruídos, simulando um ambiente ruidoso real, a WPT apresentou resultados promissores em relação a Fourier.

Os autores do artigo (HUANG, 2000) apresentam o problema da detecção e classificação de alvos submarinos, através dos sinais emitidos por um sonar, sinal que vai até os alvos e voltam (ecoam) para a fonte trazendo informações sobre o objeto (assinatura do objeto). O sistema consiste num extrator de características e um classificador não linear; a WPT e Redes Neurais do tipo MLP respectivamente. Os objetos alvos com os quais o sistema

foi treinado são basicamente minas e não minas (outros objetos diferentes de minas). Com 50 simulações o resultado foi de 98% de taxa de sucesso.

No trabalho de (De YAO, 2002) são apresentados diversos subsistemas extratores de características cooperando juntamente com a WPT (a base é Symlet 4) e ANN no contexto de classificação de alvos submarinos. O objetivo é determinar qual é a melhor banda para o funcionamento de um sonar para que o mesmo seja utilizado em discriminar alvos militares, como minas submarinas, e outros objetos de tamanho similar, com o menor erro. A performance da classificação é demonstrada e testada com uma base de dados composta por sinais acústicos emitidos pelo sonar em largura de banda de 40kHz e 80kHz. Um MLP foi utilizado treinado com o BKP para a classificação, obtendo resultados promissores. Foi determinado que a melhor banda para o sonar é de 80kHz com probabilidade de acerto de 95.83% para o sistema como um todo.

Em (SCHUCK, 2002) foi demonstrada a viabilidade de utilizar a WPT e o BBA como extrator de características para sistemas de classificação de voz disfônica. Foi mostrado que a melhor escolha da wavelet foi a Symlet5 e a melhor função custo foi a Entropia de Shannon. Também um discriminador linear foi desenvolvido, avaliando a existência ou não de certos nós da árvore podada da WPT em 5 níveis de dilatação. Com estes critérios, os autores conseguiram um percentual de acerto de 76,92% para pacientes sem patologia (falsos positivos 23,07%), e um percentual de acerto de 85,42% para pacientes com patologia (falsos negativos 14,58%).

Os autores do artigo (HONG, 2003) utilizam a Transformada Wavelet Discreta (DWT) em conjunto com ANN para reconhecimento de sons provenientes do coração. A rede neural utilizada é uma MLP de 15 camadas ocultas treinadas com o algoritmo BKP, tendo como entradas os próprios coeficientes de detalhe e de aproximação da DWT. Os resultados são comparados com a técnica da *Principal Component Analysis* (PCA), sendo que a taxa de sucesso do algoritmo proposto foi superior, com 98% de sucesso em classificar os sons do

coração em normais e patológicos; e 88% de sucesso em separar os sons em 6 grupos, o grupo de sons Normais, e 5 grupos de sons anormais.

Em (SCHUCK, 2003) foi utilizado uma ANN em adição a WPT e o BBA com 5 níveis de dilatação utilizando a entropia de Shannon como função custo e a Symlet5 como a base de decomposição conforme (SCHUCK, 2002), para produzir um classificador não linear. O objetivo era realizar a separação automática entre vozes patológicas e normais. Os valores de entropia de cada nó da árvore da *Best Basis Tree* (BBT) constituíram as entradas da rede neural do tipo MLP, formada de 4 neurônios na camada oculta e 2 neurônios na camada de saída, treinada com o algoritmo *Backpropagation*, utilizando a tangente hiperbólica como função de ativação. Foram utilizadas as vozes de 61 pacientes previamente diagnosticados para o treinamento e verificação da ANN. Os resultados mostram um erro de classificação de 4,6% de falsos negativos, 10,9% de falsos positivos, com uma taxa de sucesso de 84%. Mostrou-se uma melhoria de performance em relação ao sistema proposto tanto em (SCHUCK, 2002) quanto em (PARRAGA, 2002), apontando a existência de não linearidades no espaço de entradas definido pelos nós da BBT. O artigo conclui com a sugestão de incrementar o banco de vozes e separar as vozes por patologia, ou grupo de patologia.

3 FUNDAMENTOS TEÓRICOS

3.1 INTRODUÇÃO

Os fundamentos teóricos estão divididos em três partes. Na Seção 3.2, será apresentado um breve resumo da fisiologia da voz, assim como das doenças relacionadas ao sistema fonador, mais especificamente na laringe. Também é apresentado o modelo acústico utilizado neste trabalho.

Na Seção 3.3 será mostrada a teoria da Transformada Wavelet Packet (WPT), e brevemente o algoritmo *Best Basis* (BBA). Na Seção 3.4 é apresentado um breve histórico das Redes Neurais Artificiais (ANN), assim como também o modelo de rede utilizado neste trabalho, o *Multilayer Perceptron* (MLP).

3.2 FISIOLOGIA DA VOZ E PATOLOGIAS

3.2.1 Fisiologia Da Voz

Segundo (KENT, 1992) o sistema fonador pode ser dividido em três subsistemas principais:

- a) Respiratório;
- b) Laringeal;
- c) Articulatório.

3.2.1.1 Subsistema Respiratório

O subsistema respiratório é formado pelos pulmões, traquéia, diafragma e brônquios, conforme mostra a Figura 2. Este sistema, além de fornecer ventilação para o apoio a vida, também produz energia aerodinâmica para a fala. Os sons são formados em quase todas as

línguas no momento da expiração. Em termos simples, o subsistema respiratório é uma bomba de ar que fornece energia aerodinâmica aos subsistemas da laringe e articulatório para a formação dos sons.

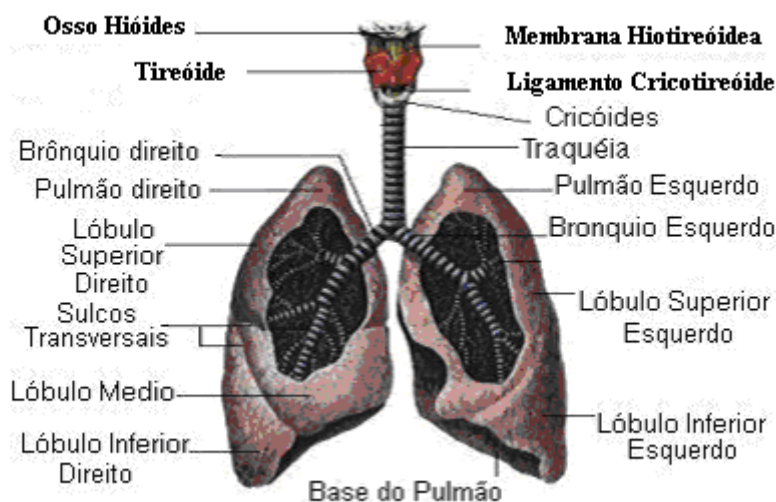


Figura 2: Diagrama esquemático que mostra as partes do sistema respiratório e a laringe. (extraído de (ROCHA, 2004)).

3.2.1.2 Subsistema da Laringe

A laringe está localizada acima da traquéia e debaixo da faringe. Ela é composta por varias cartilagens, entre elas, a cricóide, tireóide, aritenóides, epiglote, conforme a Figura 3.

A cricóide separa a laringe da traquéia. A tireóide (também chamada de “pomo-de-Adão”) é a de maior tamanho e pode ser apalpada. A epiglote é posicionada na parte superior da laringe, e fecha a laringe ao se deglutir, impedindo desta maneira a passagem de líquido e/ou alimentos para a via aérea.

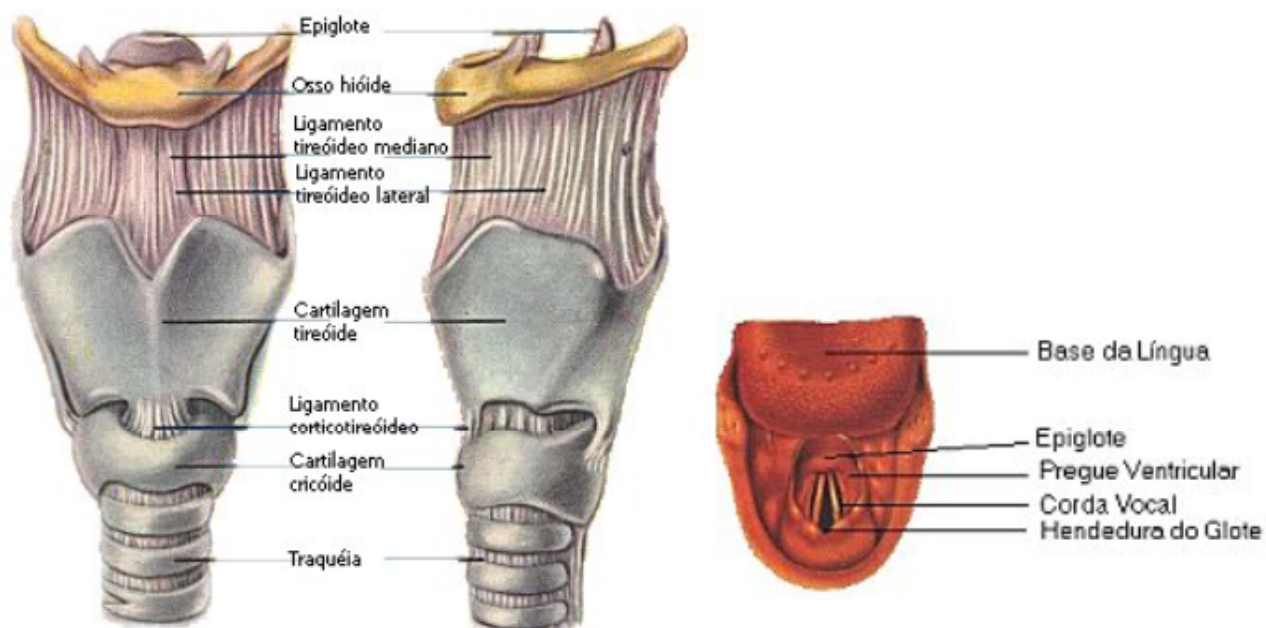


Figura 3: Partes que compõem a laringe (extraído (ROCHA, 2004)).

Entre a parte interna da base das aritenóides e da tireóide, estão as cordas vocais (ou também chamadas pregas vocais), que são 2 paredes de lábios simétricos formados por um músculo e um tecido elástico. As extremidades das cordas vocais estão fixadas sobre as aritenóides. Ao espaço, normalmente triangular compreendido entre as cordas vocais, dá-se o nome de glote. A vista superior da laringe se encontra na Figura 4.

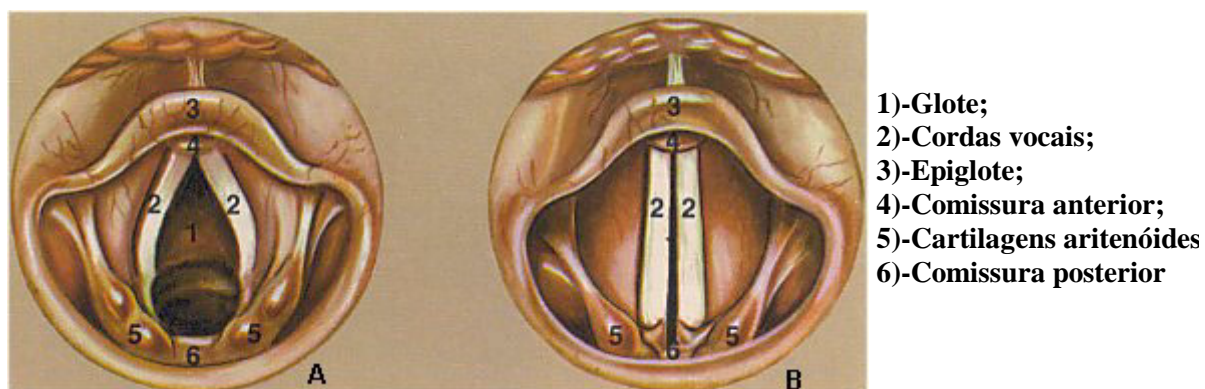


Figura 4: Vista superior da laringe, mostrando as posições da glote.

A) Glote na posição de repouso; B) Glote durante a atividade (extraído (ROCHA, 2004)).

As vogais são sons produzidos com a vibração das cordas vocais (sendo o ar do sistema respiratório a fonte de energia). No ato da fala, na fase de expiração, um conjunto de músculos abduz as pregas vocais até que elas se encostem uma à outra. O ar proveniente dos pulmões faz com que as cordas vocais vibrem, produzindo um som chamado som glotal. Do ponto de vista acústico, o ciclo de abertura e fechamento da glote dá origem às vibrações sonoras, que produzem a voz.

A frequência de vibração do ar supra-glótico depende de vários fatores, tais como comprimento e espessura das cordas vocais, pressão da corrente de ar, etc. Numa conversação normal, esses fatores são constantemente modificados, para se obter as frequências desejadas, que variam na fala normal de 60 a 300 Hertz. Essa possibilidade de regular a taxa de vibração das cordas é até certo ponto individual, estando relacionada com a idade, sexo e as particularidades dos falantes. Quanto mais longas e espessas as cordas vocais, mais lentas as vibrações; e quanto mais curtas e finas, maior a frequência. É por isso que a voz da mulher e das crianças é mais aguda que a dos homens. A velocidade de vibração das cordas vocais oscila na voz masculina entre 100 a 150 Hz (ciclos/segundo), e na voz feminina, entre 200 a 300 Hz. O tempo entre duas aberturas consecutivas das cordas vocais é chamado de período fundamental T_0 , enquanto que a taxa de vibração é chamada de frequência fundamental da fonação $F_0=1/T_0$. O período fundamental é dependente do tamanho e da tensão das cordas vocais do falante em um determinado instante (DELLER, 1993). Este subsistema é o responsável pela produção dos sons vocálicos (sons de todas as vogais e algumas consoantes), através da vibração das cordas vocais.

3.2.1.3 O Subsistema Articulatório

O som laríngeo atravessa a faringe, que lhe oferece duas vias de acesso ao exterior: a cavidade bucal e a nasal. Essas cavidades articulatórias incluem a língua, os lábios, a mandíbula, palato mole e palato duro, como mostra a Figura 5. O conjunto destas estruturas forma o sistema vocal, e o seu movimento determina as propriedades ressonantes. O subsistema articulatório não produz sons vocálicos, mas modifica os sons globais, atuando como cavidade ressonante variável que altera o conteúdo harmônico dos sons globais.

A cavidade faríngeo-bucal pode ser associada à cavidade nasal quando o palato mole não está colado à parede posterior da faringe. Nesse caso é permitida a passagem do ar, ao mesmo tempo, pela boca e pelas fossas nasais, e é a articulação do palato mole que determina se um som é oral ou nasal (KENT, 1992).

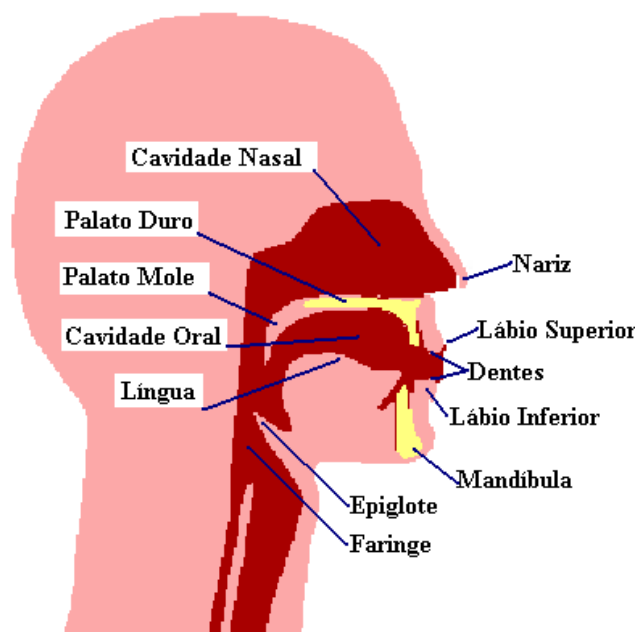


Figura 5: Representação das partes do sistema articulario.

3.2.2 Modelos Acústicos da Produção da Fala e Alguns Aspectos Espectrais

A análise acústica torna-se viável graças a considerações de linearidade e invariância no tempo. Um entendimento dos métodos modernos de análise acústica requer, no mínimo, uma apreciação da chamada teoria do filtro fonte (*source filter theory*) (KENT, 1993). Este conceito propõe que a energia acústica gerada por uma fonte sonora, na verdade passou através de um sistema de transmissão dependente em frequência, um filtro. A tarefa de análise da fala portanto pode ser grosseiramente sumarizada em:

- a) Identificar uma fonte de energia sonora;
- b) Descrever uma função, ou filtro correspondente.

A literatura destaca três principais fontes sonoras a serem consideradas:

- a) Voz Laringeal (ex. vogais);
- b) Ruído Turbulento (ex. consoantes friccionadas);
- c) Transientes (ex. explosão das consoantes de parada).

A chamada voz laringeal, fonte de energia acústica tipicamente considerada para a geração das vogais, é o som gerado pelas cordas vocais, antes de passar pelo trato vocal e ser irradiado pelo meio (geralmente pelo ar) (KENT, 1993). Esta energia sonora tem muita importância por trazer grande quantidade de informação sobre as cordas vocais, permitindo assim a identificação isolada de disfunções das mesmas. Nos próximos parágrafos apresentam-se de forma bastante simplificada os fundamentos da teoria do filtro fonte para a geração de vogais.

A teoria do filtro fonte para produção de vogais é ilustrada na Figura 6 e pode ser sumarizada numa análise no domínio frequência, indicada pela Equação (1):

$$P(f) = U(f) \cdot T(f) \cdot R(f) \quad (1)$$

onde:

f) Indica a frequência em Hertz;

$P(f)$ É o espectro do som emitido;

$U(f)$ Espectro do som glotal;

$T(f)$ Função de transferência do sistema vocal;

$R(f)$ É a característica de radiação (interface boca/meio exterior).

A Equação (1) descreve o comportamento da pressão sonora no domínio frequência, denotada como $P(f)$, tal como pode ser medida por um microfone próximo à boca de um falante. A fonte de energia sonora para a geração de vogais é o sinal vozeado gerado pela vibração das cordas vocais (KENT, 1993).

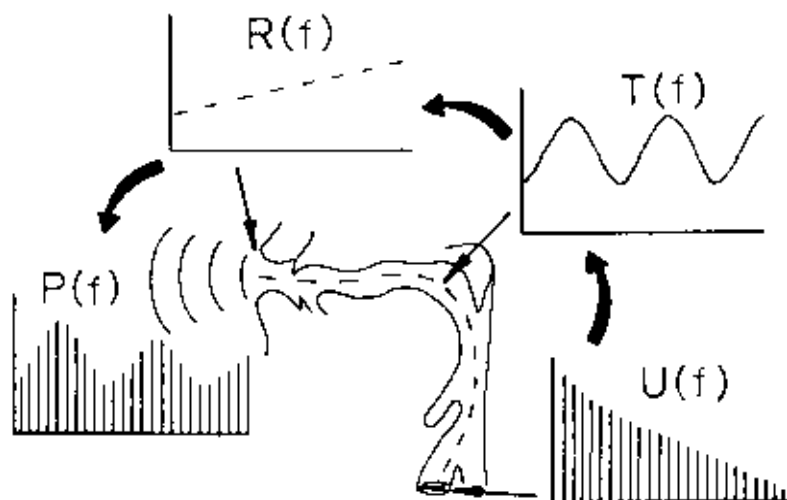


Figura 6: Representação do modelo filtro-fonte para vogais (extraído de (KENT, 1993)).

A fonte de energia acústica para vogais é tipicamente o sinal de voz gerado pela vibração das cordas vocais, aqui chamado $U(f)$. Este sinal é filtrado no trato vocal (uma parte do filtro) e sofre um efeito de radiação característica da mecânica da propagação do som (a outra parte do filtro). A forma de onda como é gerada na glote, representada por $U(f)$, idealizada e simplificada é uma série de pulsos triangulares espaçados entre si, de um período fundamental. Em frequência esta forma de onda é representada por um espectro de harmônicos cujas componentes diminuem em amplitude a uma taxa de 12dB por oitava (KENT, 1993). A análise do sinal, permite avaliar o bom funcionamento das cordas vocais e os indícios da existência de disfunções das mesmas. No entanto, os pulsos acústicos gerados pela vibração das cordas vocais propagam-se pelo trato vocal onde ocorre filtragem e depois ocorre a radiação do som modulando e atenuando o sinal.

A função de transferência do trato vocal (o primeiro filtro) é definida primeiramente pelas frequências de ressonância deste. Estas frequências de ressonância são também denominadas de formantes. Cada formante é caracterizado por uma frequência central e uma largura de banda. Tipicamente de 3 a 5 formantes são consideradas na análise da fala, mas na verdade são infinitos em quantidade (KENT, 1993). Estes formantes podem ser determinados matematicamente através de um conhecimento preciso da forma do trato vocal ou podem ser estimadas por medidas do sinal acústico. Estimativas matemáticas dos formantes são possíveis através do estudo acústico de um modelo simplificado. Dado que os formantes são função do

comprimento do trato vocal e de sua seção transversal, e fazendo a forma da seção transversal do trato vocal também em função deste comprimento (por um polinômio), boas aproximações são possíveis. Desta forma o comprimento do trato vocal determina o espaçamento médio das frequências formantes. Isto deriva de um simples modelo acústico de um tubo fechado em um lado e aberto do outro, como mostra a Figura 7.

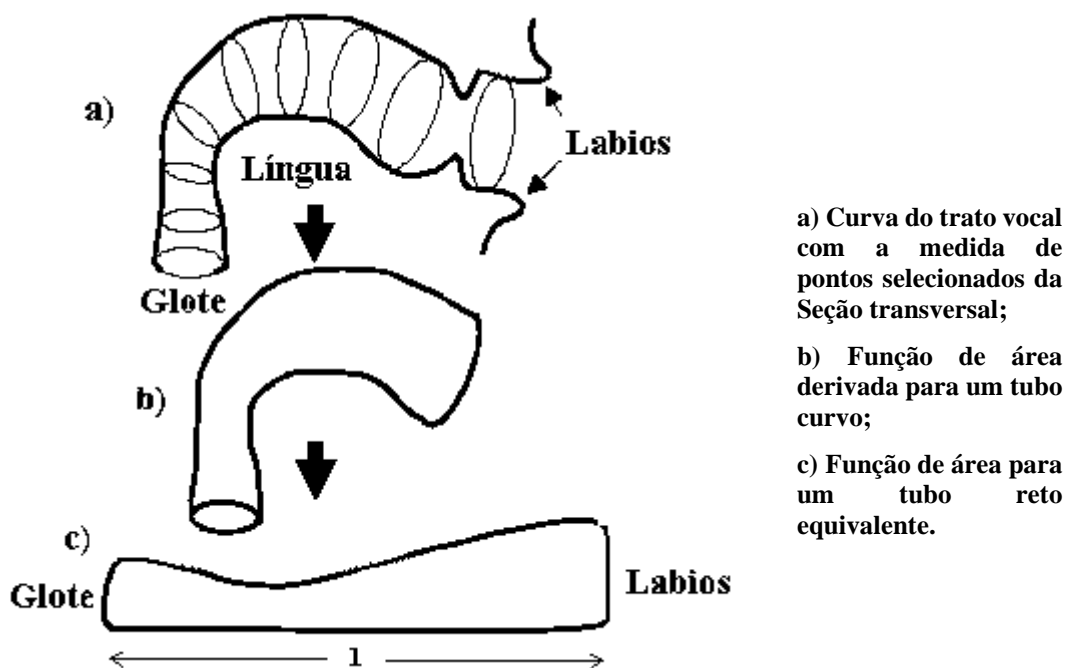


Figura 7: Esquema simplificado do trato vocal. (extraído de (KENT, 1992)).

Os formantes podem ser estimados pela Equação (2) (KENT, 1993):

$$F_n = \frac{(2n-1)c}{4l} \quad (2)$$

onde:

F_n É uma frequência formante em particular;

$(2n-1)$ Dá os inteiros ímpares;

c É a velocidade do som no meio em questão;

l É o comprimento do trato vocal.

Os formantes variam conforme o comprimento do trato vocal e, conseqüentemente, com as características do falante, como idade e sexo. Alguns artifícios, como a protusão de lábios e encolhimento da laringe, alteram l e são usados por muitos falantes e cantores. A Figura 8 ilustra os três primeiros formantes, $F1$, $F2$ e $F3$. Este modelo simplificado, precisa ser modificado para servir para a produção de diferentes vogais. Em particular, o modelo precisa permitir que a área da seção transversal varie ao longo do comprimento do trato vocal.

Esta função é obtida medindo-se a seção transversal do trato vocal em alguns pontos característicos selecionados e interpolando a partir destes um polinômio aproximado. A teoria permite relacionar perturbações na configuração do trato vocal com seus respectivos efeitos sobre as freqüências formantes.

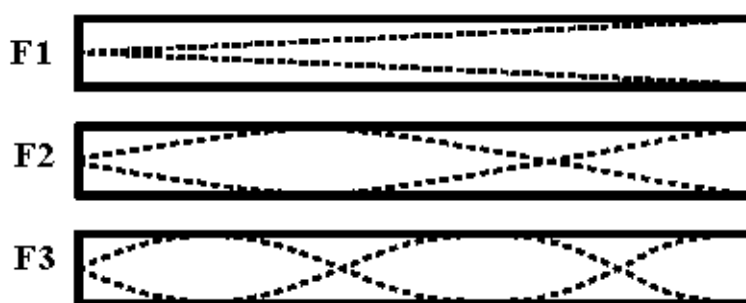


Figura 8: Modos de ressonância dos formantes.

Esta teoria é substancial não só no campo conceitual mas também na prática, descrevendo melhor a acústica fonética e permitindo o desenvolvimento de sintetizadores. Apesar de tudo, a teoria linear é apenas uma aproximação, pois já se comprovaram não-linearidades mesmo para vogais, devido à interação de jatos de ar que geram vértices turbulentos (KENT, 1993).

Uma das conclusões primárias baseadas na teoria acústica acima é que uma vogal em particular está associada com um padrão de formantes característico. A idéia de que uma vogal está associada com um padrão de formantes em particular dá uma idéia de como é a teoria de reconhecimento de voz, claro que também existem outros fatores dinâmicos a serem considerados, os quais podem ter papel significativo na identificação de vogais (KENT, 1993). Como as freqüências formantes para as vogais variam de acordo com o comprimento do trato

vocal de falante para falante, a normalização das frequências formantes tem sido tópico de pesquisa.

Normalmente apenas os três primeiros formantes, *F1*, *F2* e *F3* são considerados para este tipo de estudo, mas isto não significa que os formantes de mais alta frequência não tenham importância para algum propósito. Estudos comprovam que o acréscimo dos formantes mais altos em sintetizadores de voz dá maior naturalidade à voz sintetizada. Sabe-se também que os mais altos formantes, principalmente o quarto formante, *F4*, é responsável por um fator de distinção do falante em particular, o chamado timbre (KENT, 1993).

3.2.3 Patologias

A laringe tem funções biológicas, emocionais e lingüísticas. Elas protegem a via aérea e mantém a vida. A partir dos 3 meses de idade, o bebê parece expressar emoções através de sons laríngeos. A emotividade das pessoas pode ser ouvida na sua voz (BOONE, 1994).

Além de a laringe fornecer o sinal acústico para a fala, o canto, a e exposição dramática, o riso, o choro, a voz reflete a identidade individual, a personalidade e o estágio da vida. Um entendimento da fisiologia fonatória, bem como os fatores que podem perturbá-la, é necessário para todos os profissionais envolvidos no atendimento da voz (COLTON, 1996).

Alguns mecanismos de proteção da via aérea são mecânicos e agem para fechá-la; outros são expulsivos e servem para forçar substâncias estranhas para fora da via aérea. Todos esses mecanismos são reflexos e operam sob controle do subconsciente (COLTON, 1996).

A voz, ou vocalização é produzida pela vibração das pregas vocais, quando o fluxo de ar que vem dos pulmões passa entre elas. A maioria dos problemas que contribuem para uma voz inadequada tem mais a ver com a aproximação falha das pregas mais do que com dificuldades de respiração (BOONE, 1994).

Os distúrbios da voz resultam de estruturas orgânicas em mau funcionamento no trato vocal, ou pelo mau uso dos mecanismos vocais. Quando a voz muda de alguma forma negativa, tais mudanças possuem muitos nomes comuns, tais como: rouquidão, rudeza, aspereza, estridência, entre outras, mas termo mais genérico é disfonia (BOONE, 1994).

3.2.3.1 Principais sintomas

Um sintoma é um relato de alguma alteração percebida pelo paciente ao seu médico. Estes podem ser descritos de diversos modos, como uma queixa de sensações associadas à fonação ou dores na região da garganta. Alguns sintomas podem ser verificados, outros não. Outros sintomas podem referir-se às características perceptuais da voz, como a rouquidão, garganta arranhando ou voz trêmula (COLTON, 1996). Tabela 2, encontram-se os oito principais sintomas que os pacientes com distúrbios vocais apresentam. Estes sintomas em geral não ocorrem individualmente, aparecendo, conseqüentemente, de forma combinada (COLTON, 1996).

Tabela 2: Principais Sintomas (extraído de (COLTON, 1996)).

Rouquidão	Este sintoma reflete uma vibração aperiódica das pregas vocais (voz “rouca”, “áspera” ou “raspada” são sinônimos utilizados).
Fadiga Vocal	Pacientes queixam-se de cansaço após a fala prolongada e que a fala contínua requer muito esforço.
Soprosidade	Pacientes queixam-se de que são incapazes de pronunciar sentenças completas sem ficar sem ar e precisar reabastecer o suprimento de ar para continuar falando.
Extensão fonatória reduzida	Este sintoma está em geral associado a cantores que se queixam de experimentar dificuldades em produzir notas que, anteriormente, não apresentavam. Tipicamente, estas notas ocorrem na extremidade superior de sua extensão de canto.
Afonia	Afonia significa ausência de voz. O paciente fala em sussurro e pode, às vezes, queixar-se de uma variedade de sintomas, inclusive secura na garganta, dor e uma grande dificuldade para tentar falar.
Quebras de freqüência	Um paciente pode queixar-se de saltos periódicos de voz e de Quebras de voz. A voz parece fora de controle, e o paciente relata jamais saber que som sairá. Também pode ser descrito como o uso inadequado de falsete ou puberfonia. Estes jovens adolescentes que utilizam uma freqüência inapropriadamente aguda como voz habitual ao invés da voz masculina típica de sintoma é relatado por freqüência mais grave.
Voz tensa/ Comprimida	Estes pacientes relatam que é difícil falar. Isto pode incluir incapacidade de fazer com que a vocalização comece ou se mantenha. Eles relatam que é um esforço para falar, experimentam muita tensão ao fazê-lo e tornam-se fatigados devido ao esforço envolvido.
Tremor	Os pacientes podem queixar-se de que a voz está cambaleante Ou trêmula. Eles são incapazes de produzir voluntariamente um som estável sustentado.

Em (COLTON, 1996), são apresentados os diferentes sinais que acompanham a os distúrbios da voz. Estes sinais podem ser observados ou testados. E podem ser perceptuais, acústicos ou fisiológicos, todos eles intimamente ligados um com o outro. A seguir um resumo dos principais:

Tabela 3: Sinais de distúrbios na voz.

Perceptuais	Acústicos	Fisiológicos
Frequência	Frequência fundamental	Aerodinâmica
Frequência inapropriada	Frequência fundamental média da fala	Fluxo de ar aumentado ou reduzido
Quebras de frequência	Variabilidade da frequência	Pressão de ar aumentada ou reduzida
Intensidade	Extensão fonatória	Variabilidade do fluxo de ar
Variações de intensidade (suave, excessiva ou descontrolada)	Perturbação	Comportamento vibratório
Qualidade	Amplitude	Atividade Muscular
Rouquidão	Nível médio de pressão sonora	
Soprosidade	Variabilidade de amplitude	
Tensão	Extensão dinâmica	
Tremor	Ruído espectral	
Tensão/compressão	Tempo de início e fim de sonorização	
Quebra de soprosidade	Início e fim da emissão	
Diplofonia	Tempo de fonação	
Comportamento não-fonatório	Quebras de sonoridade	
Estridor	Quebras de frequência	
Pigarro	Acústica normal	
Afonia		
Consistente		
Episódica		

3.2.3.2 Principais Patologias das Pregas Vocais:

3.2.3.2.1 Nódulos Vocais

Desenvolvem-se na porção anterior das pregas vocais. São minúsculas elevações esbranquiçadas geralmente bilaterais. Consiste de tecido edematoso e/ou fibras colágenas. A lesão é geralmente simétrica e pode impedir que a glote se feche totalmente durante a fonação. Eles são causados pelo abuso contínuo da laringe e pelo mau uso da voz. Estes abusos podem ser: cantar, berrar, tossir e limpar a garganta excessivamente, por exemplo. Os nódulos podem variar em número (até quatro nódulos), e sua resposta a terapia vocal é muito boa. Também a massa aumentada dos nódulos nas pregas vocais contribui para uma altura de voz mais grave e maior periodicidade (julgada como rouquidão) (BOONE, 1994). Isso leva a um tipo de voz soprosa, monótona, que, muitas vezes, parece carecer de ressonância apropriada. As características acústicas do nódulo apresentam *Jitter* e *Shimmer* elevados. Outro fato são as evidências de ruído no espectro, grau do qual depende a severidade da rouquidão e do tamanho da lesão (COLTON, 1996).

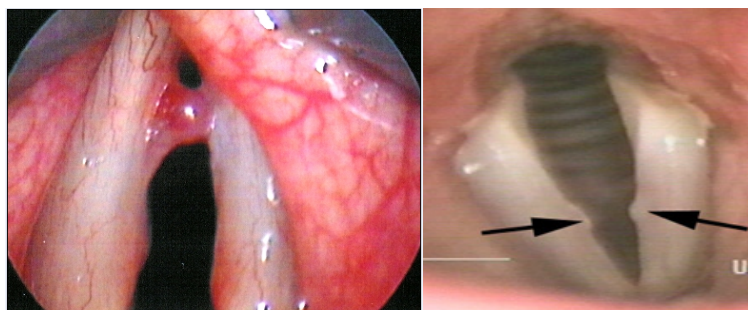


Figura 9: Fotos de nódulo. (Extraído de (GHORAYEB, 2004) e (UPMC, 2004)).

3.2.3.2.2 Pólipos Vocais

Os pólipos vocais comumente ocorrem na mesma região das pregas vocais como os nódulos, e são mais frequentes unilaterais do que bilaterais. Os pólipos e os nódulos são ambos relacionados ao hiperfuncionamento vocal e apresentam algumas semelhanças físicas. Estes têm a mesma etiologia e diferem apenas em grau. Um pólipo é maior e mais vascularizado, edematoso e inflamatório do que um nódulo. As vozes de pacientes com

pólipos unilaterais caracterizam-se por disfonia severa. A prega vocal normal vibra em uma frequência, enquanto a lesão abafa seriamente a vibração da prega envolvida, resultando no que é percebido como rouquidão e soprosidade, muitas vezes, requerendo limpeza continuada da garganta. As características acústicas são semelhantes ao nódulo, *Jitter* e *Shimmer* aumentados são esperados, assim como o aumento do ruído espectral. O principal sintoma vocal é a rouquidão (BOONE, 1994).

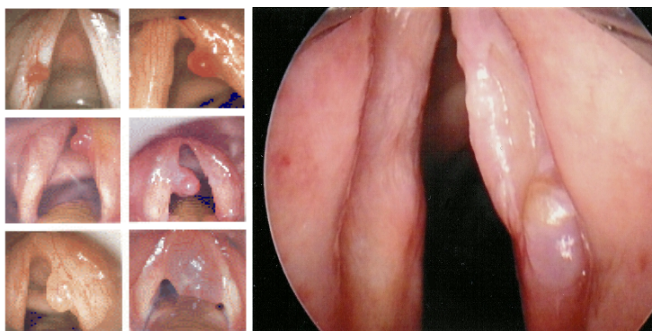


Figura 10: Diversas imagens mostrando tipos de pólipos. (Extraído de (GHORAYEB, 2004)).

3.2.3.2.3 Cistos

Os cistos intracordais parecem como pequenas esferas nas margens das pregas vocais. Eles podem ser confundidos com nódulos, porque pequenos crescimentos semelhantes a nódulos podem aparecer em uma prega, porém não na outra, e os nódulos são geralmente bilaterais. Os cistos intracordais podem ser causados por bloqueio de um duto glandular no qual há retenção de muco. Por não haver meio para o muco escapar, um cisto pode, com o tempo, aumentar. O principal sintoma do cisto é a rouquidão. Não há dados disponíveis sobre características acústicas das vozes de pacientes com cistos (COLTON, 1996). Os nódulos contêm acúmulo de gorduras, enquanto que os cistos contêm substâncias aquosas.

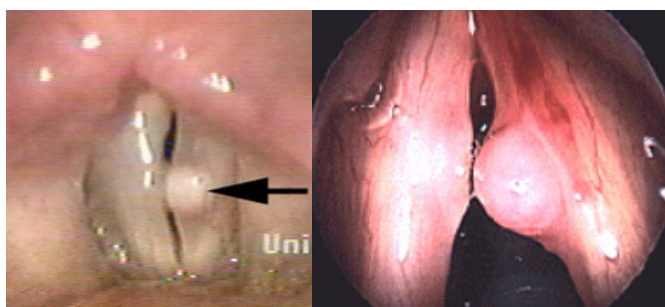


Figura 11: Fotos de cisto. (Extraído de (GHORAYEB, 2004)).

3.2.3.2.4 Edema

O Edema refere-se a um acúmulo de fluido em algum lugar na prega vocal. Ele pode ocorrer profundamente na prega vocal ou em camadas mais superficiais. Quando ocorre na primeira camada da lâmina própria, é referido como edema de Reinke. Algum inchaço aparente das pregas vocais, especialmente quando localizado, pode também ser um sinal de cisto.

Edema é uma reação natural do tecido a trauma e mau uso. Além do abuso vocal, o Edema de Reinke crônico é mais freqüentemente associado ao fumo. Dos sinais acústicos, a freqüência fundamental da fonação é mais baixa do que esperada para o sexo e a idade do paciente. Pode haver freqüência aumentada e perturbação da amplitude, bem como a presença de ruído espectral devido à rouquidão (COLTON, 1996).

Edema de Reinke é o edema bilateral da camada sub-epitelial no espaço de Reinke, isto é, na região onde o epitélio cilíndrico das cordas vocais se transforma em plano estratificado. As cordas vocais ficam enormes e flácidas (KUHL, 1982) com um edema pobre em vasos o que modifica completamente o aspecto anatômico da região glótica.

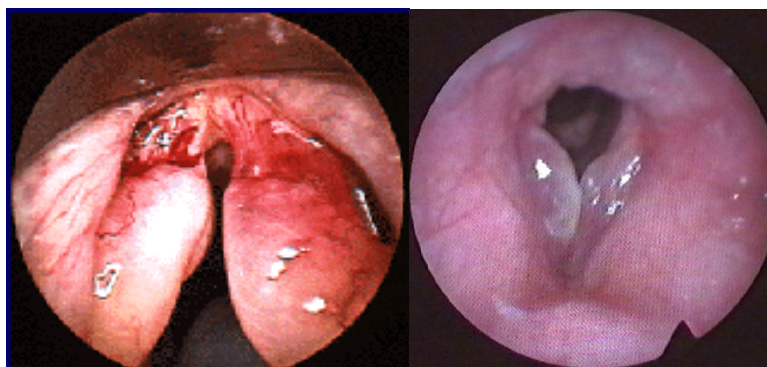


Figura 12: Fotos de edema de Reinke. (Extraído de (GHORAYEB, 2004) e (UPMC, 2004))

3.2.3.2.5 Laringite

A Laringite é uma inflamação das pregas vocais e da laringe. Pode ser decorrente de exposição a agentes nocivos (tabaco, drogas, álcool), agentes ambientais (poeira) ou abuso vocal. A laringite também pode decorrer de infecções respiratórias superiores que exercem

um efeito generalizado sobre a mucosa do trato respiratório, incluindo a laringe. Freqüência e perturbação de amplitude acima do normal são esperadas, assim como o ruído espectral aumentado. O principal sintoma vocal é a rouquidão (COLTON, 1996).

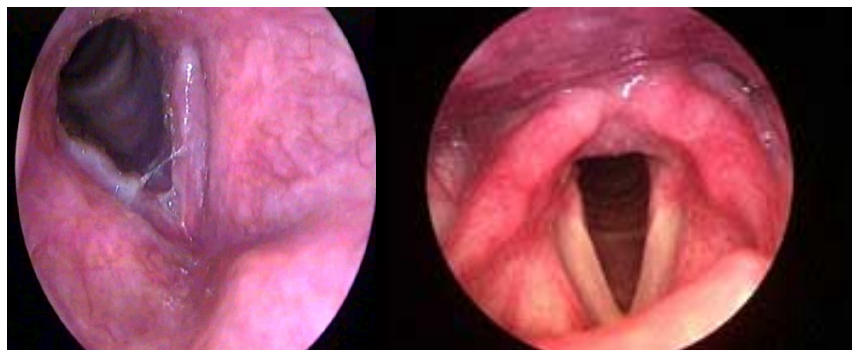


Figura 13: Fotos de laringite. (Extraído de (KAVANAGH, 2004))

3.2.3.2.6 Laringectomia

Uma das funções primárias da laringe humana é prevenir a aspiração de alimentos e líquidos na via aérea. A laringe contém três regiões das válvulas que permitem fechar a via aérea para prevenir a inalação de líquidos, alimentos, saliva, muco e quaisquer outros corpos estranhos que possam ser direcionados pela laringe para dentro da via aérea. Quando a laringe está tão comprometida por doença (como câncer avançado) ou trauma que não pode desempenhar com segurança seu papel de válvula, o paciente torna-se candidato a uma laringectomia, ou seja, a remoção total da laringe. Por isso, é criada uma abertura na traquéia através do qual o paciente deve desenvolver uma fonte de vocalização substituta, como uma laringe artificial ou fala esofágica (BOONE, 1994).

3.2.3.2.7 Câncer e Carcinoma

Câncer e carcinoma são neoplasias malignas de etiologia desconhecida. Carcinoma no sistema vocal é uma doença ameaçadora à vida e que requer manejo médico-cirúrgico. Os cânceres de lábio e intra-orais, raramente, contribuem para mudanças vocais, porém, podem exercer efeitos negativos óbvios sobre articulação e podem ser fatais. Lesões orais extensas, envolvendo a língua, ou câncer palatino ou velar, podem afetar seriamente a articulação e a

ressonância vocal. A rouquidão persistente é conhecida como um dos principais sintomas do câncer. Em geral, há 3 classificações de câncer laríngeo, dependendo da região da lesão supra-glótica, envolvendo estruturas como as pregas ventricular e ariepiglótica, a epiglope, as cartilagens aritenóideas e as paredes da hipofaringe; glótica, da comissura anterior até as extremidades do processo vocal das aritenóideas; e sub-glótica, envolvendo a cartilagem cricóidea e a traquéia. Os sinais do câncer da laringe podem incluir: nódulo no pescoço, um alargamento na laringe, aumento de volume no pescoço e rouquidão.

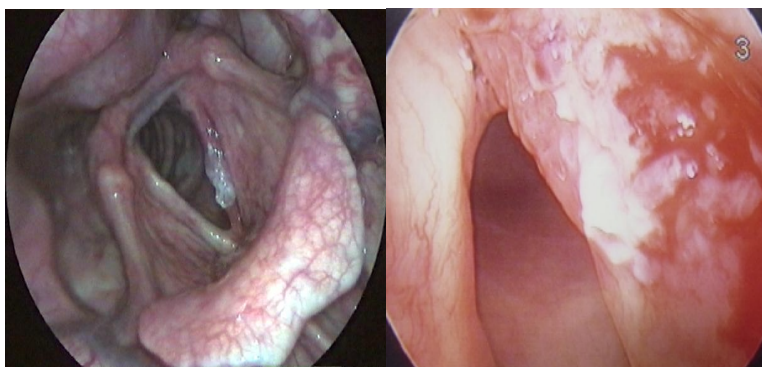


Figura 14: Foto de câncer. (Extraído de (GHORAYEB, 2004)).

3.2.3.2.8 Leucoplasia

Leucoplasia são lesões de acréscimo à membrana superficial da mucosa e sinais esbranquiçados que, muitas vezes, estendem-se por baixo da superfície para o espaço sub-epitelial. Embora as lesões sejam classificadas como tumores benignos, são consideradas lesões pré-cancerosas e devem ser observadas cuidadosamente. A causa mais comum é o fumo. Embora a leucoplasia sobre a língua exerça efeitos mínimos sobre a voz, a leucoplasia sobre as cordas vocais pode alterar drasticamente a voz. A massa de lesão acrescentada às pregas reduz a altura da voz e causa rouquidão. Pelo fato da leucoplasia ser aleatória em relação ao tamanho e localização, faz com que as pregas fiquem assimétricas, o que pode resultar em diplofonia, pois cada prega vibra em uma velocidade diferente devido ao seu tamanho ou massa diferentes. A leucoplasia que ocupa espaço na borda glótica pode impedir a aproximação das pregas, contribuindo para a sopro, intensidade reduzida e disfonia geral. O tratamento da leucoplasia é médico-cirúrgico e a terapia vocal apenas contribui para desenvolver a melhor voz possível (COLTON, 1996).

3.2.3.2.9 Paralisia

Lesões que afetam o nervo vago em alguma parte ao longo do seu curso, desde a base até a laringe podem resultar em uma paresia (fraqueza) ou paralisia nos músculos. O nervo vago é o principal nervo que supre a laringe através de fibras sensoriais que controlam todos os músculos da laringe. Este possui duas ramificações, os nervos laríngeos superior (NLS) ou recorrente (NLR). O nervo laríngeo superior controla os músculos cricotireóideos e o recorrente os músculos restantes da laringe. As lesões do NLS ou NLR podem afetar a posição das cordas vocais. Na paralisia NRL, a lesão pode ser unilateral ou bilateral e pode ser do tipo adutor ou abductor, dependendo dos músculos afetados. A etiologia da paralisia da prega vocal é variada. Os sintomas perceptuais mais comuns da paralisia unilateral são a sopro e a rouquidão. Ocasionalmente, a diplofonia pode estar presente. A paralisia bilateral do tipo adutor causará sopro severo ou afonia, porém, uma voz quase normal pode estar presente no tipo abductor. Acusticamente falando, uma maior aperiódicidade (*Jitter* e *Shimmer*), uma extensão de frequência reduzida, níveis de ruídos mais elevados e uma extensão de intensidade vocal reduzida estão presentes (BOONE, 1994), (COLTON, 1996).

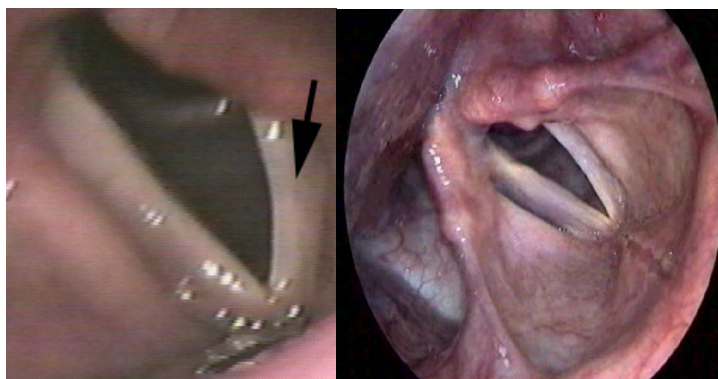


Figura 15: Fotos de paralisia. (Extraído de (UPMC, 2004)).

3.2.3.2.10 Granuloma

Há três tipos de granulomas, um resultando de intubação durante a cirurgia, um de trauma glótico por abuso/mau uso e um terceiro resultando do refluxo gástrico. Cada uma destas etiologias sugere um tratamento diferente. Os granulomas localizam-se no processo vocal da cartilagem aritenóidea ou na parede lateral da glote posterior. A prega vocal

membranosa não é afetada. Histologicamente, os granulomas consistem em geral de capilares proliferados, fibroblastos e fibras colágenas. A lesão pode ser unilateral ou bilateral. Os granulomas impedem o fechamento glótico apenas quando são muito grandes. Não há efeitos sobre as propriedades mecânicas de qualquer camada da prega vocal (COLTON, 1996).

Pacientes com granuloma podem experimentar disфонia severa, caracterizada por rouquidão, soprosidade e a necessidade de limpar a garganta frequentemente (BOONE, 1994).

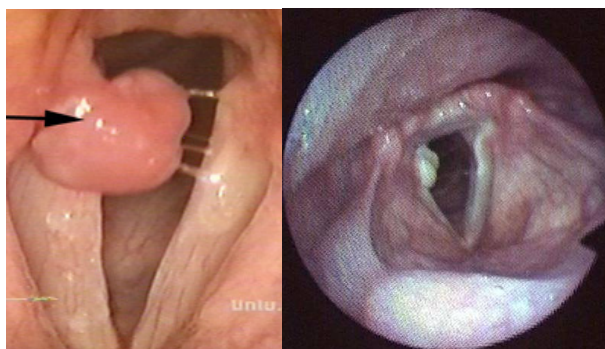


Figura 16: Foto de granuloma. (Extraído de (KAVANAGH, 2004)).

3.2.3.2.11 Sulcos-Fendas

Sulco vocal refere-se a uma condição na qual se observa uma invaginação ao longo da margem mediana superior das pregas vocais, cuja extensão longitudinal é variável, podendo em alguns casos parecer dividir a prega pela metade. Normalmente é resultante do rompimento de um cisto. Pacientes com sulco apresentam ronquidão, voz soprosa aparentemente devido ao fechamento incompleto das pregas. *Jitter* e *Shimmer* ligeiramente acima do normal. O sulco é uma patologia mais rara (COLTON, 1996).

3.2.3.2.12 Disфонia Funcional

Se o distúrbio da voz não tem como origem alguma patologia orgânica estrutural, então a causa é funcional, ou seja, devido ao mau uso do sistema fonador. Os pacientes podem aproximar as pregas de uma maneira frouxa, produzindo soprosidade ou de forma apertada, produzindo sintomas de aspereza ou constrição alterando assim a qualidade da voz.

3.3 TRANSFORMADA WAVELET PACKET

3.3.1 Introdução

Em 1807, Jean Baptiste Joseph Fourier, afirmou que qualquer forma de onda pode ser representada como uma soma infinita de ondas senoidais de diferentes frequências. Na época existiam serias dúvidas sobre a exatidão dos argumentos de Fourier por isso seu trabalho só foi publicado 15 anos depois (MACKENZIE, 2001). Hoje em dia a série de Fourier e a Transformada de Fourier estão onipresentes nas ciências, e são ideais para análise de ondas sonoras e de luz; porém não são eficazes para o estudo de fenômenos transitórios como a voz e outros sinais biológicos, pois os mesmos, muitas vezes, possuem características não estacionárias (AKAY, 1998). A análise do sinal utilizando-se a Transformada de Fourier (FT) é a mais adequada para análise de sinais estacionários (BENDAT, 1986). Ao aplicar-se a TF a um sinal, a informação de tempo é perdida e apenas a informação de frequência está presente. A FT é definida como:

$$FT(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi \cdot f \cdot t} dt \quad (3)$$

Em 1946, Dennis Gabor adaptou a Transformada de Fourier com uma técnica chamada de janelamento (*windowing*) do sinal. A adaptação de Gabor, conhecida como *Short Time Fourier Transform* (STFT), coloca o sinal em uma função de duas dimensões, tempo e frequência (RIOUL, 1991), conforme a Equação (4).

$$STFT(\tau, f) = \int_{-\infty}^{\infty} x(t) g(t - \tau) e^{-j2\pi \cdot f \cdot t} dt \quad (4)$$

Esta função pode ser descrita como a Transformada de Fourier do sinal $x(t)$, previamente limitada por uma função janela $g(t)$ ao longo do tempo τ . Este método assume estacionariedade do sinal dentro desta janela limitada $g(t)$. A STFT possui um compromisso entre tempo e frequência na análise do sinal. Esta Transformada fornece a informação de quando e em qual frequência ocorreu um evento no sinal. Contudo, esta informação tem precisão limitada pelo tamanho da janela. Uma vez escolhido o tamanho da janela, este será o mesmo para todas as frequências. Uma janela muito estreita resulta em uma boa resolução

temporal, mas ao mesmo tempo tem associado no espectro uma função com lóbulo principal largo, o que reduz a resolução entre frequências próximas.

Em 1927, Werner Heisenberg afirmou que a posição e a velocidade de um elétron não poderiam ser determinadas exatamente ao mesmo tempo, nem sequer na teoria. Em termos de sinais, isto significa que é impossível conhecer de forma simultânea a frequência exata e o momento exato em que ocorre esta frequência num sinal. Para conhecer a frequência, o sinal deve ser dilatado no tempo, e vice-versa (MACKENZIE, 2001). Por isso a resolução no tempo e na frequência não podem ser arbitrariamente pequenas, e seu produto fica limitado pela inequação de Heisenberg (RIOUL, 1991):

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \quad (5)$$

onde Δt e Δf são as resoluções no tempo e na frequência respectivamente.

Uma vez determinada a resolução no tempo e a resolução em frequência, esta fica fixa por todo o plano tempo-frequência. Desta forma, o sinal poderá ser analisado com uma boa resolução em tempo ou boa resolução em frequência, mas geralmente não em ambas.

Porém, muitos sinais exigem uma aproximação mais flexível, onde o tamanho da janela seja variável, determinando mais precisamente informações sobre tempo ou frequência dos mesmos (RIOUL, 1991).

A análise com a Transformada Wavelet (WT) ajudou a superar a limitação de janela fixa que existe na STFT, utilizando uma técnica de janelamento variável. A resolução Δt e Δf variam no plano tempo-frequência de forma a obter uma Análise Multi-Resolução (RIOUL, 1991), desta forma a WT serve como ferramenta de análise de fenômenos transientes, não-estacionários, ou variantes no tempo (BURRUS, 1998)

A WT descrita acima é denominada Transformada Wavelet Contínua, e é definida por:

$$CWT(s, \tau) = \gamma(s, \tau) = \int_{-\infty}^{\infty} x(t) \psi_{s,\tau}^*(t) dt \quad (6)$$

As Wavelets $\psi_{s,\tau}(t)$, são geradas a partir de uma única Wavelet, denominada Wavelet Mãe, escalonada e transladada:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (7)$$

onde s é o fator de escala, τ é o fator de translação e $1/\sqrt{s}$ é um fator para normalização da energia através das diferentes escalas. As Wavelets Mães podem assumir diversas formas. A escolha da Wavelet mãe é um fator importante, mas o critério para a escolha da Wavelet para uma determinada aplicação em particular ainda é uma área em pesquisa (BENTLEY, 1994).

Um espaço de funções é um espaço vetorial (de dimensão finita ou infinita) onde os vetores são funções. A Wavelet deve pertencer ao espaço de Hilbert, denotado por $L^2(\mathfrak{R})$ (ou simplesmente L^2). Neste espaço todas as funções $f(t)$, são quadraticamente integráveis:

$$\int_{-\infty}^{\infty} |f(t)|^2 dt < \infty \quad (8)$$

Em outras palavras, as funções devem ter energia finita. No caso das Wavelets a energia das mesmas deve ser constante para qualquer nível de dilatação:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1, \text{ para qualquer nível de dilatação.} \quad (9)$$

Além disso, as Wavelets devem satisfazer a condição de admissibilidade, que garante a existência da Transformada Inversa (SHENG, 1996) dada por:

$$c_h = \int_{-\infty}^{\infty} \frac{|\Psi(w)|^2}{|w|} dw < +\infty \quad (10)$$

onde $\Psi(w)$ é a Transformada de Fourier da Função Wavelet $\psi(t)$, e w é a frequência angular em radianos por segundo.

A condição de admissibilidade implica que a FT da Wavelet deve ser zero na componente de frequência $w=0$:

$$\left. |\Psi(w)|^2 \right|_{w=0} = 0 \quad (11)$$

O que, por sua vez, implica em que a Wavelet deve ser oscilatória com média zero no domínio tempo:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (12)$$

Existem três características da CWT que dificultam o seu uso de forma prática. A primeira é a redundância existente devido a que os fatores de escala e translação mudam de forma contínua (não discreta); a segunda é que existem infinitas Wavelets a serem consideradas, a terceira é que nem todas as transformadas inversas (ainda não definidas) têm solução analítica e mesmo com solução numérica seriam necessários algoritmos rápidos para explorar de forma eficiente a transformada.

Para resolver o problema da redundância da CWT, as Wavelets são discretizadas, no sentido de que as mesmas não são escaláveis e transladáveis de forma contínua, mas sim de forma discreta. Isto é possível modificando a Equação (7), (DAUBECHIES, 1992).

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right) \quad (13)$$

onde j e k são inteiros que representam os índices que controlam a dilatação e translação respectivamente, $s_0 > 1$ é o ‘passo’ fixo de dilatação, o fator de translação fixo é τ_0 . O efeito de discretizar as Wavelets desta maneira é que o espaço tempo-escala se torna discreto. Usualmente $s_0=2$ e $\tau_0=1$ obtendo a chamada amostragem dyadica, desta forma:

$$\begin{aligned} \psi_{j,k}(t) &= \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k2^j}{2^j}\right) = 2^{-j/2} \psi\left(\frac{2^{-j}t}{2^{-j}} - \frac{k2^{-j}}{2^{-j}}\right), \\ \psi_{j,k}(t) &= 2^{-j/2} \psi(2^{-j}t - k) \quad j, k \in \mathbb{Z} \end{aligned} \quad (14)$$

Quando as “Wavelets Discretas” $^1(\psi_{j,k})$ são utilizadas para transformar um sinal contínuo f , o resultado será uma serie de coeficientes denominados coeficientes de decomposição em Wavelets, $\gamma(j,k)$, que correspondem a DWT (SIMPSON, 1993).

$$\gamma(j,k) = \int_{-\infty}^{\infty} f(t) \psi_{j,k}^*(t) dt \quad (15)$$

A reconstrução do sinal f é possível se existem dois limites positivos A e B tais que (DAUBECHIES, 1992):

$$A\|f\|^2 \leq \sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2 \leq B\|f\|^2 \quad (16)$$

¹ As “Wavelets Discretas”, são contínuas no tempo, porém as dilatações e translações é que são discretas.

onde $\langle f, g \rangle$ significa o produto interno entre duas funções f e g , ambas pertencentes ao espaço L^2 , e definido em (WICKERHAUSER, 1994):

$$\langle f(t), g(t) \rangle = \int f(t) g^*(t) dt \quad (17)$$

E $\|f\|$ significa a norma, ou o ‘tamanho’ de uma função f , é definida em termos do seu produto interno (BURRUS, 1998):

$$\|f(t)\| = \sqrt{\langle f(t), f(t) \rangle} \quad (18)$$

O conjunto de $\psi_{j,k}$ constitui um *frame* de A a B. E esse conjunto é chamado *frame compacto (tight frame)* se os coeficientes A e B da Equação (16) forem iguais, neste caso, as Wavelets se comportam como bases ortogonais.

Para eliminar a redundância as Wavelet Discretas devem ser ortogonais a suas próprias dilatações e translações pela especial escolha da Wavelet Mãe, isso implica que:

$$\langle \psi_k, \psi_l \rangle = 0 \text{ somente se } k \neq l \quad (19)$$

A ortogonalidade não é essencial para representar um sinal, e em alguns casos as Wavelets precisam ser não-ortogonais, já que a redundância ajuda a reduzir a sensibilidade ao ruído (SHENG, 1996) ou implementam a chamada invariância ao deslocamento (*shift invariance*) da transformada (BURRUS, 1998).

Mesmo com as $\psi_{j,k}$ ao invés das $\psi_{s,\tau}$, é necessário um número infinito de translações e escalonamentos para calcular a WT. Porém, na prática, as translações das Wavelets dependem da duração do sinal que esta sendo transformado, não é necessário continuar transladando as Wavelets se não existe sinal fora de certos limites. Para determinar quantas escalas de dilatações são necessárias para analisar o sinal, a WT deve ser analisada de um ângulo diferente, como será visto nos parágrafos seguintes.

Pela condição de admissibilidade, observa-se que as Wavelets são em espectro, filtros passa-faixas. Pela teoria de Fourier é sabido que a compressão no tempo equivale à expansão e deslocamento em frequência

$$FT\{f(at)\} = \frac{1}{|a|} FT\left(\frac{w}{a}\right) \quad (20)$$

Uma compressão da Wavelet por um fator de 2, significa uma expansão no espectro pelo fator 2 e também deslocamento pelo fator 2. Utilizando esta propriedade se tem o seguinte gráfico (Figura 17):

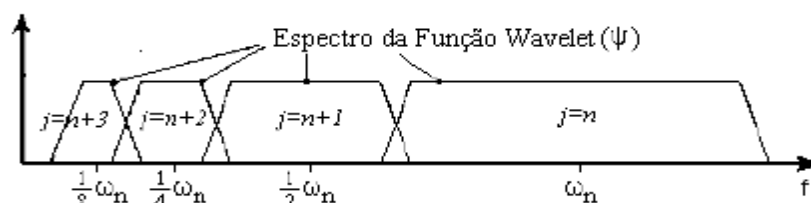


Figura 17: Espectro das Funções Wavelets

Se uma Wavelet pode ser vista como um filtro passa-banda, então uma série de Wavelets dilatadas pode ser vista como um banco de filtros passa-faixa. A razão entre a frequência central e a largura de banda deve ser constante para todas as Wavelets. Esta razão é chamada de fator de qualidade Q do filtro. No caso das Wavelets se fala de bancos de filtros com Q constante.

Cada vez que a Wavelet é contraída pelo fator 2, no espectro a largura de banda é reduzida a metade, ou seja é necessário um número infinito de Wavelets para abranger o espectro todo, até o nível DC. Uma solução para este problema é simplesmente não tentar abranger todo o espectro com funções Wavelets, mas sim adicionar uma função cujo espectro cubra as baixas frequências desde zero hertz até um nível determinado de dilatação, denominada Função de Escalonamento (*Scaling Function*) e que é um filtro passa-baixas. A Função de Escalonamento foi introduzida por (MALLAT, 1989) e devido a sua natureza passa-baixa ele é referido como um Filtro de Média (*averaging filter*).

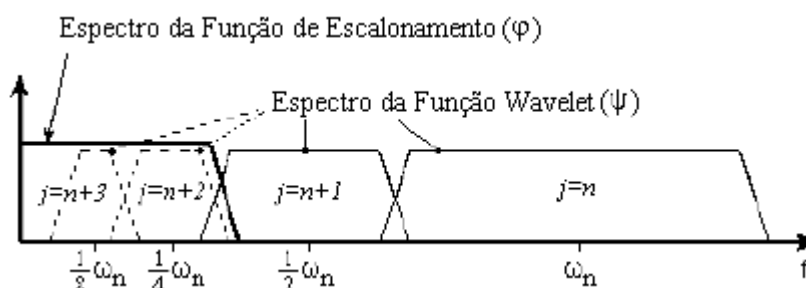


Figura 18: Espectro das Funções Escalonamento

A Função de Escalonamento pode ser decomposta nas suas componentes Wavelets:

$$\varphi(t) = \sum_{j,k} \gamma(j,k) \psi_{j,k}(t) \quad (21)$$

Desta forma podemos analisar o sinal utilizando uma combinação de Funções Wavelets com Funções Escalonamento.

As Funções de Escalonamento abrangem o espectro desde zero hertz até uma certa escala j . O restante do espectro é das Funções Wavelets. Com este artifício foi limitado o número de Funções Wavelets de um valor infinito para um valor finito, graças a Função de Escalonamento. Quando se utiliza Funções de Escalonamento ao invés de Funções Wavelets, perde-se informação mas não no sentido de representação, pois ainda é possível a reconstrução perfeita do sinal, mas sim no sentido que se descarta possíveis escalas de Wavelets que poderiam trazer mais informação do sinal, por isso que o tamanho da Função de Escalonamento no espectro é um parâmetro de projeto muito importante.

A Função de Escalonamento não satisfaz a condição de admissibilidade, mas sim:

$$\int_{-\infty}^{\infty} \varphi(t) dt = 1 \quad (22)$$

A Função de Escalonamento, junto com a série de Funções Wavelets dilatadas, podem ser vistas como um banco de filtros.

3.3.2 A Transformada Wavelet Discreta (DWT) e a Análise Multi-Resolução (MRA)

A análise de um sinal, passando por um banco de filtros não é uma idéia nova, tem sido utilizada por muitos anos com o nome de codificação subbanda (*subband coding*). Uma maneira de implementar a codificação subbanda é dividir o espectro do sinal em duas partes iguais, uma passa alta e uma passa baixas. Usualmente o número de bandas é limitado pelo poder computacional que se dispõe. A vantagem, é que pode-se dividir o espectro do sinal, utilizando apenas 2 filtros, como mostra a Figura 19:

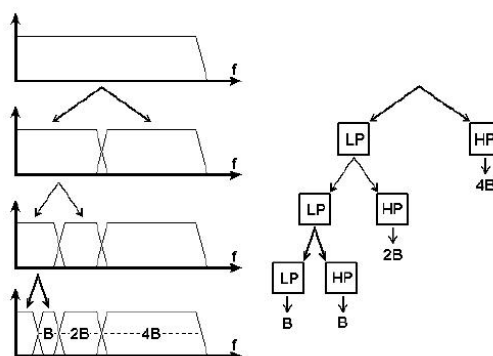


Figura 19: Codificação Subbanda

Como os sinais analisados pertencem a L^2 , os mesmos possuem energia limitada. Desta forma o filtro passa-alta utilizado na codificação subbanda pode ser substituído por um passa-faixas devido a largura limitada do sinal. Em outras palavras, pode-se realizar a codificação subbanda pela passagem do sinal por um banco de filtros passa-bandas cada um deles com a metade da largura de banda do anterior, e um filtro passa baixa.

Em 1986, Stéphane Mallat, um antigo aluno de Yves Meyer (um dos propositores da teoria de Wavelets), no seu trabalho de doutorado, vinculou a teoria de Wavelets à literatura existente sobre codificação subbandas e filtros QMF e mostrou que as Wavelets estão implícitas naquilo que a comunidade de processamento de sinais chama de Análise Multi-Resolução (MRA) (SHENG, 1996).

Observa-se através da Equação (21) que é possível expressar a Função de Escalonamento em termos de Wavelets que vão desde menos infinito até uma certa escala j . Se for somado o espectro de uma Função Wavelet ao espectro da Função Escalonamento, obter-se-ia uma nova Função de Escalonamento, com um espectro do dobro do tamanho da primeira. O efeito dessa adição é que pode-se expressar a primeira Função de Escalonamento em função da segunda. Formalmente é chamado formulação Multi-Resolução (WICKERHAUSER, 1994) ou *two-scale relation* (SHENG, 1996).

$$\varphi\left(\frac{1}{2^{j+1}}t\right) = \sum_k h_j(k)\varphi\left(\frac{1}{2^j}t - k\right) \quad (23)$$

Segundo a Equação (23) a Função de Escalonamento numa certa escala $j+1$ pode ser representada em termos de Funções de Escalonamento de escalas menores (escalas j menores

= mais detalhes), em outras palavras, funções mais dilatadas podem ser compostas por funções menos dilatadas.

A primeira Função de Escalonamento substitui um conjunto de Wavelets e desta forma pode-se expressar as Wavelets em termos de Funções de Escalonamento transladadas e na escala seguinte. Mais especificamente, para a Wavelet no nível j :

$$\psi\left(\frac{1}{2^{j+1}}t\right) = \sum_k g_j(k)\varphi\left(\frac{1}{2^j}t - k\right) \quad (24)$$

Desde que $f(t)$ possa ser representado em termos de Wavelets transladas e dilatadas até uma escala $j+1$, então $f(t)$ pode também ser expressa em termos das funções Escalonamento dilatadas e transladadas até uma escala j :

$$f(t) = \sum_k \lambda_j(k)\varphi\left(\frac{1}{2^j}t - k\right) \quad (25)$$

Para ser consistente com a notação empregada deve-se neste caso falar de Funções de Escalonamento somente com dilatações e translações discretas.

Se na Equação (25) for utilizada escala $j+1$, será necessário adicionar Wavelets para manter o mesmo nível de detalhes. Desta forma a função pode ser representada também por:

$$f(t) = \sum_k \lambda_{j+1}(k)\varphi\left(\frac{1}{2^{j+1}}t - k\right) + \sum_k \gamma_{j+1}(k)\psi\left(\frac{1}{2^{j+1}}t - k\right) \quad (26)$$

Se a Função de Escalonamento e a Função Wavelet forem ortogonais (ou *frame* compacto) então os coeficientes $\lambda_{j+1}(k)$ e $\gamma_{j+1}(k)$ são calculados pelo produto interno:

$$\lambda_{j+1}(k) = \langle f(t), \varphi_{j,k}(t) \rangle \quad (27)$$

$$\gamma_{j+1}(k) = \langle f(t), \psi_{j,k}(t) \rangle \quad (28)$$

Substituindo a Função de Escalonamento e a Função Wavelet de (27) e (28) por versões dilatadas e transladadas de (23) e (24), e manipulando as equações, tendo em mente que o produto interno é uma integral, chega-se ao importante resultado (WICKERHAUSER, 1994):

$$\lambda_{j+1}(k) = \sum_m h(m-2k)\lambda_j(m) = c_{j+1}(k) \quad (29)$$

$$\gamma_{j+1}(k) = \sum_m g(m-2k)\gamma_j(m) = d_{j+1}(k) \quad (30)$$

As Equações (29) e (30) mostram que os coeficientes das Funções Wavelets e Escalonamento numa certa escala podem ser determinadas pela soma ponderada dos coeficientes nas escalas pr evias; por outro lado,   sabido que uma soma ponderada deste tipo, caracteriza um filtro digital do tipo FIR (*Finite Impulse Response*). Os coeficientes $\lambda_j(k)$ (ou coeficientes de aproxima  o $c_j(k)$) s o da parte passa-baixas do espectro de sinal e os fatores $h(k)$ formam os coeficientes do filtro passa-baixas; por outro lado os coeficientes $\gamma_j(k)$ (ou coeficientes de detalhes $d_j(k)$) s o da parte passa-altas do espectro de sinal, os fatores $g(k)$ formam os coeficientes do filtro passa-altas. Isso significa que as Equa  es (29) e (30) juntas, formam um estagio no banco de filtros iterativos, e nos referiremos a $h(k)$ e $g(k)$ s o referidos como os coeficientes do filtro de escalonamento e coeficientes do filtro wavelet.

Nas Equa  es (29) e (30) pode-se ver a propriedade da decima  o por dois (guarda-se uma amostra de cada duas). Cada vez que se itera com o banco de filtros digital, o n mero de amostras do estagio seguinte cai pela metade, at  o caso extremo de ter apenas 1 amostra. Isto mostra claramente quando   o ponto de parada das itera  es e determina a largura do espectro da fun  o de escalonamento. Normalmente as itera  es param quando o n mero de amostras se torna t o pequeno quanto o tamanho do filtro h ou do filtro g (o tamanho do maior entre eles).

A opera  o de decima  o pelo fator 2,   definida (pictoricamente) por (OPPENHEIM, 1989):

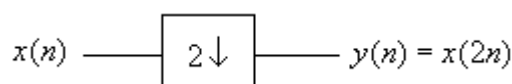


Figura 20: Simbologia da decima  o

Desta forma nota-se que as Equa  es (29) e (30), desenvolvem a convolu  o e a decima  o, conforme a Figura 21:

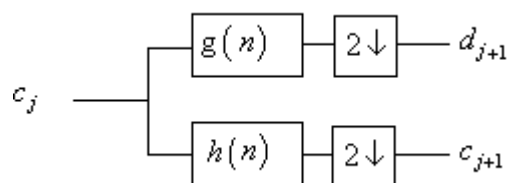
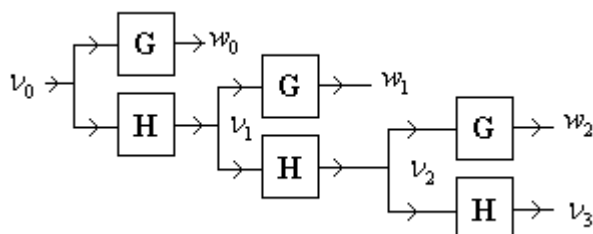


Figura 21: Análise com banco de filtros em duas bandas

A implementação da Transformada Wavelet Discreta (DWT) pode ser interpretada como uma decomposição em sub-bandas, ou seja, o processo da transformada é iterativo com sucessivas decomposições nos coeficientes de aproximação. Desta forma, obtém-se uma árvore de decomposição da Wavelet. Esta decomposição em sub-bandas pode ser vista como uma estrutura de banco de filtros e esta pode ser implementada como mostra a Figura 22. O sinal x pertence ao espaço ν_0 ; a medida que se aumenta o nível de dilatação j ao aplicar os operadores G e H , os sinais resultantes passam a pertencer a outros espaços, por exemplo Gx pertence ao espaço w_0 , Hx pertence ao espaço ν_1 , HHx pertence ao espaço ν_2 , e assim sucessivamente. Os espaços w e ν são subconjuntos do espaço L^2 .



Sinal x (no espaço ν_0)		nível 0, j
Hx		nível 1, $j+1$
Gx		
HHx	GHx	nível 2, $j+2$
HHHx		GHHx

Figura 22: Decomposição em sub-bandas, mostrando os espaços gerados.

A projeção de um sinal na base ortogonal $\phi(t)$ forma o subespaço ν_0 que é uma aproximação “borrada” do sinal x (SHENG, 1996).

O complemento ortogonal de v_j em v_{j+1} é definido como w_j . Desta forma o espaço das Funções Wavelets é a diferença entre dois espaços consecutivos da Função Escalonamento, como mostra a Figura 23:

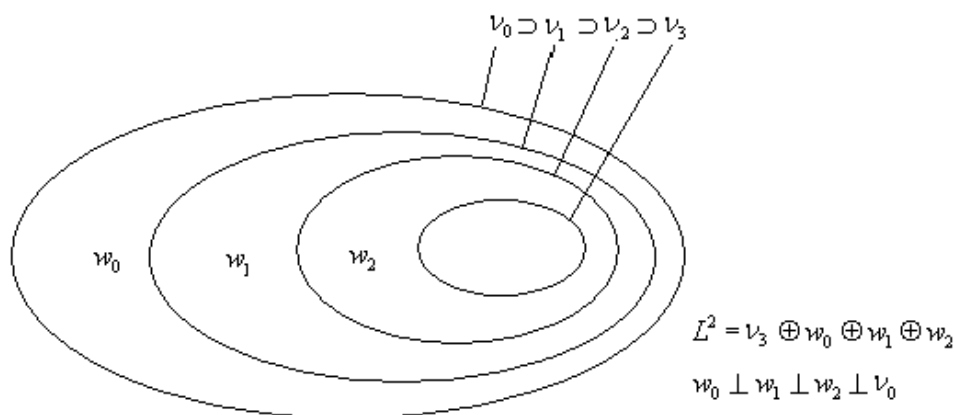


Figura 23: Espaços de sinais gerados pelas Funções de Escalonamento.

A função Wavelet $\psi(t)$ está relacionada a um filtro FIR passa-alta (g), a qual produz os coeficientes de detalhes da decomposição (d_{jk}). A função de escalonamento $\phi(t)$ está associada com os coeficientes de aproximação da decomposição (c_{jk}), através de um filtro FIR passa-baixa (h).

Os filtros digitais utilizados na DWT Multi-Resolução, chamados *Quadrature Filter* (QF), tem como característica filtrar e decimar o sinal, respectivamente e também permitir a reconstrução do sinal. Através do par de filtros de Quadratura (QMF) é possível obter uma cobertura aproximada do espectro de frequências (BENTLEY, 1994).

3.3.3 Transformada Wavelet Packet (WPT)

A WPT é uma generalização do conceito da DWT, na qual a resolução tempo-freqüência pode ser escolhida de acordo com o sinal, embora isto seja feito dentro dos limites do princípio de incerteza de Heisenberg (KARRAKCHOU, 1998).

A Transformada Wavelet Discreta (DWT) é equivalente à implementação de um banco de filtros não uniformes com uma melhor resolução em freqüência em baixas freqüências e melhor resolução em tempo para as altas freqüências. Na Análise Multi-Resolução da Wavelet Discreta, o sinal é dividido em coeficientes de aproximação e detalhes, como visto na Figura 22. Apenas os coeficientes de aproximação são novamente divididos em coeficientes de aproximação e detalhes, gerando o segundo nível de decomposição. E assim pode ser feito N vezes, onde $2N$ é o tamanho do sinal amostrado. Já no primeiro nível de decomposição da Wavelet Packet, os coeficientes de detalhes também são divididos em coeficientes de aproximação e detalhes, decompondo o sinal em forma de árvore binária, chamada árvore Wavelet Packet. Desta forma, tem-se o sinal representado em bandas de freqüências com diferentes resoluções (COIFMAN, 1998).

De forma análoga às demais transformadas, o cálculo dos coeficientes da WPT é a seqüência resultante do produto interno de um sinal $x(t) \in L^2(\mathfrak{R})$ com as funções Wavelet Packet $w_{jkf}(t)$:

$$\lambda_{jf}(k) = \langle x, w_{jkf} \rangle = \int_{\mathfrak{R}} x(t) 2^{-j/2} w_f(2^{-j}t - k) dt \quad j, k \in \mathbb{Z}, e f \geq 0 \quad (31)$$

onde λ_{jf} são os coeficientes da expansão de $x(t)$ nas funções Wavelet Packet, $w_{jkf}(t)$ são as funções Wavelet Packet básicas, j é o índice de escala (ou dilatação), k é o índice de posição (ou translação), e f é o índice de freqüência (ou também denominado por n , “número de oscilações”, dependendo da literatura).

Ao contrário da DWT, que é função de duas variáveis (escalamento e translação), a Wavelet Packet é uma função de três variáveis. O conjunto das Waveletes Packet é construído de forma semelhante às anteriores transformadas, só que a decomposição em $L^2(\mathfrak{R})$ com múltiplas escalas também é possível. A função transladada, dilatada e normalizada das funções Wavelet Packet é obtida pela equação:

$$w_{j kf}(t) = 2^{-j/2} w_f(2^{-j} t - k) \quad (32)$$

Seja H e G o par conjugado de filtros de quadratura (*Quadrature Filters*). A biblioteca de funções Wavelet Packet é construída com um algoritmo recursivo, e as funções $w_f(t)$ são definidas, considerando a escala fixa (WICKERHAUSER, 1994), conforme as Equações (33) e (34):

$$w_{2f}(f) = \sum_k h(k) w_f(2t - k) = H w_f \quad (33)$$

$$w_{2f+1}(f) = \sum_k g(k) w_f(2t - k) = G w_f \quad \text{para } f=0,1,2,3... \quad (34)$$

onde h e g são os Filtros de Quadratura H e G (operadores que convoluem e decimam por 2). A coleção destas funções com $f = 0, 1, 2, \dots$, é chamada de Wavelet Packet associada aos filtros de quadratura H e G . A função inicial w_0 para $f=0$ é a própria função de Escalonamento $\phi(t)$. Da mesma forma, w_1 para $f=1$ é a função Wavelet ψ_1 da DWT (WICKERHAUSER, 1994). Diferentemente da DWT, a Transformada Wavelet Packet utiliza mais operadores G em cada nível de dilatação.

A decomposição Wavelet Packet pode ser implementada aplicando os filtros H e G na seqüência de coeficientes λ da Equação (31), satisfazendo a relação recursiva (WICKERHAUSER, 1994):

$$\lambda_{j+1, 2f}(k) = \sum_{k \in Z} h(k) \lambda_{jf}(2j - k) = H \lambda_{jf}(k) \quad (35)$$

$$\lambda_{j+1, 2f+1}(k) = \sum_{k \in Z} g(k) \lambda_{jf}(2j - k) = G \lambda_{jf}(k) \quad (36)$$

Esta implementação é mostrada na Figura 24, onde são destacados os espaços de funções para cada nível de dilatação j . O sinal x pertence ao espaço ν_0 ; a medida que se aumenta o nível de dilatação j ao aplicar os operadores G e H , os sinais resultantes passam a pertencer a outros espaços, por exemplo Gx pertence ao espaço w_0 , Hx pertence ao espaço ν_1 , GGx pertence ao espaço w_{01} , GHx pertence ao espaço w_{00} , e assim sucessivamente. Os espaços w e ν são subconjuntos do espaço L^2 .

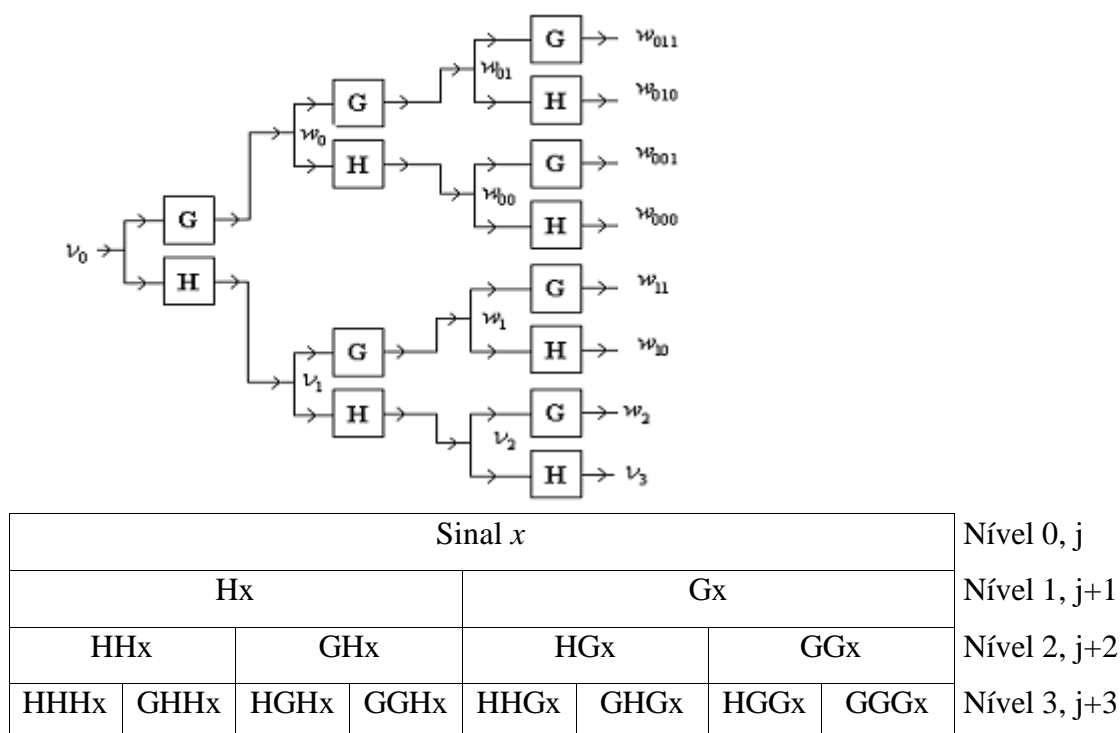


Figura 24: Decomposição em sub-bandas WPT, mostrando os espaços gerados.

3.3.4 Algoritmo *Best Basis* (BAA) e *Best Basis Tree* (BBT)

As árvores de decomposição Wavelet Packet, contém diversas bases admissíveis para representar o sinal (diversas possibilidades de reconstrução do sinal a partir de decomposição escolhida). O BBA é um algoritmo elaborado para escolher a melhor base para representar um sinal. Esta escolha é baseada na análise de uma função discriminante que avalia o custo de utilizar tal o qual base para representar o sinal; este discriminante recebe o nome de Função Custo. Desta forma ao escolher uma determinada base todas as sub-bases derivadas dela são podadas da árvore de decomposição. A propriedade de ortogonalidade garante que mesmo podando a árvore e obtendo-se uma sub-árvore, ainda assim existe uma base ortogonal com a qual o sinal original pode ser reconstruído (KARRAKCHOU, 1998).

Seja β uma biblioteca de bases, e A uma base qualquer pertencente a β . Seja também, uma seqüência de números, x_a , resultantes da projeção de x na base A . O custo de representar x na base A é definido por $M(A.x)$. A BBT é definida para x , de forma a ser escolhida uma base A na qual $M(A.x)$ é mínimo. (WICKERHAUSER, 1994)

Existem diversas Funções Custo encontradas na literatura; as denominadas “aditivas”, por definição satisfazem (37) (COIFMAN, 1992):

$$M(x) = \sum_i M(x_i) \quad \text{e} \quad M(0) = 0 \quad (37)$$

$M(x)$ é a função custo da seqüência de números x , i é o índice da seqüência .

Alguns exemplos são de Funções Custo aditivas são:

a) Entropia de Shannon

$$M_1(x_i) = -\frac{|x_i|^2}{|x|^2} \log_2 \frac{|x_i|^2}{|x|^2} \quad \text{onde } |x|^2 = \sum_i |x_i|^2 \quad (38)$$

$$M_1(x) = -\sum_i \frac{|x_i|^2}{|x|^2} \log_2 \frac{|x_i|^2}{|x|^2} \quad (39)$$

A entropia de Shannon fornece uma forma de estimar o número mínimo de bits necessários para codificar a seqüência x , baseados na freqüência de aparecimento dos seus elementos. Por definição $(0)\log(0)=0$.

b) Concentração em l^p

$$M_2(x_i) = \left(\|x_i\|_p\right)^p, \quad \forall x \in l^p, \quad 1 < p < 2 \quad (40)$$

$$M_2(x) = \sum_i \left(\|x_i\|_p\right)^p \quad (41)$$

Se duas seqüências possuem igual energia, ou seja $\|u\| = \|v\|$, mas $M_2(u) < M_2(v)$, então u possui maior concentração de energia em menos elementos.

c) Logaritmo da Energia

$$M_3(x_i) = \log|x_i|^2 \quad (42)$$

$$M_3(x) = \sum_i \log|x_i|^2 \quad (43)$$

Por definição $\log(0) = 0$.

d) *Threshold*

Se escolhe um limiar arbitrário ε que limita o número de coeficientes necessários para representação do sinal x na base que está sendo considerada.

$$M_4(x_i) = \begin{cases} 1, & \text{se } x_i > \varepsilon \\ 0, & \text{se } x_i \leq \varepsilon \end{cases} \quad (44)$$

$$M_4(x) = N, \text{ onde } N \text{ é o número de vezes em que } x_i > \varepsilon. \quad (45)$$

As Funções Custo “não aditivas” são utilizadas maiormente para calcular uma base próxima a *Best Basis (Near Best Basis)*. Mais detalhes em (TASWELL,1994).

Na árvore de decomposição WP existe sempre a situação da Figura 25, onde os coeficientes x_a e x_b são chamados nós filhos do nó pai x , e representam as projeções de x nas bases A e B respectivamente.

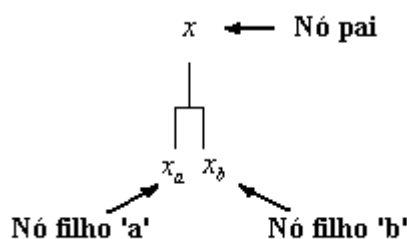


Figura 25: Diagrama de um nó pai e nós filhos

Para aplicar o BBA se desenvolve a WPT até o nível de dilatação j que se deseja. Depois se calcula o valor funcional M (função custo escolhida) em todas as projeções do sinal x . Desta forma, por exemplo, $M(x_a)$ é o valor funcional da projeção do sinal x na base A e $M(x_b)$ é o valor funcional da projeção do sinal x na base B .

No último nível de dilatação é avaliado o custo de armazenar a informação de cada par de nós filhos e ao mesmo tempo é avaliado o custo de armazenar apenas os seus correspondentes nós pais, que estão a um nível acima. Se for menos custoso armazenar a informação dos nós filhos então eles permanecem, por outro lado se o nó pai for menos ‘custoso’ ele permanecerá e os filhos serão podados da árvore.

Como mostra a Figura 26, se $M(x) > M(x_a) + M(x_b)$ o nós filhos permanecem, se $M(x) \leq M(x_a) + M(x_b)$ os nós filhos são podados, e o $M(x)$ assume o valor $M(x_a) + M(x_b)$ para

futuras comparações. Isto é feito até atingir o topo da árvore ($j=0$), formando então uma base completa de acordo com a função custo escolhida. A esta árvore resultante dá-se o nome de *Best Basis Tree (BBT)*.

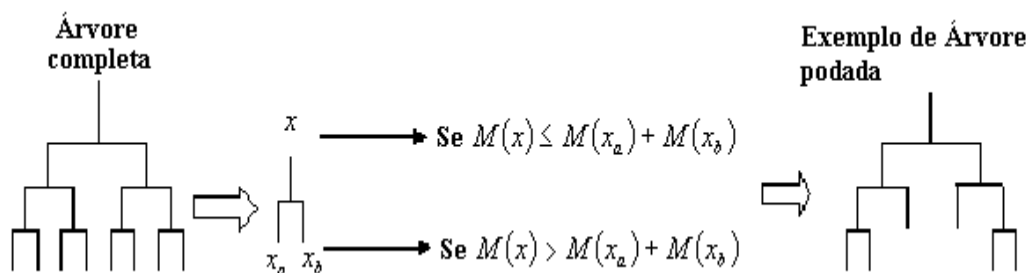


Figura 26: Ilustração do princípio de seleção *Best Basis*.

Originalmente, o BBA foi elaborado para compressão de dados, ou seja, guarda-se apenas as decomposições que gastam menos bits para sua representação. Porém; a morfologia da BBT obtida, depende do sinal original x , e esta característica faz com que a aplicação do algoritmo BBA seja uma forma de pré-classificação, conforme discutido em (PARRAGA, 2002).

3.4 REDES NEURAIS ARTIFICIAIS

3.4.1 Introdução

Em (HAYKIN, 2001) encontra-se uma definição para rede neural artificial (ANN). É um processador maciçamente e paralelamente distribuído constituído de unidades de processamento simples (neurônios), que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos:

- a) O conhecimento é adquirido pela rede a partir de seu ambiente através de um procedimento de aprendizagem.
- b) Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

Algumas características que tornam atraente o uso de ANN são:

- a) Habilidade de tratar sistemas não-lineares. Esta habilidade das redes neurais as tem tornado útil para modelar controladores não lineares.
- b) Tolerância a falha e processamento paralelo. Devido a que nos modelos de redes neurais normalmente se tem muitos neurônios interconectados por pesos com capacidade de adaptação. Estes neurônios podem ser arranjados em uma estrutura paralela. Por causa deste paralelismo, a falha de alguns neurônios não causa efeitos significantes para a performance de todo o sistema.
- c) Adaptabilidade e Aprendizagem. Graças a estas características os modelos de redes neurais podem lidar com dados imprecisos e situações não totalmente definidas.
- d) Treinamento e Generalização. Uma rede treinada de maneira razoável tem a habilidade de generalizar quando é apresentada a entradas que não estão presentes em dados já conhecidos por ela.

Os primeiros estudos das redes neurais se remontam aos trabalhos de Hipócrates. Mas a era moderna das redes neurais começou com os trabalhos pioneiros do psiquiatra e neuroanatomista Warren S. McCulloch e do matemático Walter Pitts em 1943, onde descrevem um cálculo lógico de redes neurais que unificava os estudos de neurofisiologia e da lógica matemática. Com o modelo de neurônio por eles proposto seria possível (com um número adequado de neurônios) realizar qualquer função computável. Von Neumann se inspirou no modelo de neurônio para desenvolver o computador EDVAC (*Electronic Discrete Variable Automatic Computer*) nos anos 1943 a 1946 (KOVÁCS,1996).

Embora não diretamente relacionados com Redes Neurais, na década de 40 vários conceitos importantes e teorias foram sendo desenvolvidas em outros campos de aplicação, mas que acabaram dando bases mais sólidas para uma teoria melhor sobre ANN. Mas sem dúvida a primeira grande onda de entusiasmo com as redes neurais surgiu com os trabalhos de Rosenblatt. O Perceptron (que era o modelo proposto) consegue aprender a classificação de padrões a partir de exemplos. Assim, pela primeira vez, tinha-se a configuração de um modelo de aprendizado e percepção com resultados concretos. Só que para funcionar adequadamente, os padrões a serem classificados devem estar suficientemente separados no espaço de entradas para garantir que nesse espaço o delimitador entre classes seja um hiperplano. Isto implica em que as classes devem ser linearmente separáveis.

Em 1960 Widrow e Hoff introduzem o algoritmo LMS e o usaram para formular o *Adaptative Linear Element* (ADALINE) e posteriormente o *Many Adaptative Linear Elements* (MADALINE), que são modelos semelhantes ao Perceptron de Rosenblatt, porém com outro tipo de processo de aprendizagem (WIDROW, 1990).

Em (MINSKY, 1988) se fez uma grande crítica aos Perceptrons, mostrando que estes seriam incapazes de resolver problemas simples, como por exemplo, executar a função XOR (que é não linearmente separável). Também se faz uma crítica pessoal e errônea sobre o modelo MLP, inferindo que teria a mesma dificuldade em classificar classes não linearmente separáveis. A consequência desta afirmação desencorajou o investimento na área de ANN durante aproximadamente dez anos. Poucos pesquisadores continuaram a publicar artigos sobre ANN (HAYKIN, 2001).

Em 1972, Kohonen, Anderson e Nakano, de modo independente, introduzem a idéia de “matriz por correlação”, que é uma matriz que multiplicada por um vetor em teste, permite

saber se o vetor de teste é semelhante a um padrão “memorizado”, que se encontra intrinsecamente na matriz. Desta forma implementa-se o que é conhecido pela comunidade de redes neurais como “memória associativa”.

Na década de oitenta um grupo de pesquisadores publica o artigo clássico (RUMELHART,1986) do PDP (*Parallel Distributed Processing*) onde é utilizado o algoritmo "*error back-propagation*" em uma rede de Perceptrons Multicamadas (MLP). A partir deste marco observa-se uma explosão de aplicações de ANNs, nos mais variados campos do conhecimento.

Em (WIDROW, 1990) é feita uma análise (até então) dos últimos 30 anos desde o aparecimento dos algoritmos chamados “de mínima perturbação”, entre outros são:

- a) Regra do Perceptron
- b) Algoritmo LMS
- c) Algoritmo do Madaline
- d) A técnica *Backpropagation*

Existem vários tipos de ANN desenvolvidas ao longo do tempo, elas são classificadas entre outras coisas, pela sua topologia, arquitetura, modelo de neurônio e algoritmo de treinamento. A topologia diz respeito a como os neurônios estão interconectados (todos interconectados, ou não). A arquitetura diz respeito ao número de camadas (única ou múltiplas), número de nós por camada e tipo de conexão (sem ou com realimentação) . Sobre o modelo de neurônio, existem vários, neste trabalho foi utilizado o modelo do Perceptron. Dependendo do modelo de neurônio a filosofia de aprendizado muda totalmente, é o caso das Redes de Hopfield, as Redes de Kohonen, aprendizado por reforço, redes competitivas, redes convolutivas. O algoritmo de treinamento diz respeito a como os parâmetros são ajustados (HAYKIN,2000).

Em existem algoritmos supervisionados (*Backpropagation, Perceptron Rule, LVQ*), não-supervisionados (SOM) e híbridos (Redes de Base Radial). Os algoritmos supervisionados também são conhecidos como “algoritmos com um professor”, a figura do professor, conceitualmente, é uma entidade que tem conhecimento sobre o ambiente, com esse conhecimento sendo representado por um conjunto de exemplos entrada-saída.

Entretanto, o ambiente é desconhecido para a rede neural de interesse. Ambos, o professor e a rede são expostos a um vetor de treinamento (i.e., exemplo) retirado do ambiente. Em virtude do seu conhecimento prévio, o professor é capaz de fornecer à rede neural uma resposta desejada para aquele vetor de treinamento. A resposta desejada representa a ação ótima a ser realizada pela rede neural. Os parâmetros da rede são ajustados sob a influência combinada do vetor de treinamento e do chamado sinal de erro, que é a diferença entre a resposta desejada e a resposta real da rede. Este ajuste é realizado passo a passo, iterativamente, com o objetivo de fazer emular o professor, no sentido estatístico. Desta forma o conhecimento do ambiente é transferido do professor para a rede neural através de treinamento, quando isto acontece pode-se dispensar do professor e deixar a rede neural lidar com o ambiente por si mesma.

Já na aprendizagem não supervisionada, também conhecida como, aprendizagem auto-organizada, ou também como “aprendizagem sem um professor” como o nome implica, não há um professor para supervisionar o processo de aprendizagem. Isto significa que não há exemplos rotulados da função a ser aprendida pela rede neural. Este tipo de redes utiliza o que se conhece como aprendizagem competitiva. Os neurônios da rede competem pelo “oportunidade” de responder às características contidas nos dados de entrada. Na forma mais simples a rede opera de acordo com uma estratégia do tipo “o vencedor leva tudo” nesta estratégia o neurônio que contém seus pesos o mais semelhantes ao espaço de entrada “ganha” a competição e se torna “ligado”, e os demais se tornam desligados. Depois do treinamento, cada vetor que descreve o ambiente produz a ativação de algum neurônio na rede, desta forma é mapeado o ambiente para um outro espaço que o descreve.

3.4.2 Reconhecimento de Padrões

O termo "reconhecimento de padrões" foi introduzido no início da década de 60. Mas sem dúvida a definição mais aceita, foi proposta por Duda e Hart, em 1973, como sendo "campo interessado no reconhecimento por máquinas de regularidades significativas em ambientes ruidosos ou complexos". Dentre algumas abordagens de reconhecimento de padrões, as mais comuns são as simbólicas e as numéricas. A abordagem simbólica define componentes primitivos (símbolos) e regras de como estes componentes serão agrupados para formar os padrões desejados. Já a abordagem numérica define que todo padrão pode ser

representado por um vetor numérico; quanto maior a similaridade entre os vetores, maior a chance de pertencerem ao mesmo tipo (classe).

Diversas técnicas têm sido amplamente cobertas pela literatura. Neste trabalho, foram utilizados métodos numéricos para classificar padrões.

3.4.3 Modelo do Neurônio dos Perceptrons Muticamadas (MLP)

Um neurônio é a unidade de processamento de informação de uma rede neural. Na Figura 27 identificam-se os três elementos básicos do neurônio do MLP, o qual tem inspiração no neurônio biológico:

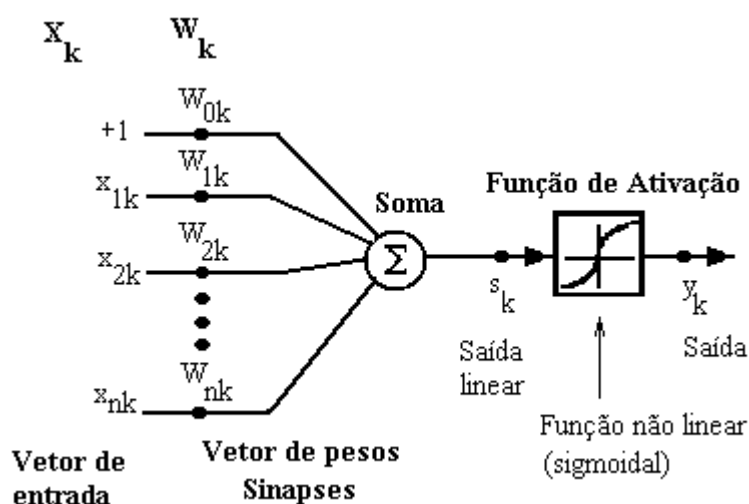


Figura 27: Modelo de neurônio utilizado. (Adaptado de (WIDROW, 1990)).

a) O Conjunto de Sinapses ($w_{0k}, w_{1k}, \dots, w_{nk}$): Cada uma caracterizada pelo “peso” ou força própria. Especificamente o sinal de entrada x_j , na entrada da sinapse j conectada ao neurônio k é multiplicada pelo peso sináptico w_{kj} . Ao contrario das sinapses do cérebro, as sinapses deste modelo podem adquirir valores positivos e negativos. O conjunto de sinapses também recebe o nome de vetor de pesos.

b) O Soma (ou Somador): Soma os sinais de entrada (vetor x), ponderados pelas respectivas sinapses do neurônio, as operações até aqui descritas constituem um combinador linear, a saída do somador recebe o nome de campo induzido.

c) Função de Ativação: É normalmente do tipo sigmóide. Serve para restringir a amplitude da saída do neurônio. Normalmente no modelo se escolhe valores entre (0 e 1) ou (-1 e 1) para estas funções .

Na Figura 28 encontra- se a representação simplificada do neurônio. O peso w_{0k} recebe também o nome de *Bias* na literatura, é um peso que ajusta a entrada líquida da função de ativação. Mais detalhes em (HAYKIN, 2001) e (WIDROW, 1990).

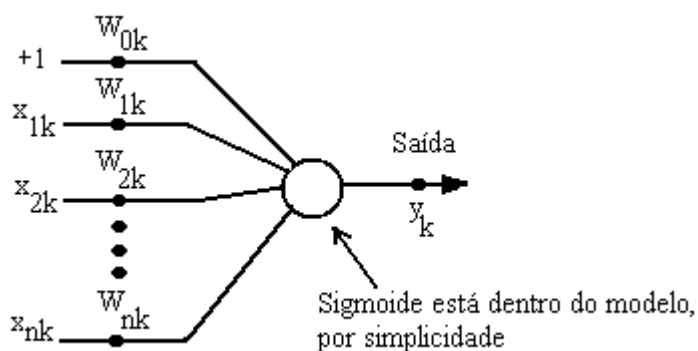


Figura 28: Representação simplificada do modelo do neurônio

O neurônio descrito nos parágrafos anteriores, em alguns casos também é chamado de Perceptron quando for treinado com o algoritmo de Rosenblatt. Quando este modelo for utilizado em camadas, recebe o nome de MLP, porém ele não é treinado, necessariamente, com o algoritmo de Rosenblatt; mas geralmente é treinado com o algoritmo BKP.

As redes MLP estão constituídas por um conjunto de unidades sensoriais (nós de fonte), que formam a camada de entrada; uma ou mais camadas ocultas de nós computacionais e uma camada de saída de nós computacionais. O sinal de entrada se propaga para frente através da rede, camada por camada.

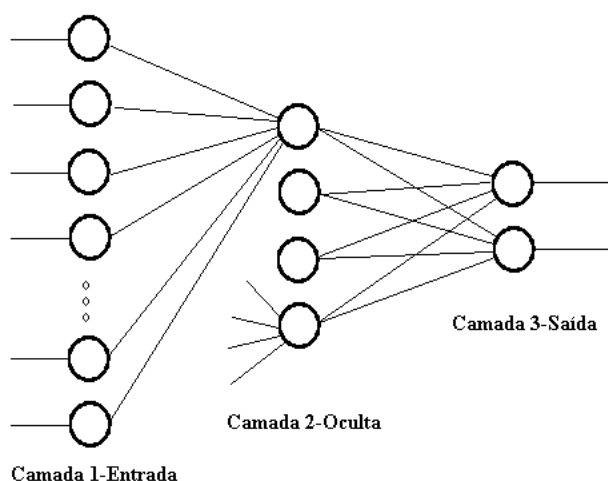


Figura 29: Rede neural do tipo Multi-Layer-Perceptron

Em (FUNAHASHI,1998) é mostrado que uma rede MLP com uma camada oculta e com funções de ativação sigmoidais pode aproximar qualquer função contínua multivariável com qualquer grau de precisão desejado apenas escolhendo o número de neurônios adequado na camada oculta. No entanto, em (FREEMAN, 1992), o autor comenta que existem problemas particulares que são resolvidos 'mais facilmente' adicionando mais camadas intermediárias. Neste caso 'mais facilmente' significa 'mais rapidamente'.

O mapeamento de entrada-saída que é realizado numa rede MLP, é definido como uma serie de mapeamentos intermediários entre as camadas. A cada camada a complexidade não linear aumenta devido aos pesos e às funções de ativação (KOSKO, 1992).

3.4.4 Algoritmo *Backpropagation*

O algoritmo BKP representa um marco nas ANN pois fornece um método computacional eficiente para o treinamento de MLPs. Não é possível afirmar que o BKP forneça sempre uma solução ótima para todos os problemas resolúveis, mas acabou com o pessimismo sobre MLPs que havia sido causado no livro de Minsky e Papert em 1969.

Em algoritmos supervisionados, se tenta adaptar a ANN de forma que suas saídas se aproximem a os alvos desejados para um determinado conjunto de treinamento (WERBOS, 1990).

Na aplicação do BKP, distinguem-se dois passos distintos de computação. O primeiro é o conhecido como propagação e o segundo como retropropagação. A grosso modo, no primeiro passo, os pesos sinápticos permanecem inalterados e se calcula a saída de cada neurônio desde a camada de entradas até a camada de saídas. No segundo passo, se começa avaliando o erro desde a camada de saídas em direção à camada de entradas, avaliando também o chamado gradiente local do erro de cada neurônio. Esse processo recursivo permite que os pesos sinápticos sofram modificações de forma a minimizar o erro. Em uma aplicação pratica do BKP, o aprendizado resulta das muitas apresentações de um determinado conjunto de exemplos de treinamento para o MLP. A apresentação completa do conjunto é chamada de 'época'. O treinamento acontece de época em época até que os pesos se estabilizem e até que o erro médio quadrático sobre o conjunto de treinamento convirga para um valor mínimo. Normalmente os exemplos são apresentados em forma aleatória de uma época para a seguinte para tornar aleatória a busca no espaço de pesos e evitar ciclos limitados na evolução mesmos.

Não se pode demonstrar que o BKP convergiu durante o treinamento e não existem critérios bem definidos para finalizá-lo. Existem critérios razoáveis que podem ser utilizados para encerrar o algoritmo, normalmente o que se utiliza é verificar se a taxa absoluta de variação do erro médio quadrático por época for suficientemente pequena.

Pictoricamente, pode-se representar o algoritmo BKP na Figura 30.

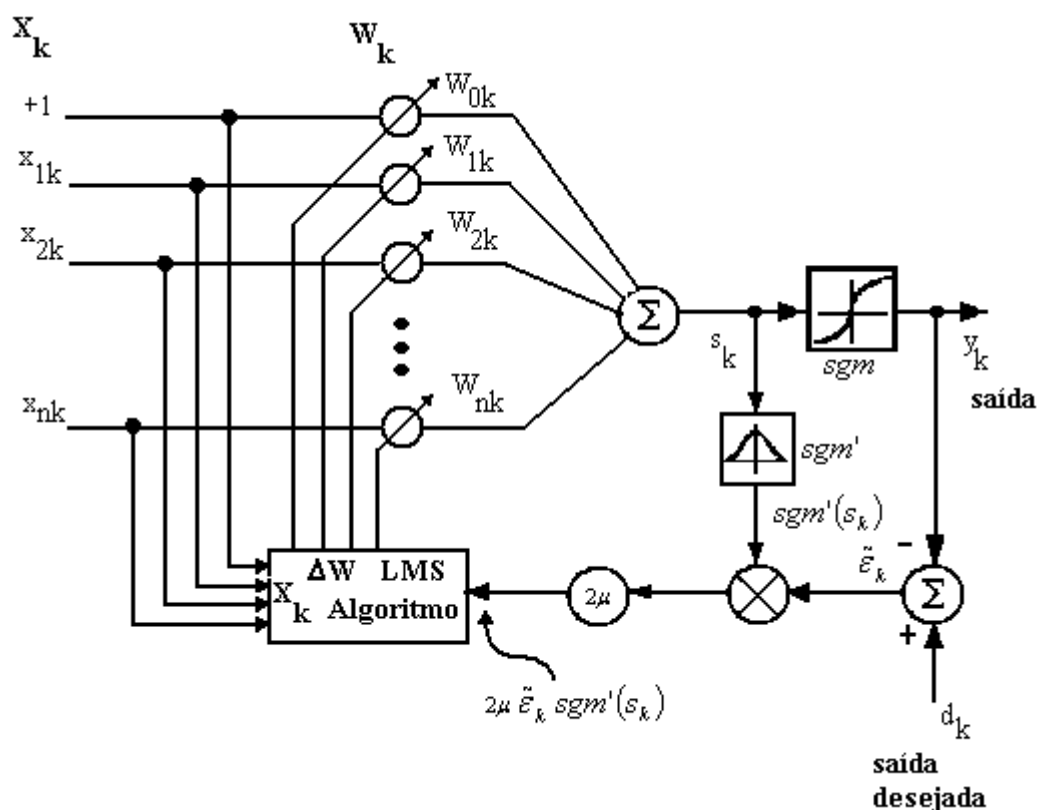


Figura 30: Algoritmo *Backpropagation* exemplificado. (Adaptado de (WIDROW, 1990)).

O sinal de erro na saída de um neurônio j na iteração n (i.e. no n -ésimo exemplo de treinamento) é definido por:

$$e_j(n) = d_j(n) - y_j(n) \quad (46)$$

Onde, d significa valor desejado, e y é o valor real obtido na saída a função de ativação de um neurônio, neste caso é o neurônio genérico j que se encontra na camada de saída do MLP. Definindo o valor instantâneo da energia do erro para o neurônio j como $\frac{1}{2}e_j^2(n)$, o valor instantâneo total de todos os neurônios da camada de saída “C” seria:

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (47)$$

A energia média do erro, sobre todos os N exemplos de treinamento é:

$$\varepsilon_{med}(n) = \frac{1}{N} \sum_{n=1}^N \varepsilon(n) \quad (48)$$

Para um conjunto de treinamento dado, a ε_{med} representa a função custo, ela depende dos parâmetros livres da rede (pesos+*bias*), O processo de aprendizagem ajusta os parâmetros da rede de forma a minimizar ε_{med} .

Seja $v_j(n)$ o campo induzido do neurônio j , com m entradas (sem *bias*):

$$v_j(n) = \sum_{i=0}^m w_{ji}(n) y_i(n) \quad (49)$$

O sinal funcional, $y_j(n)$, que aparece na saída do neurônio j :

$$y_j(n) = \varphi_j(v_j(n)) \quad (50)$$

onde φ_j é a função sigmóide utilizada.

De forma similar ao algoritmo LMS, o algoritmo BKP aplica uma correção aos pesos, proporcional à derivada parcial da função custo. Pela regra da cadeia:

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = \frac{\partial \varepsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (51)$$

Diferenciando ambos lados da Equação (48) em relação a $e_j(n)$ obtém-se:

$$\frac{\partial \varepsilon(n)}{\partial e_j(n)} = e_j(n) \quad (52)$$

Diferenciando ambos lados da Equação (47) em relação a $y_j(n)$ obtém-se:

$$\frac{\partial e(n)}{\partial y_j(n)} = -1 \quad (53)$$

Diferenciando a Equação (51) em relação a $v_j(n)$ obtém-se:

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi'(v_j(n)) \quad (54)$$

Diferenciando a Equação (50) em relação a $w_{ji}(n)$ obtém-se:

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \quad (55)$$

Substituindo a Equação (53), (54), (55) e (56) na Equação (52) obtém-se:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = -e_j(n) \phi'_j(v_j(n)) y_i(n) \quad (56)$$

Pela regra Delta a correção nos pesos deve ser:

$$\Delta w_{ji} = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \quad (57)$$

onde η é a taxa de aprendizagem da rede neural. Correspondentemente o uso da Equação (57) na Equação (58) produz:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad (58)$$

Sendo que o gradiente local é definido por:

$$\delta_j(n) = -\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = -\frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = e_j(n) \phi'_j(v_j(n)) \quad (59)$$

No caso do neurônio j ser um neurônio oculto, não existe uma resposta desejada especificada para ele. Conseqüentemente o erro deve ser determinado recursivamente, em termos dos sinais de erro de todos os neurônios aos quais o neurônio oculto esta conectado.

$$\delta_j(n) = -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \phi'_j(v_j(n)) \quad (60)$$

A energia do erro na camada de saída k é:

$$\mathcal{E}(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad (61)$$

Diferenciando a Equação (62) em relação ao funcional $y_j(n)$, obtemos:

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} \quad (62)$$

Utilizando a regra da cadeia:

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_k(n)} \quad (63)$$

O sinal de erro é definido por:

$$e_k(n) = d_k(n) - y_k(n) = e_k(n) = d_k(n) - \varphi_k(v_k(n)) \quad (64)$$

Assim,

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi'_k(v_k(n)) \quad (65)$$

Da mesma figura, nota-se que o campo local induzido é:

$$v_k(n) = \sum_{j=0}^m w_{kj}(n) y_j(n) \quad (66)$$

onde m é o número total de entradas aplicadas ao neurônio k . Derivando a Equação (66) em relação a funcional $y_j(n)$ produz:

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n) \quad (67)$$

Utilizando as Equações (65), (66), (67) e (68) na Equação (64) obtém-se a derivada parcial desejada:

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = -\sum_k e_k(n) \varphi'_k(v_k(n)) w_{kj}(n) = -\sum_k \delta_k(n) w_{kj}(n) \quad (68)$$

Finalmente, utilizando a Equação (66) e (69) obtém-se a fórmula de retropropagação para o gradiente local, descrito como:

$$\delta_j(n) = -\varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (69)$$

Resumindo as relações que derivamos para o algoritmo de *backpropagation*, a correção do peso sináptico conectando o neurônio i ao neurônio j é:

$$\begin{pmatrix} \text{Correção de peso} \\ \Delta w_{ji}(n) \end{pmatrix} = \begin{pmatrix} \text{Parâmetro da taxa de} \\ \text{aprendizagem} \\ \eta \end{pmatrix} \cdot \begin{pmatrix} \text{Gradiente} \\ \text{local} \\ \delta_j(n) \end{pmatrix} \cdot \begin{pmatrix} \text{Sinal de entrada} \\ \text{do neurônio } j \\ y_j(n) \end{pmatrix} \quad (70)$$

O algoritmo BKP fornece uma aproximação para a trajetória no espaço dos pesos calculada pelo método da descida mais íngreme. Quanto menor for a taxa de aprendizado,

menores serão as variações dos pesos sinápticos da rede e menor as chances da rede ficar instável, (i.e. oscilatória).

Na aprendizagem por BKP esperamos que a rede MLP seja capaz de generalizar. Generalizar significa que o mapeamento de entrada-saída computado pela rede for correto ou aproximadamente correto para dados de testes não utilizados para a criação do treinamento da rede. O termo “generalização” é emprestado da psicologia. Assume-se que os dados de teste são retirados do mesmo ambiente ou da mesma população usada para gerar os dados de treinamento. O processo de aprendizagem (i.e. treinamento da rede) pode ser visto como um problema de “ajuste de curva”. A própria rede pode ser vista como um mapeamento não-linear de entrada-saída. Com este ponto de vista, a generalização é apenas o efeito de uma boa interpolação. Quando uma rede é treinada em excesso (i.e. exemplos são apresentados muitas vezes) a rede pode acabar memorizando os dados de treinamento. Ela pode fazer isto, encontrando uma característica (devido ao ruído, por exemplo) presente nos dados de treinamento que não estão na função subjacente que deve ser modelada. Neste caso temos uma generalização pobre.

4 MATERIAIS E MÉTODOS

4.1 INTRODUÇÃO

No capítulo 4 estão relatados os materiais e os métodos utilizados nos experimentos feitos neste trabalho. Na Seção 4.2 estão apresentadas informações sobre o banco de vozes patológicas utilizadas neste trabalho. O equipamento e o método de aquisição das vozes é brevemente descrito, sendo que informações adicionais encontram-se em (SCHUCK, 1998) e em (PARRAGA, 2002). Na Seção 4.3 é apresentada a formatação dos dados obtidos ao aplicar a WPT e o BBA no banco de vozes. Na Seção 4.4 é apresentado brevemente o teste do algoritmo BKP para validação do mesmo. Na Seção 4.5 é apresentado o sistema utilizado para a classificação de vozes, as virtudes e limitações, assim como também se faz um comparativo com o trabalho de (SCHUCK, 2003).

4.2 BANCO DE VOZES PATOLÓGICAS

A descrição do banco de vozes utilizado se encontra no Apêndice A; as mesmas foram gravadas e diagnosticadas no consultório do Dr. Nédio Steffen, do Hospital da PUCRS. O diagnóstico preciso da existência e tipo de patologia foi obtido pela avaliação dos pacientes através da Videolaringoscopia direta e da Videoestrobolaringoscopia. Digitalizou-se toda a emissão da vogal sustentada /a/ pelo paciente (vogal cujas estruturas do trato vocal se mantém mais estáveis (SCHUCK, 1998)) e foi escolhido o trecho de três segundos que era o mais estável. No total foram gravadas vozes de 63 pessoas, incluindo normais e patológicas.

4.2.1 Equipamento utilizado

Para a gravação das vozes foi utilizado um microfone marca LeSon, modelo LM-58, com resposta em frequência linear de 50 Hz a 17KHz. O sinal do microfone foi amplificado graças a um circuito baseado em 2 amplificadores operacionais em cascata, na configuração não-inversor, com alimentação por bateria 9V. O sinal foi então levado a um gravador cassete

Technics model RSB105, que possui um sistema de redução de ruído da fita do tipo Dolby-NR. As fitas cassetes utilizadas foram do tipo II (óxido de cromo) da Sony, modelo UX60.

Para a reprodução, conversão analógica-digital e pré-processamento dos sinais de voz, foi utilizado o seguinte equipamento do Laboratório de Processamento Digital de Sinais e Imagens (LaPSI-DELET-UFRGS): duplo deck Tascam 302, mesa de mixagem Mackie 1402 VLZ, um conversor analógico-digital (A/D) de 16 bits, uma placa de processamento de sinais usando o processador DSP TMS320C25 da Texas Instruments, um microcomputador Pentium 133MHz. A placa DSP e o conversor A/D D/A foram desenvolvidos no LaPSI.

As principais características da placa DSP são:

- a) Processador TMS0C25, com *clock* de 40 MHz e ciclo de instruções de 100 ns.
- b) Memória de 32 kBytes, permitindo uso conjunto de RAM e EPROM.
- c) Barramento para expansão de memória.
- d) Duas interfaces de entrada e saída, uma com 8 e outra com 16 bits.
- e) Uma interface de comunicação serial, do tipo RS232.
- f) Interface de comunicação com o micro computador, no padrão PC-AT, com acesso em 16 bits.

As principais características do conversor A/D e D/A, também desenvolvido no LaPSI, são:

- a) Conversor A/D de 14 bits, e capacidade de até 80.000 amostras por segundo (frequência de amostragem, f_s), baseado no "chip" AD7871.
- b) Duas saídas D/A de 8 bits, baseadas no "chip" PM-7226.
- c) Comunicação serial com a placa de processamento de sinais DSP.
- d) Três entradas de sinal: 2 normais (RCA) e 1 do tipo diferencial balanceada.

O sinal a ser analisado foi limitado em banda através de um filtro passa-baixa *anti-aliasing* de baixa ordem, descrito no trabalho de (SCHUCK,1998), este filtro garante uma resposta plana na banda do sinal. O processo de aquisição foi feito através da sobre-amostragem do sinal em 3 vezes a frequência de amostragem final (f_s), filtragem digital com um filtro do tipo passa-baixas, sub-amostragem (decimação pelo fator 3) do sinal filtrado (de cada 3 amostras, 2 são desprezadas) de forma a ter um sinal amostrado à f_s . Para um sinal cuja banda é 12kHz, a frequência de amostragem mínima seria de 24 kHz, seguindo o critério

de Nyquist. No sistema proposto, o sinal é filtrado por um filtro analógico passa-baixas de 1ª ordem com frequência de corte em 25Khz e amostrado em 75 kHz, filtrado digitalmente e então sub-amostrado para uma $f_s=25\text{kHz}$, obtendo-se uma banda de sinal efetiva de 12,25 kHz. O filtro passa-baixas digital utilizado foi um filtro FIR (*Finite Impulse Response*) do tipo janela *Hamming* de 63 pontos, com banda de passagem de $0,16.f_s$ (12kHz) (SCHUCK, 1998).

4.2.2 Organização do Banco de Vozes

Observando o banco de vozes utilizado, nota-se que o número de amostras por patologias seria estatisticamente insuficiente para o teste adequado de classificadores, entretanto no artigo (ROSA, 1998) os autores propõem um agrupamento de patologias baseado na acústica e nas similaridades patológicas da voz ao invés de patologias individuais.

Os grupos e as patologias pertencentes a cada grupo encontram-se na Tabela 4, um novo grupo (o 7º grupo) foi incluído; é aquele que representa todas as patologias não englobadas de forma direta pelos grupos anteriores.

Tabela 4: Grupos de patologias utilizado neste trabalho

Grupo	Patologias
Laringite crônica	Laringite
Degenerativa	Carcinoma, granuloma, leucoplasia,
Mobilidade Incorreta	Paralisia unilateral, disfonia unilateral e disfonia hiperkinética
Alterações orgânicas	Edemas de Reinke e edemas generativos
Crescimentos orgânicos	Pólipos, nódulos e cistos
Normal	Normais
Nenhum dos Anteriores	Indeterminada

A mesma idéia dos autores (ROSA, 1998) foi aplicada ao banco de vozes que está descrito no Apêndice A. Os pacientes foram agrupados conforme a Tabela 5. Na primeira coluna está o número do grupo, na coluna do meio os pacientes daquele grupo e na terceira está número total de amostras para cada grupo. Existem pacientes que pertencem a mais de um grupo de patologias.

Tabela 5: Grupos de patologias para o Banco de Vozes descrito no Apêndice A.

Grupo	Pacientes	Total de Amostras
Laringite crônica	4, 9, 15, 30, 32, 51, 53, 56, 62, 63, 66, 67, 73	13
Degenerativa	15,22,41,58,59,62,63	7
Mobilidade Incorreta	10,13,26,27,28, 29,30,31,43,46	10
Alterações orgânicas	6,23,32,35,36	5
Crescimentos orgânicos	1,21,24,26,50,69,71	7
Normal	2,18,19,44,45,47,48, 49,55,61,64,68,70	13
Nenhum dos Anteriores	5,8,11,14,16,17,25,37, 42,54,57,60,72,74	14

4.3 APLICAÇÃO DA WPT NAS VOZES PATOLÓGICAS

A ferramenta utilizada para executar todos os cálculos envolvidos, inclusive a Transformada Wavelet Packet foi o MatlabTM v5.3 (Matlab, é uma marca registrada da MathWorks Inc.) o motivo principal do uso deste software foi a facilidade que a linguagem de programação apresenta ao uso de cálculos matriciais, além de ser uma ferramenta altamente difundida no meio científico. Ao aplicar a WPT e o BBA (com o Wavelet Toolbox do Matlab v.5.3) a um sinal de voz (paciente) do banco de vozes, obtém-se uma estrutura denominada árvore binária. Esta estrutura foi re-organizada de forma que cada elemento (nó da árvore) seja um elemento de um vetor que caracterizara ao paciente em questão. Cada vetor dos pacientes foi agrupado de forma a obter uma matriz.

Foi mostrado em (PARRAGA, 2002) que a WPT junto com o BBA, são ferramentas capazes de separar patológicos de normais utilizando um classificador linear. Entre as bases investigadas, a base escolhida foi a Symlet 5 pois com ela, as árvores da WPT geradas, se mostraram invariantes ao sexo e à frequência fundamental da voz e do paciente. Foi escolhida como função custo a entropia de Shannon para gerar a BBT (*Best Basis Tree*) pois com ela a topologia da árvore gerada, apresentou maiores características de separabilidade linear entre as classes patológico e normal. Isto justifica a escolha desta base, a Symlet 5 com a entropia de Shannon como função custo, neste trabalho.

Nota-se também que em (PARRAGA, 2002) o discriminante entre classes foi a existência ou não de determinados nós na BBT (árvore podada), já em (SCHUCK, 2003) o discriminante entre classes é a similaridade vetorial entre o padrão de treinamento (aquilo que a rede neural aprendeu a diferenciar) e o padrão apresentado (o vetor que representa a BBT do paciente.) Ou seja, em (SCHUCK, 2003) se estenderam os fatores diferenciadores das classes. Além da existência ou não de nós na árvore podada também conta o valor da Entropia de cada nó.

Num sistema como este a primeira etapa é a aquisição e armazenamento do sinal, porém as vozes utilizadas para treinar as redes neurais foram extraídas do banco de vozes patológicas utilizado em (SCHUCK, 1998) e em (PARRAGA, 2002).

Efetuuou-se o cálculo da WPT e a BBA, para obter a BBT, os valores de saída deste estágio estão colocados na forma matricial como será mostrado mais adiante, na Equação (72).

4.4 TESTE DO ALGORITMO BACKPROPAGATION PARA MLP

O algoritmo BKP também foi implementado no Matlab v5.3, com base nas fórmulas recursivas da Seção 3, e das heurísticas de (HAYKIN, 2001) para acelerar convergência e determinar momento de parada do treinamento.

Para testar o algoritmo, duas situações foram propostas:

a) Primeiro determinou-se como parâmetros da rede: dois neurônios de entrada e dois na camada oculta, um neurônio na saída. A tarefa foi “aprender” a tabela verdade da função booleana XOR (OR exclusivo). Foi escolhida esta função booleana por ela ser simples e por ser não linearmente separável. Verificou-se que a rede em teste após poucas épocas convergiu sem problemas para um erro de classificação de 0%.

b) Após verificar que o algoritmo funcionou para um problema de dimensão baixa, o segundo passo foi repetir as experiências de (SCHUCK, 2003), diferenciando vozes patológicas de normais, utilizando o mesmo Banco de Vozes e as mesmas configurações de rede. Os resultados obtidos foram os mesmos, portanto validando o algoritmo.

4.5 SISTEMA DE REDES NEURAIS INDEPENDENTES POR GRUPO DE DOENÇAS

O sistema proposto pode ser sumarizado na Figura 31:

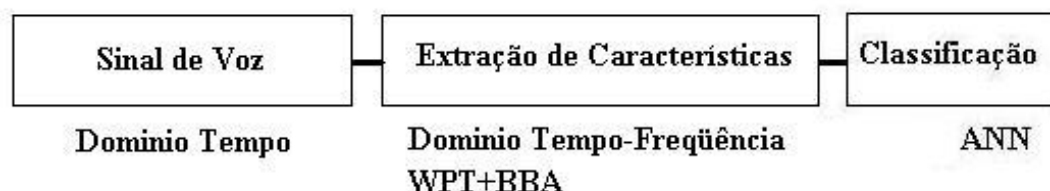


Figura 31: Diagrama de blocos do sistema proposto

A análise realizada é um apoio ao diagnóstico de doenças da laringe. Consiste na extração das características da voz através da transformação tempo-freqüência (aplicação da WPT juntamente com o algoritmo BBA), para gerar a BBT. A classificação é realizada através de um sistema composto por 6 redes neurais independentes, especialmente treinadas para identificar cada grupo de doenças da Tabela 4.

Cada rede neural, neste contexto é chamada de "especialista" e é do tipo MLP, treinada com o algoritmo BKP. O número de camadas é 3; uma de entrada com 63 neurônios², cada neurônio recebe um sinal de entrada proveniente BBT, especificamente o valor de entropia de cada nó é passado como uma característica discriminante do espaço de entradas (voz do indivíduo). A camada intermediária (oculta) possui um número de neurônios que varia segundo a necessidade de cada especialista. E, finalmente, a camada de saída possui 1 neurônio; ou seja, a saída pode ser interpretada como o erro/acerto da rede. Neste contexto a rede neural descrita no trabalho (SCHUCK, 2003), constitui-se em um "especialista" em diferenciar patológicos de normais.

O uso de seis ANNs com saídas independentes ao invés de uma ANN com seis saídas é justificado porque desta forma é possível avaliar o grau de não linearidade presente no espaço de entradas. Quanto maior o número de neurônios ocultos de um especialista, mais não-linear, e portanto, mais complexa se torna a avaliação que a ANN tem que realizar sobre

²Soma de todos os nós de cada nível da árvore da WPT: $2^0 + 2^1 + 2^2 + 2^3 + 2^4 + 2^5 = 63$

os parâmetros de entrada para a tomada de decisão. Isto pode ser útil, por exemplo, em trabalhos futuros onde se busque entender porque é mais “fácil” reconhecer certo tipo de patologias. Outra utilidade de se ter seis ANN independentes ao invés de uma só, pode ser exemplificado da seguinte maneira:

Um sistema integrado para utilização em consultório, seria mais eficiente se permitisse que novos elementos fossem incorporados na base de conhecimento das ANN que o compõem. Por isso, seis redes neurais independentes podem ser mais atraentes, já que o tempo de treinamento seria apenas sobre uma rede neural menor, ao invés de ser sobre uma rede neural maior com seis saídas. Por exemplo, se um paciente diagnosticado com laringite possuir alguma característica que o avaliador considera relevante de ser incorporada à base de conhecimento das redes neurais do sistema, será necessário re-treinar apenas a ANN especialista em pacientes do grupo de laringite crônica.

Dentre os diferentes tipos de ANN, os aproximadores universais mais utilizados são o MLP juntamente com as Redes de Base Radial (RBF). Ambos tipos de ANN possuem características atraentes para a aplicação neste trabalho. Sendo testado apenas o MLP e colocado como sugestão para trabalhos futuros o uso de RBF. Fazendo um comparativo breve sobre estes dois tipos de rede, com base na literatura pode-se concluir que, as diferenças apontam em relação ao erro final atingido e convergência. Ambas ANN são treinadas com algoritmos supervisionados, geralmente pelo algoritmo *Backpropagation*. Também diferem quanto à arquitetura:

Enquanto que os neurônios ocultos das redes MLPs avaliam a soma ponderada das entradas transformadas por funções de ativação monotônicas, comumente não lineares; nas redes RBF, a ativação de um neurônio oculto é determinada por uma função não-linear da distância entre o vetor de entrada e um vetor de referência. As funções de ativação são localizadas e apresentam base radial sobre seu domínio.

Uma MLP forma uma representação distribuída no espaço de valores de ativação para as unidades escondidas, pois para um vetor de entrada, muitas unidades escondidas contribuirão para a determinação do valor de saída. (MLPs tendem a resultar aproximações globais) (BISHOP, 1997). A interconexão entre as unidades escondidas leva a resultados no processo de treinamento da rede que são muito não-lineares, resultando em problemas de mínimos locais ou em regiões quase planas na função de erro, fatores estes que podem levar a uma convergência muito lenta. Os parâmetros de uma MLP são usualmente determinados ao

mesmo tempo, estratégia global de treinamento, envolvendo treinamento supervisionado. Este tipo de treinamento apresenta um alto custo computacional, pela necessidade de retropropagação do erro, o que faz as MLPs terem um aprendizado muito lento. Porém o desempenho de generalização é bom (BISHOP, 1997).

RBFs possuem funções de base localizadas que formam uma representação no espaço de unidades escondidas que é local em relação ao espaço de entrada porque, para um vetor de entrada, tipicamente apenas algumas unidades escondidas apresentarão ativações significativas. Por isso, as RBFs tendem a produzir aproximações locais (BISHOP, 1997);

As RBF são treinadas em dois estágios, com as funções de base radial sendo determinadas primeiramente por técnicas não-supervisionadas, usando para tal os dados de entrada e a segunda camada (pesos) sendo depois determinada por métodos lineares supervisionados, de rápida convergência.

No contexto de aproximação de funções, sob idênticas condições do ambiente no qual estão inseridas, de uma forma geral pode-se afirmar que (BISHOP, 1997), (HAYKIN, 2000):

- a) O erro final atingido por uma RBF é menor que o de uma MLP;
- b) A convergência de uma RBF pode chegar a uma ordem de grandeza mais rápida do que a convergência de uma MLP;
- c) A capacidade de generalização da MLP é, em geral, superior a capacidade de generalização da RBF.

A seguir serão apresentados vários diagramas de blocos exemplificando o sistema proposto.

Na primeira etapa, utilizou-se os sinais do banco de vozes utilizado em (SCHUCK, 1998) e em (PARRAGA, 2002). Neste banco de vozes, os sinais se encontram como mostrado no primeiro bloco da Figura 32.

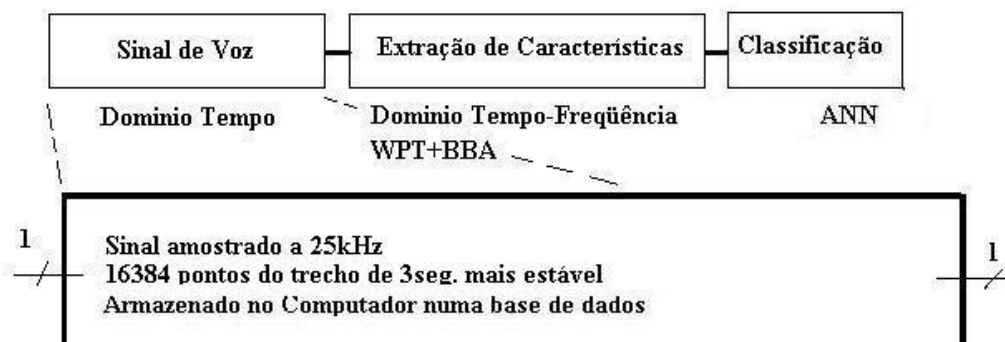


Figura 32: Bloco de aquisição do sinal de voz.

A seguir efetua-se o cálculo da WPT e o BBA, para obter o BBT (ver a Figura 33).

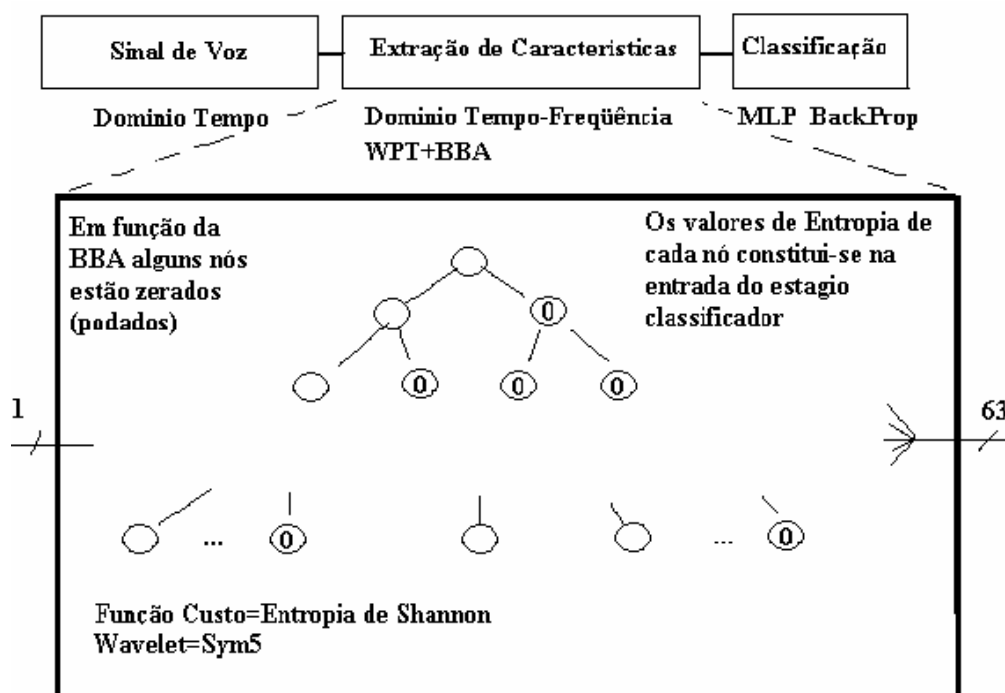


Figura 33: Bloco de extração de características.

Os valores de saída deste estágio estão colocados na forma matricial, cada linha da matriz corresponde a uma característica do paciente (existem 63 características extraídas do

senal de voz, pois existem 63 nós na árvore gerada pela WTP e o BBA), e cada coluna corresponde a um paciente em especial; como exemplificado na Equação (71):

$$M = \begin{bmatrix} Entropia01Paciente01 & Entropia01Paciente02 \dots Entropia01Paciente63 \\ Entropia02Paciente01 & Entropia02Paciente02 \dots Entropia02Paciente63 \\ \dots & \dots \dots \dots \\ Entropia63Paciente01 & Entropia63Paciente02 \dots Entropia63Paciente63 \end{bmatrix} \quad (71)$$

A matriz de treinamento (derivada a partir de M), apresenta peculiaridades que podem dificultar a aprendizagem das ANNs que vêm no estagio de treinamento. Existe muita variação da ordem de grandeza dos números entre as diferentes linhas (características), assim, o procedimento adotado foi "normalizar" esta matriz antes de entrar no estagio seguinte. Segundo (HAYKIN, 2001) a normalização dos dados apresentados a um MLP facilita a tarefa de classificação. Alguns dos procedimentos de normalização testados foram:

- a) Dividir cada linha da matriz pela norma euclidiana da linha, desta forma os dados de características terão sempre norma unitária.
- b) Normalizar os dados de cada paciente (colunas de M), dividindo cada coluna pelo máximo absoluto encontrado na coluna.
- c) Substituir cada elemento da matriz M por 1's e -1's, segundo existência ou não de nós na BBT, de forma que quando a entropia em um determinado nó for zero (nó podado) designa-se o valor -1 , senão designa-se 1. Depois de treinar as redes neurais os resultados foram piores que em (SCHUCK, 2003) verificando um aspecto importante de classificação: não apenas a morfologia da árvore podada é necessária para diferenciar as diferentes patologias, mas se torna imprescindível avaliar o valor da entropia em cada nó da BBT.
- d) Para cada característica do paciente, converter o máximo valor em +1 e o menor valor em -1 conforme regra de três direta. Este tipo de normalização muda a topologia da árvore BBT, o que não é interessante. Por exemplo, se o menor valor de entropia for zero (i.e. nó podado), este vai assumir o valor -1 .
- e) Cada variável de entrada é subtraída a sua média e dividida pelo seu desvio padrão, este método foi sugerido em (HAYKIN, 2001), de forma a ter dados com média zero e variância unitária. Os resultados com este método de normalização foram bons, só que para o cálculo de média e variância de cada variável seria necessário um

número suficientemente grande de amostras (vozes), para inferir quanto vale a média e a variância da população.

f) Escalonar a matriz de treinamento dividindo cada elemento pelo maior elemento em módulo, desta forma o maior valor (em módulo) da matriz normalizada será 1. Este método não apresentou bons resultados pois ao dividir todas as linhas (características) pelo mesmo valor, estar-se-ia fazendo uma mistura dimensional. Seja, por exemplo:

$$M = \begin{bmatrix} 20^\circ & 75^\circ & 123^\circ \\ 2Km & 100Km & 25Km \\ 1s & 2s & 4s \end{bmatrix}, \text{ cada linha representa uma característica}$$

diferente, o maior valor da matriz é 123 graus, dividir cada valor de M por 123 causa um problema de coerência com as unidades (grau, quilômetro, segundo). Embora neste trabalho sempre que se fala em entropias se fala da mesma unidade para cada nó, mas as ANN interpretam cada entrada como características diferentes.

Finalmente a normalização adotada foi dividir cada característica (linha de M), pelo máximo valor da linha de M correspondente a essa característica, com isso evitamos o problema de unidades descrito acima. Com este escalonamento obtiveram-se os melhores resultados para os grupos de patologias até o momento testadas. Cada vetor de características deve ser dividido pelo seu máximo valor absoluto. Em aplicações práticas também é difícil conhecer os valores máximos e mínimos que as entropias de cada nó podem assumir, apenas um conhecimento melhor das BBT permitirá elaborar uma forma melhor de normalizar as entradas das ANN, isso será possível aumentando a banco de vozes disponível.

Na Figura 34 é apresentado o último bloco, onde se efetua a classificação:

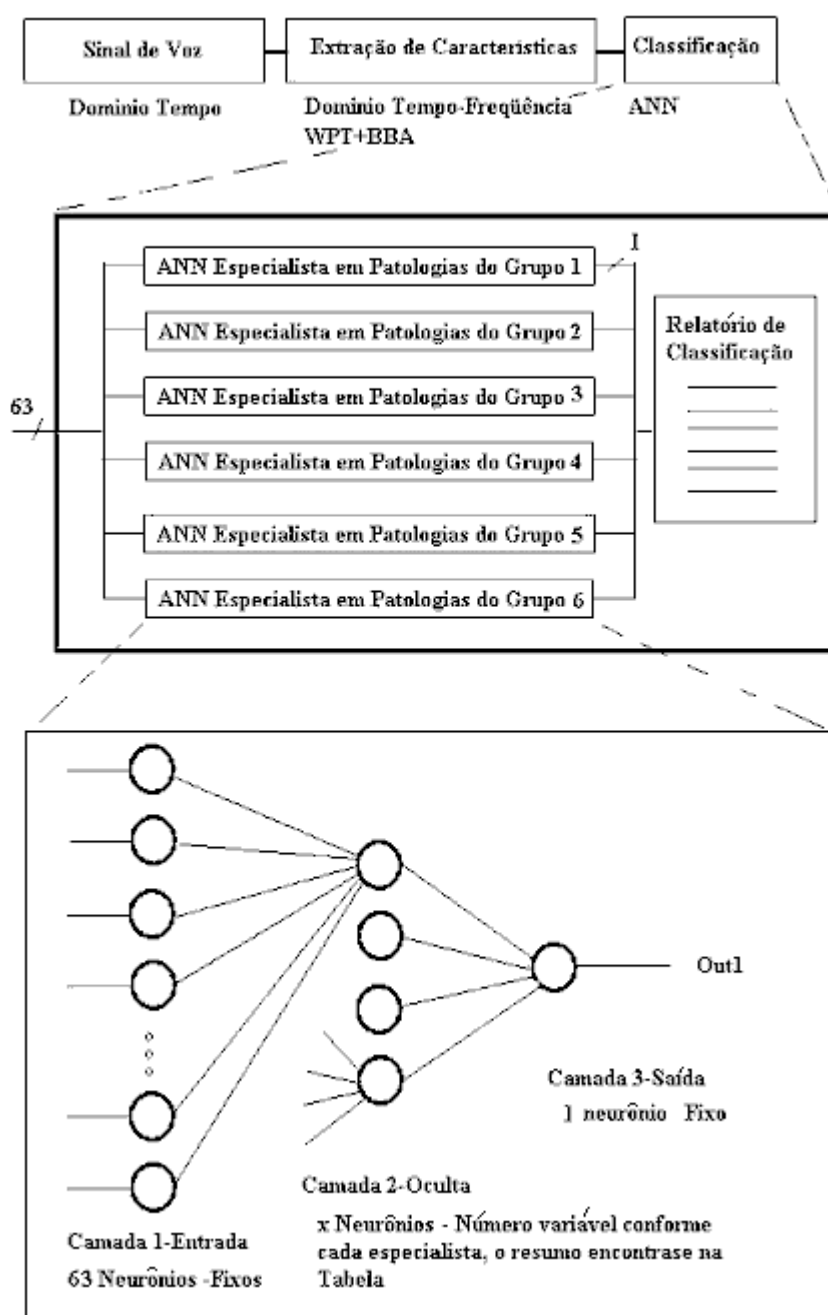


Figura 34: Bloco de classificação, rede neural “especialista”.

Neste ponto, antes de proceder à classificação, é interessante fazer um estudo sobre a dimensionalidade do problema. Muitos autores utilizam algum método para reduzir a dimensionalidade dos dados passados a uma rede neural, com o objetivo de reter informação

importante para a discriminação entre classes e descartar a informação desnecessária. Um classificador com poucas entradas, tem menos parâmetros adaptativos a serem ajustados e sendo assim possui um poder maior de generalização (HAYKIN, 2001).

Depois de efetuar a WPT e a BBA obteve-se uma árvore “podada”, com 63 nós (entradas do MLP), ou seja, o MLP deve ser capaz de classificar um problema 63-dimensional. Muitos autores, que utilizam a WPT e a BBA juntamente com o MLP, tentam reduzir ainda mais a dimensionalidade do problema, pois como foi dito antes, quanto mais variáveis ajustáveis tenha o MLP, mais difícil será sua tarefa de generalização. Existem vários métodos estatísticos e de redes neurais para reduzir dimensionalidade, os mais conhecidos são os Mapas Auto Organizáveis de Kohonen (SOM), o algoritmo LDA, e K-means, discriminantes de Fisher, Análise de Componentes Principais (PCA) entre outros.

Num primeiro momento não foi reduzida a dimensão do problema por vários motivos:

- a) Antes de proceder a reduzir a dimensão, foi necessário verificar o comportamento do MLP sem esta redução, bons resultados sem reduzir a dimensionalidade, evitaria desperdício de tempo de simulação.
- b) Em comparação com os trabalhos pesquisados, um espaço de dimensão 63 ainda pode ser considerado pequeno (comparar por exemplo o trabalho de (ENGLEHART, 1999) onde se tem um problema 2032-dimensional), ou então em (HONG, 2003) onde as entradas do MLP são os coeficientes de detalhe e aproximação da própria DWT.
- c) Computar a PCA exige um esforço computacional grande. Sem contar o fato que o banco de vozes patológicas é pequeno e o *kernel* determinado pelo PCA, talvez não seja o melhor para descrever a população.

A “inicialização” dos pesos de cada rede neural, é randômica, mas obedece a diretiva de LeCun (HAYKIN, 2001), ou seja cada peso é “inicializado” independentemente conforme o número de neurônios assumindo valores pequenos para evitar a saturação de neurônios. Esta “inicialização” randômica determina as condições iniciais de como o treinamento irá se desenvolver, podendo ter convergência rápida ou mais lenta. Também define o “caminho” que se segue na superfície de erro.

A função de ativação tangente hiperbólica (*tanh*) tem um comportamento anti-simétrico, conforme a Figura 35, o que facilita o treinamento, em termos de número de épocas

necessário. Considerando como valores alvo (saída desejada), os valores 0.9 ou -0.9 ; tendo um ε de ± 0.1 dificulta-se a saturação de qualquer peso.

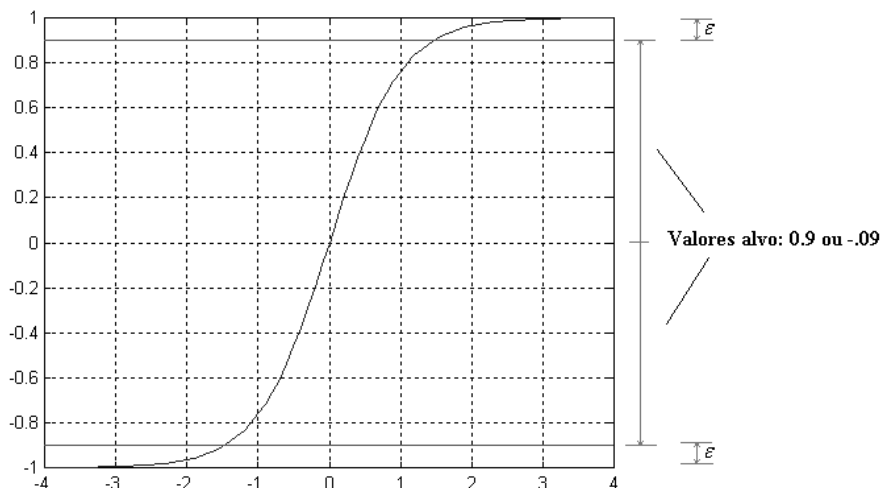


Figura 35: Função de ativação tangente hiperbólica, com valores alvo para evitar saturação.

No treinamento com o BKP foi utilizada a atualização por lote, ou seja os pesos são modificados somente no final de uma época.

Treinaram-se cada una das seis ANNs, buscando a configuração com menor erro de generalização e menor erro de classificação, tendo como parâmetro da rede o número de neurônios na camada oculta e a taxa de aprendizagem inicial, a qual era modificada segundo o decorrer do treinamento, decaindo conforme o erro de treinamento em função das épocas.

Utilizaram-se vários critérios de parada antecipada do algoritmo de treinamento:

- a) Terminar o treinamento se for verificado que o erro de generalização permanece constante por pelo menos 5000 épocas.
- b) Terminar o treinamento se for verificado que o erro de generalização aumenta consistentemente por pelo menos 5000 épocas.
- c) Terminar o treinamento se for verificado que o erro de classificação esta aumentando consistentemente mesmo que o erro de generalização continue decaindo.

A ferramenta padrão da estatística conhecida como Validação Cruzada Múltipla (MCV) foi utilizada para a seleção e teste do modelo das ANNs. Dividiu-se o banco de vozes em dois conjuntos distintos:

- a) Conjunto de Treinamento, composto por:
 - Subconjunto de Estimação (ou Subconjunto de Treinamento), usado para selecionar o modelo. Este conjunto contém os padrões usados para modificar os pesos.
 - Subconjunto de Validação, usado para validar o modelo. Este conjunto contém os padrões usados para verificar o problema de super treinamento (*overfitting*). Serve como indicador da capacidade de generalização da rede, determina o instante de parada do treinamento.
- b) Conjunto de Teste, é o conjunto de padrões para testar o desempenho do modelo final.

OBS: Como o número de indivíduos é muito pequeno, optou-se apenas por validar o modelo e deixar o teste para quando se possuíam mais amostras de indivíduos. Porém depois de validar o modelo, foram apresentados pacientes do Grupo 7 apenas para avaliar a resposta que as ANN fornecem perante este tipo de pacientes, comentários adicionais são necessários sobre a interpretação dos resultados.

Conforme a MCV o conjunto disponível de N exemplos rotulados deve ser dividido em k subconjuntos. Ao todo tem-se 64 pacientes, com 14 pertencentes do Grupo 7, então $N=50$ e o k escolhido foi 5. Desta forma obtêm-se os subconjuntos (M1,M2,M3,M4,M5) como mostra a Tabela 6.

Tabela 6: Subconjuntos criados a partir do Banco de Vozes.

SubGrupo	Pacientes (PACs)
M1	01, 02, 04, 06, 09, 10, 13, 15, 18, 19
M2	22, 23, 24, 26, 27, 29, 30, 64, 53, 66
M3	36, 41, 43, 44, 45, 46, 47, 48, 21, 73
M4	49, 50, 55, 56, 58, 59, 61, 62, 28, 35
M5	62, 63, 51, 67, 68, 69, 70, 71, 31, 32

Cada subconjunto 'M' possui pacientes com cada grupo de patologia da Tabela 1. A soma de todos corresponde ao Conjunto de Treinamento. Com os 5 subconjuntos 'M' é

possível formar 5 configurações utilizando 4 'M's para o treinamento, deixando de fora sempre um para a validação, com mostra a Tabela 7.

Tabela 7: Configurações de Estimação (Treinamento) e Validação.

Configuração	Subconjunto de Treinamento	Subconjunto de Validação
A	M2+M3+M4+M5	M1
B	M1+M3+M4+M5	M2
C	M1+M2+M4+M5	M3
D	M1+M2+M3+M5	M4
E	M1+M2+M3+M4	M5

Uma das configurações é mostrada na Figura 36, nesta figura também são mostrados os diferentes conjuntos e subconjuntos utilizados no processo de MCV:

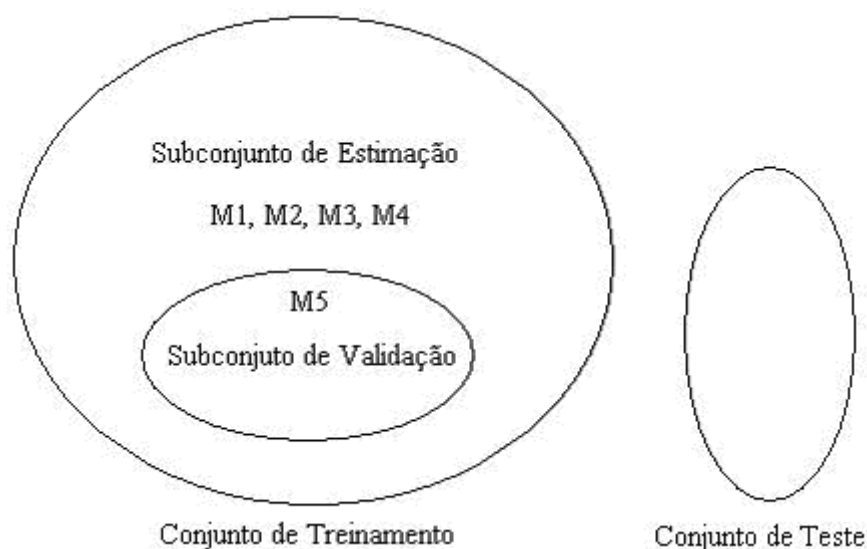


Figura 36 Conjuntos de Treinamento e Teste, para a configuração E.

Escolheu-se um valor inicial de taxa de aprendizado de $\eta = 0.0002$. O modelo da ANN foi treinado com cada configuração da Tabela 7, o erro de validação foi calculado sobre a média dos erros de validação obtidos com cada configuração.

Limitou-se o número de neurônios a um valor inferior à metade do número de entradas do sistema, atendendo assim as recomendações de (FREEMAN, 1992). ANNs com diferentes números de neurônios na camada oculta foram testadas (5, 10, 15, 20, 25, 30), escolhendo aquela topologia que apresenta menor erro de validação médio.

Escolhido o número de neurônios 'ótimo' treinou-se novamente a ANN, desta vez com todos os N exemplos durante i épocas, onde i é a média de todas as épocas necessárias

para o treinamento das k configurações. Os resultados das configurações finais são apresentados na Tabela 12.

Após determinar o número de neurônios ótimo com o método MCV. Para avaliar o poder de generalização das configurações finais encontradas, aplicou-se se novamente a MCV levada ao caso extremo em que o conjunto de validação é apenas um indivíduo. Este método é conhecido como *leave-one-out* (LOO).

Conforme o LOO o conjunto disponível de N exemplos rotulados deve ser dividido em k subconjuntos, onde $k=1$. Ao todo se têm 64 pacientes, com 14 pertencentes do Grupo 7, então $N=50$. Desta forma obtêm-se 50 subconjuntos de validação, caracterizando o LOO como um método computacionalmente pesado.

Ao aplicar o LOO, para cada grupo de patologias é possível construir uma tabela de 50 linhas como exemplificada na Tabela 8, onde cada paciente é o conjunto de validação do treinamento realizado. A média dos erros de validação é a capacidade de generalização da rede, ou seja a capacidade que tem o “especialista” de computar uma saída correta ou aproximadamente correta para dados de testes não utilizados para a criação do treinamento da rede.

Tabela 8 Exemplificação dos erros de validação LOO, Grupo 2.

Paciente (PAC)	Erro de validação(%)
1	0.0099
2	10.9058
4	16.1882
...	...
...	...
70	41.3819
71	0.2967
73	4.1665
Média →	35.9439

Os resultados completos se encontram no Apêndice E, e o resumo dos erros de generalização se encontra na Tabela 14.

A média dos erros de validação para o exemplo da Tabela 8, foi de 35,9439% ou seja em média os erros médios absolutos das saídas da rede neural especialista em patologias do Grupo 2, se afastavam em 35,9439% do valor esperado. Mesmo com a capacidade de generalização baixa, o erro de classificação também é baixo (4,68%) como mostra a Tabela 9.

4.5.1 Cálculo das saídas do sistema

a) Erro de Classificação:

$$E_{classif} = 100 \left(\frac{\text{Número de Falsos Negativos} + \text{Número de Falsos Positivos}}{\text{Número de Pacientes}} \right) \quad (72)$$

b) Taxa de Sucesso

$$T_{sucesso} = 100 - \text{Erro de Classificação} \quad (73)$$

c) Erro para Falsos Negativos:

$$E_{fn} = 100 \left(\frac{\text{Número de Falsos Negativos}}{\text{Número de Pacientes}} \right) \quad (74)$$

d) Erro para Falsos Positivos:

$$E_{fp} = 100 \left(\frac{\text{Número de Falsos Positivos}}{\text{Número de Pacientes}} \right) \quad (75)$$

e) Classificação do Sistema:

Com o sistema composto por n ANN independentes (onde $n=6$), cada uma com um neurônio de saída, a classificação pode ser avaliada observando as saídas y_i , onde i varia de 1 a 6. Estas saídas são valores que variam de -1 a 1 (conforme a função de ativação *tanh*). O limiar de aceitação para identificar se o paciente pertence ao grupo do especialista i é:

$$\begin{cases} y_i \geq 0, PAC \in Grupo_i \\ y_i < 0, PAC \notin Grupo_i \end{cases} \quad (76)$$

f) Erro de Treinamento e Erro de Validação (ou Erro de Generalização):

Durante o treinamento é avaliado de época em época o erro MAPE sobre os padrões de treinamento e também sobre os padrões de validação. O MAPE indica o valor médio do erro absoluto percentual das previsões sobre todo o conjunto de N amostras.

$$MAPE(\%) = \frac{\sum_{k=1}^N \left| \frac{a_k - y_k}{a_k} \right|}{N} 100 \quad (77)$$

onde, a_k é saída desejada para a previsão índice “ k ” e,
 y_k é saída prevista para a previsão índice “ k ”.

5 RESULTADOS

5.1 INTRODUÇÃO

No capítulo 5 são apresentados os resultados finais obtidos neste trabalho. Na seção 5.2 são mostrados os resultados da classificação obtida utilizando o treinamento supervisionado do algoritmo BKP nas redes MLP dos grupos 1 ao 6.

5.2 RESULTADOS DA CLASSIFICAÇÃO UTILIZANDO O ALGORITMO BKP NOS MLP

Depois de executar varias configurações, os resultados são apresentados nas tabelas a seguir (Tabela 9 até a Tabela 12):

Tabela 9: Erros de Classificação e Taxas de Sucesso.

Especialista	Grupo	Erro de Classificação	Taxa de Sucesso
1	Laringite crônica	12,5%	87,5%
2	Degenerativas	4,68%	95,31%
3	Mobilidade Incorreta	12,5%	87,5%
4	Alterações orgânicas	0%	100%
5	Crescimentos orgânicos	3,12%	96,87%
6*	Normais	10,93%	89,06%

Tabela 10: Pacientes (PACs) classificados como Falsos Negativos (FN) e Falsos Positivos (FP).

Especialista	Grupo	FN (PACs)	FP (PACs)
1	Laringite crônica	4, 9, 34, 62, 63, 66, 73	68
2	Degenerativas	58, 59, 63	-
3	Mobilidade Incorreta	10, 26, 28, 29, 30, 31, 43	34
4	Alterações orgânicas	-	-
5	Crescimentos orgânicos	50, 69	34
6*	Normais	48, 49, 55, 61, 64	63, 69

Tabela 11: Erro devido a Falsos Negativos (FN) e Falsos Positivos (FP).

Especialista	Grupo	Erro para FN	Erro para FP
1	Laringite crônica	10,93%	1,56%
2	Degenerativas	4,68%	0%
3	Mobilidade Incorreta	10,93%	1,56%
4	Alterações orgânicas	0%	0%
5	Crescimentos orgânicos	3,12%	1,56%
6*	Normais	7,81%	3,12%

Tabela 12: Configurações Finais.

Especialista	Número de Neurônios na Camada Oculta
1	20
2	25
3	10
4	5
5	25
6	20

*O sentido de FN se refere a pacientes que pertencem ao grupo do especialista, mas que foram classificados como não pertencendo ao mesmo. Este sentido é aplicado a todos os grupos exceto ao grupo 6 onde o significado adquire conotação oposta. O sentido de FP é análogo e oposto ao do FN.

5.2.1 Apresentação de Relatório

Utilizando a saída de cada rede conforme (76), pode ser gerado um relatório em forma de tabela. Com base nos resultados pode-se ter uma idéia do grupo ao qual pertence o paciente.

Paciente:	PAC 15
Especialista	Saída 'y'
1	0.9834
2	0.9471
3	0.7942
4	-0,9643
5	-0,9926
6	0,1950

Resultado: Com base na saída de cada especialista, as chances do paciente 15 pertencer ao grupo 1 são grandes, porém há indícios de que pertença ao mesmo tempo ao grupo 2. Quem interpreta o relatório deve inspecionar a saída y de cada ANN e julgar as chances do individuo pertencer ao grupo daquele especialista, quanto mais próximo de +-0.9, mais semelhante é aquele paciente aos apresentados durante o treinamento.

O relatório completo se encontra no Apêndice D.

5.2.2 Curvas de Aprendizado

As curvas de aprendizado são úteis para visualizar o comportamento das redes durante a fase de treinamento, e para determinar o momento de parada. A seguir apresentam-se as curvas de aprendizado para as configurações da Tabela 7, com o número de neurônios ocultos ‘ótimo’ (Figura 37 até a Figura 54). O desempenho final para diferentes números de neurônios se encontra no Apêndice C.

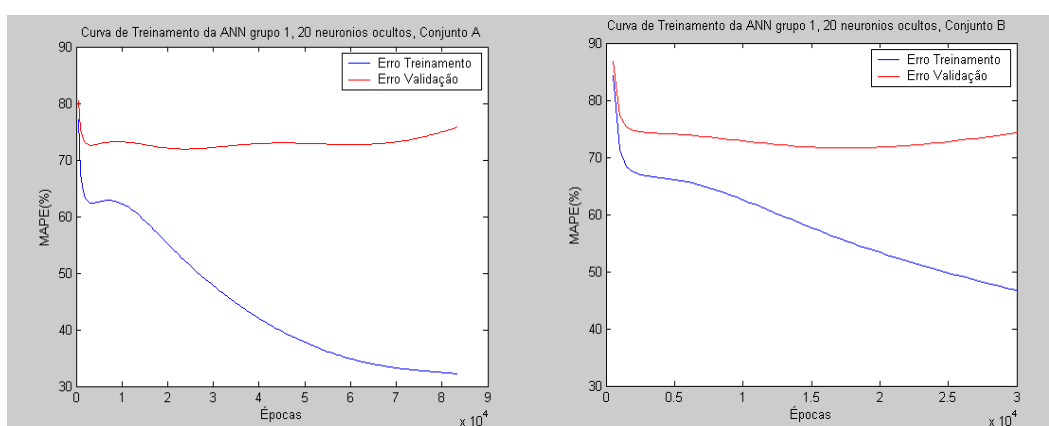


Figura 37: Curvas de aprendizado. Especialista Grupo 1 - Configuração A e B.

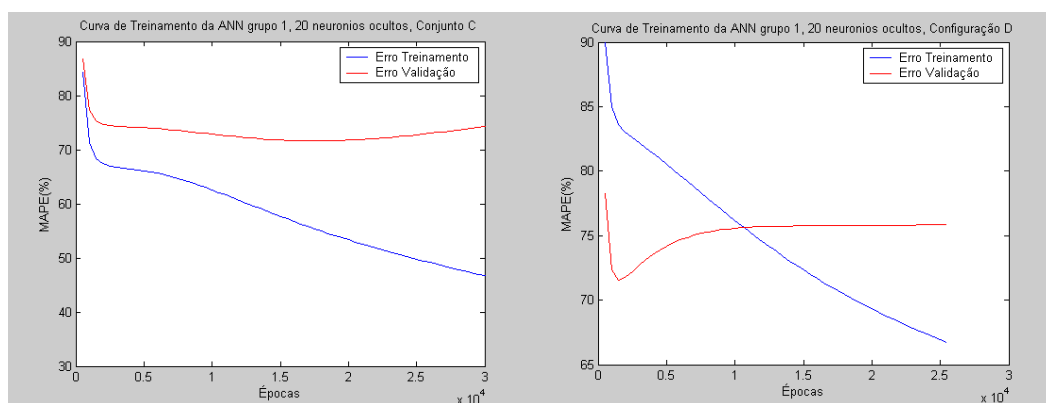


Figura 38: Curvas de aprendizado. Especialista Grupo 1 - Configuração C e D.

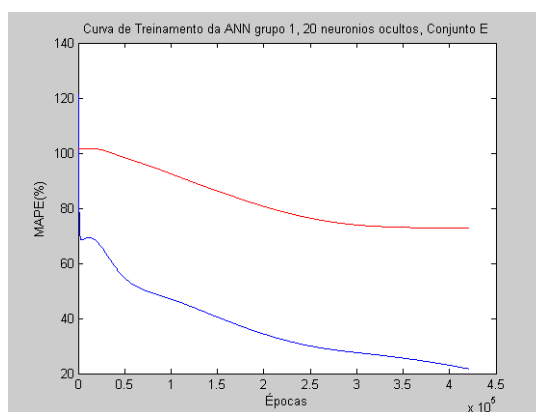


Figura 39: Curvas de aprendizado. Especialista Grupo 1 - Configuração E.

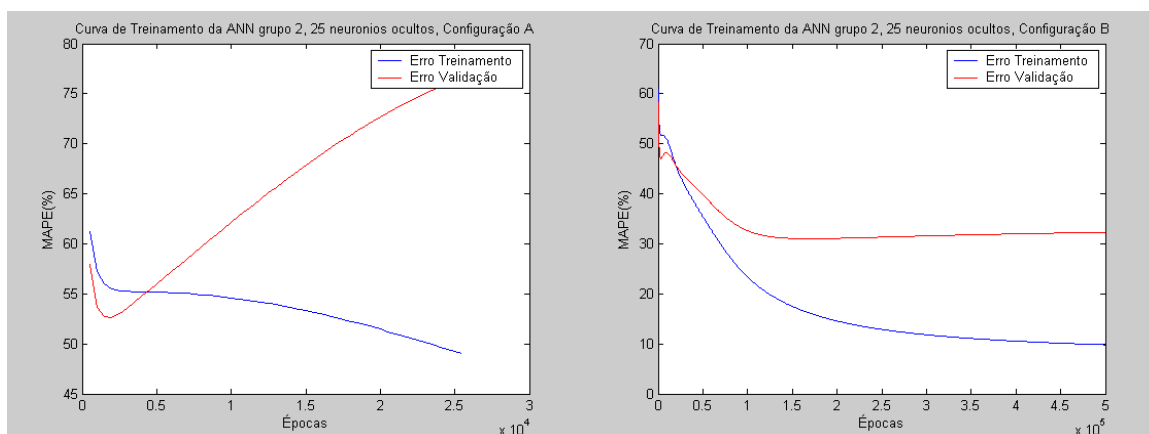


Figura 40: Curvas de aprendizado. Especialista Grupo 2 - Configuração A e B.

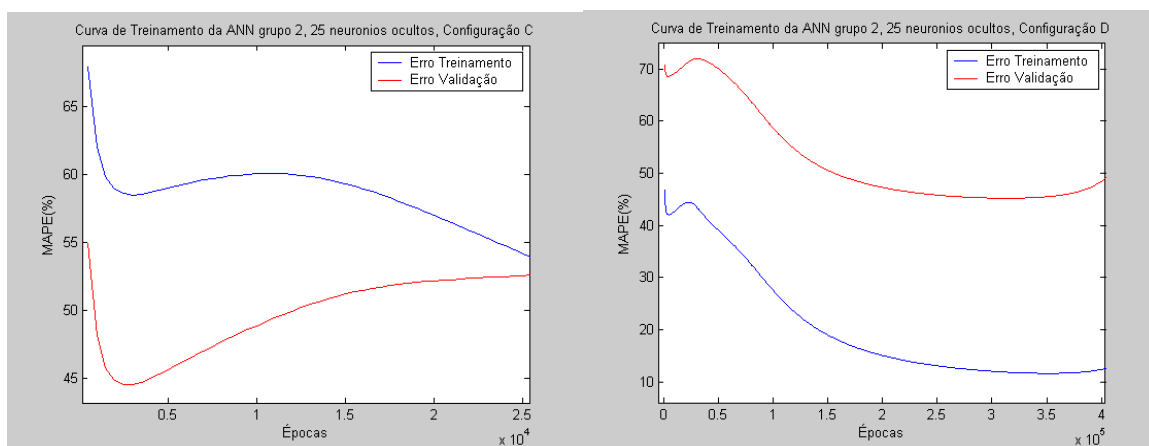


Figura 41: Curvas de aprendizado. Especialista Grupo 2 - Configuração C e D.

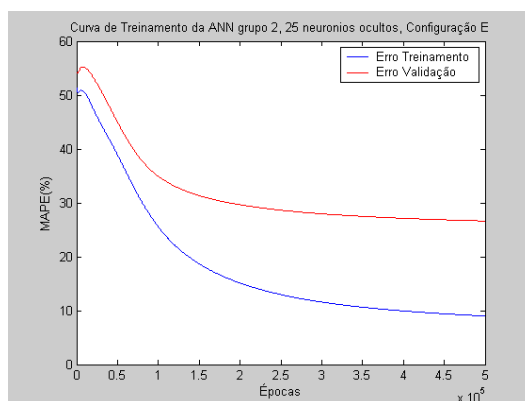


Figura 42: Curvas de aprendizado. Especialista Grupo 2 - Configuração E.

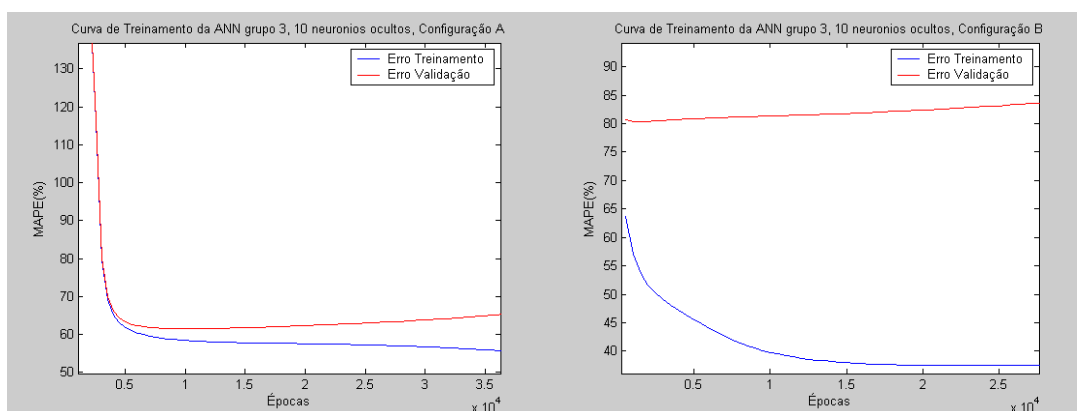


Figura 43: Curvas de aprendizado. Especialista Grupo 3 - Configuração A e B.

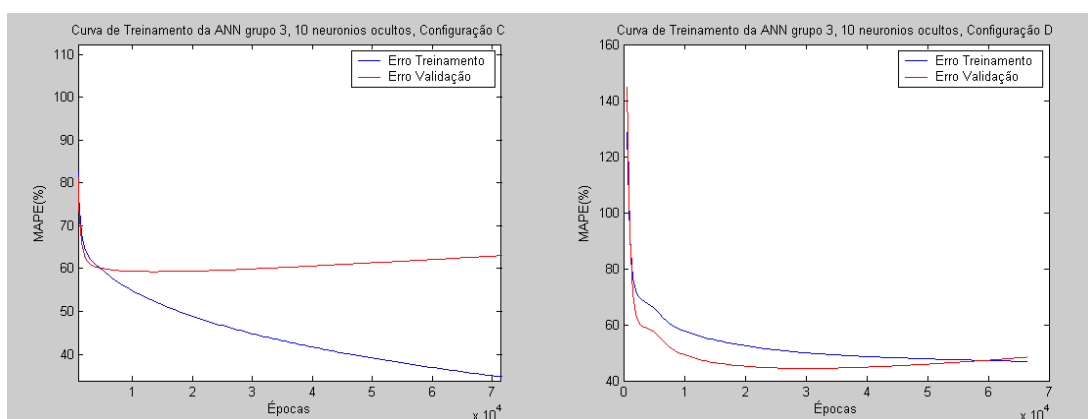


Figura 44: Curvas de aprendizado. Especialista Grupo 3 - Configuração C e D.

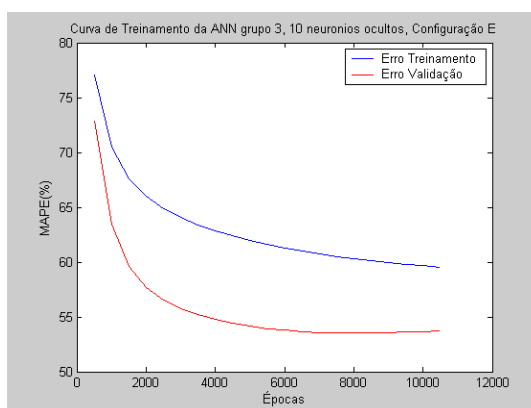


Figura 45: Curvas de aprendizado. Especialista Grupo 3 - Configuração E.

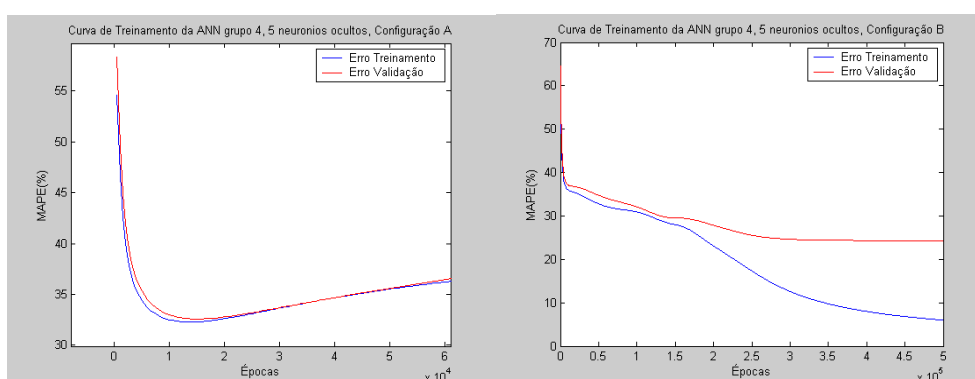


Figura 46: Curvas de aprendizado. Especialista Grupo 4 - Configuração A e B.

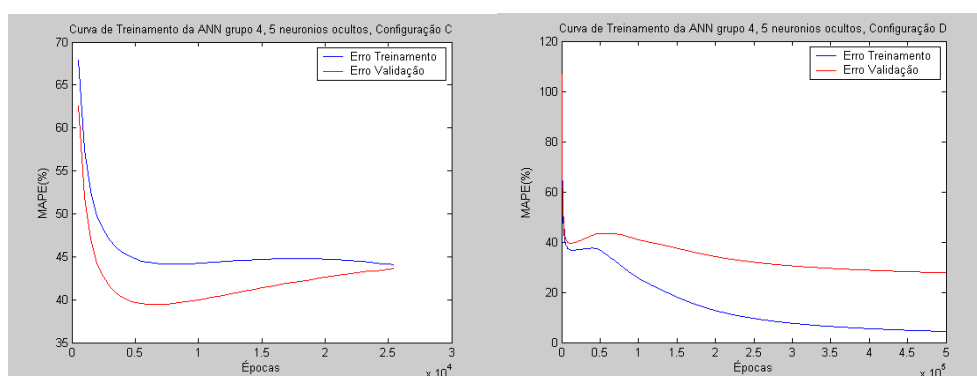


Figura 47: Curvas de aprendizado. Especialista Grupo 4 - Configuração C e D.

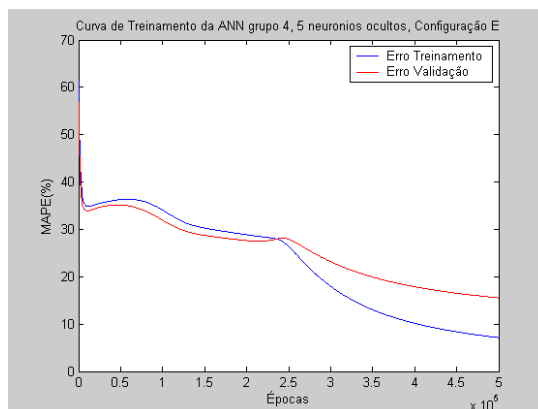


Figura 48: Curvas de aprendizado. Especialista Grupo 4 - Configuração E.

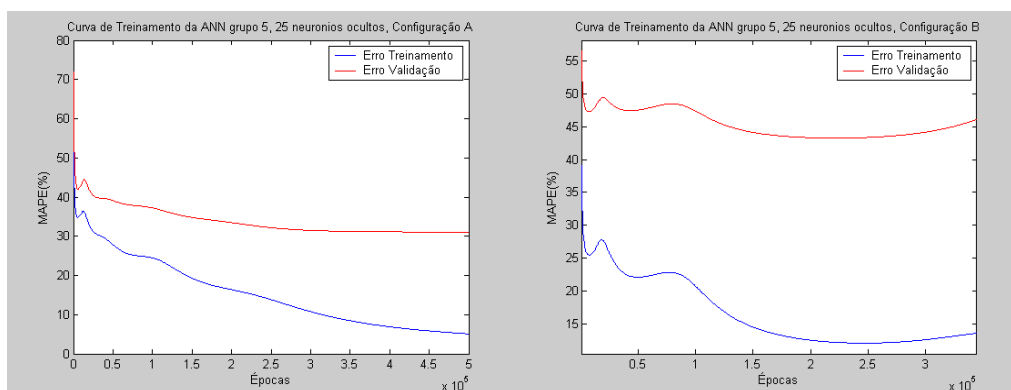


Figura 49: Curvas de aprendizado. Especialista Grupo 5 - Configuração A e B.

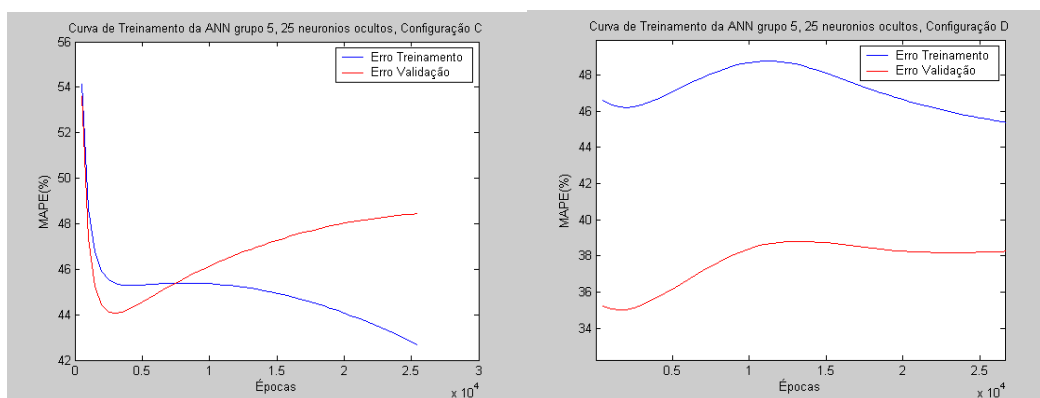


Figura 50: Curvas de aprendizado. Especialista Grupo 5 - Configuração C e D.

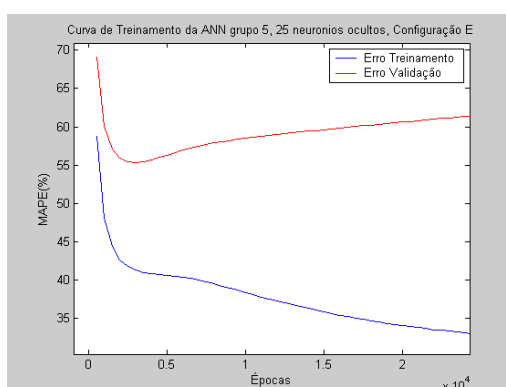


Figura 51: Curvas de aprendizado. Especialista Grupo 5 - Configuração E.

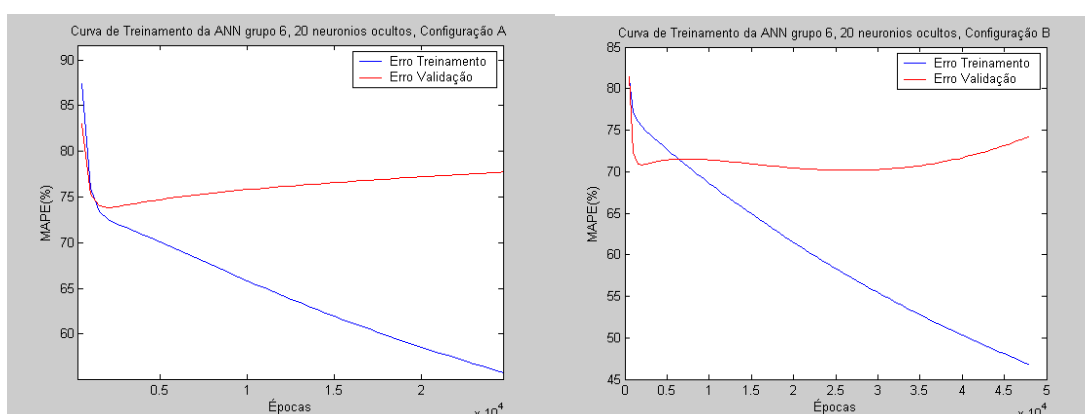


Figura 52: Curvas de aprendizado. Especialista Grupo 6 - Configuração A e B.

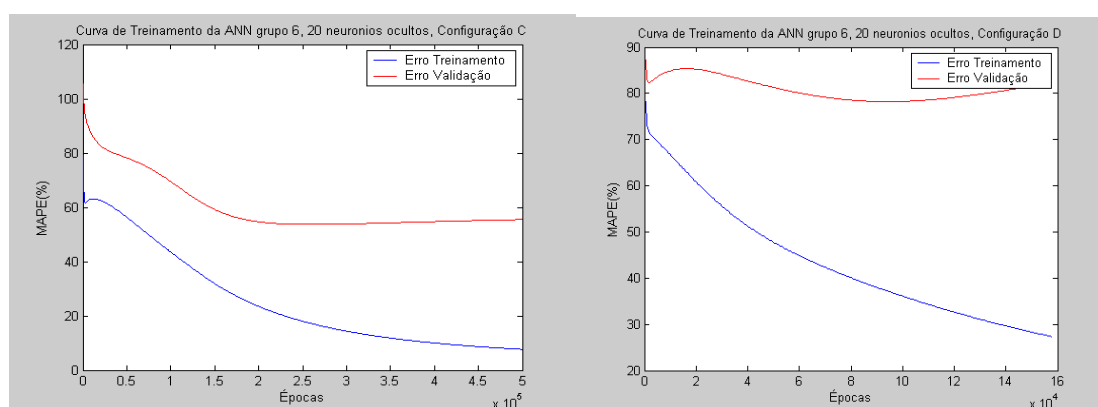


Figura 53: Curvas de aprendizado. Especialista Grupo 6 - Configuração C e D.

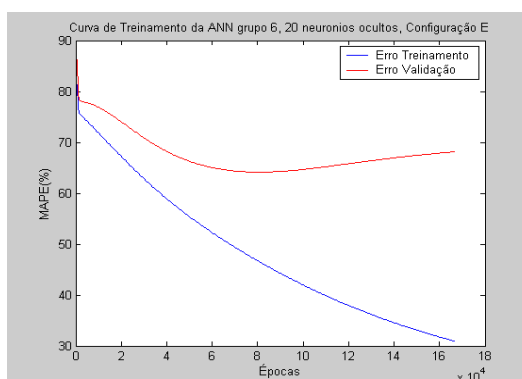


Figura 54: Curvas de aprendizado. Especialista Grupo 6 - Configuração E.

As figuras a seguir (Figura 55 até a Figura 57) mostram as curvas de aprendizado, com apenas o erro de treinamento. Todos os padrões disponíveis do banco de vozes foram utilizados nesta fase. O critério de parada de cada treinamento consiste apenas em treinar pelo número épocas médio (i_m) sobre cada configuração da Tabela 7. Ao treinar cada ANN por i_m épocas, o erro de treinamento obtido se torna aproximadamente o mesmo que a média dos erros de validação sobre cada configuração da Tabela 7.

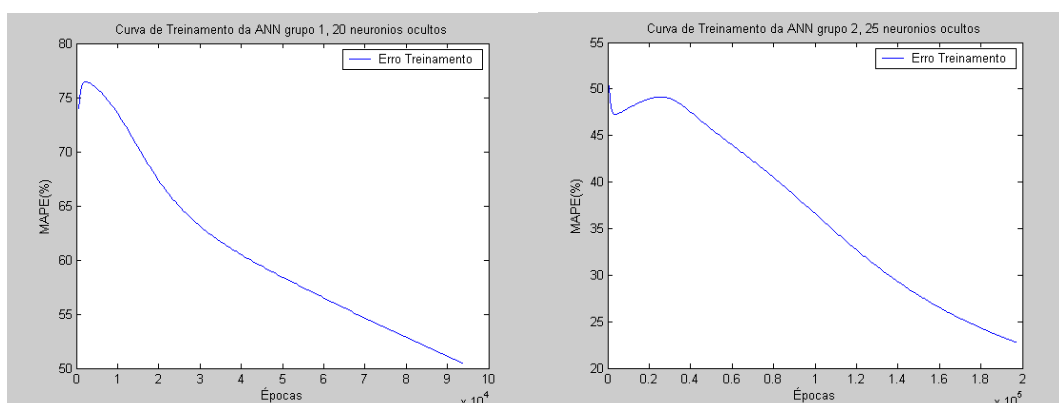


Figura 55: Curvas de Treinamento. Especialista Grupo 1 e 2 - Final.

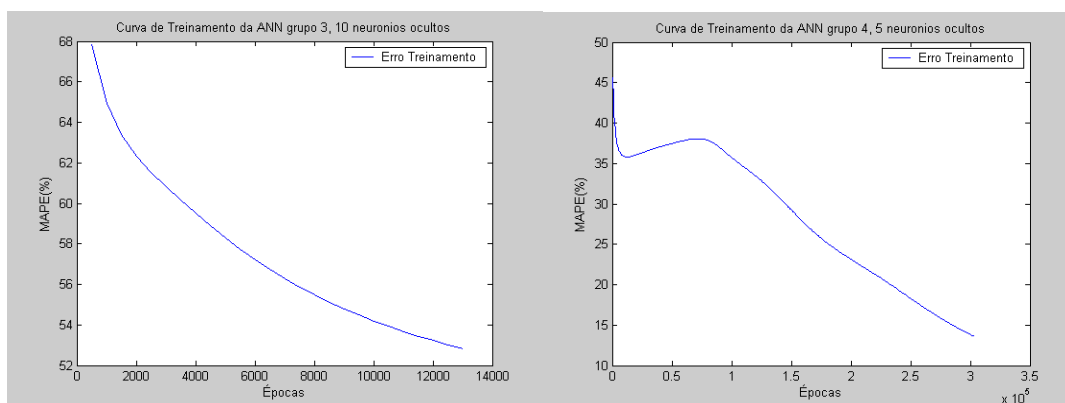


Figura 56: Curvas de Treinamento. Especialista Grupo 3 e 4 - Final.

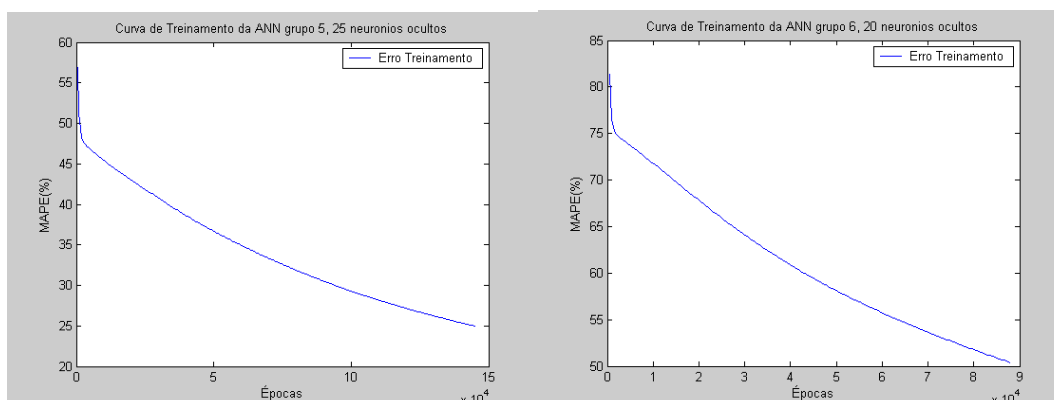


Figura 57: Curvas de Treinamento. Especialista Grupo 5 e 6 - Final.

A Tabela 13, apresenta um resumo dos erros de validação encontrados durante o treinamento de cada ANN para diferentes números de neurônios, de 5 a 30, com incremento de 5 em 5.

Tabela 13: Erros de validação (MAPE) Vs. Número de neurônios ocultos.

Neurônios	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5	Grupo6
5	73,87%	50,44%	58,66%	25,80%	41,37%	64,54%
10	75,57%	45,27%	54,65%	28,12%	44,68%	66,25%
15	72,49%	43,91%	59,95%	36,13%	41,36%	66,41%
20	64,20%	48,01%	61,69%	29,61%	40,23%	59,16%
25	74,57%	36,47%	61,13%	34,70%	37,68%	66,61%
30	73,57%	48,72%	60,60%	33,11%	44,33%	67,59%

A Tabela 14, mostra o resumo dos erros de generalização ao aplicar o *Leave-One-Out*.

Tabela 14: Erros de generalização (MAPE) das ANN aplicando o LOO.

Especialista	Erro de Generalização
Grupo1 (20 neurônios ocultos)	43.65%
Grupo2 (25 neurônios ocultos)	35.94%
Grupo3 (10 neurônios ocultos)	48.59%
Grupo4 (5 neurônios ocultos)	27.69%
Grupo5 (25 neurônios ocultos)	31.93%
Grupo6 (20 neurônios ocultos)	43.32%

Estes erros de generalização podem ser considerados muito elevados, mas os erros de classificação destas redes mostram que mesmo assim as ANN conseguem determinar o grupo de patologias de forma aproximadamente correta.

Uma forma de medir o erro de generalização do sistema como um todo, poderia ser calculando o valor médio dos erros da Tabela 14. O valor é de 38.52%.

6 DISCUSSÕES E CONCLUSÕES

No trabalho anterior desta linha de pesquisa (CROVATO, 2004), os padrões utilizados no treinamento das ANN foram aqueles que apresentavam maior dificuldade de serem generalizados. Desta forma forçava-se à rede a aprender as características mais difíceis presentes em cada grupo de patologia; assim a tarefa de generalizar ficava mais fácil, possibilitando a redução do erro de classificação, que atingiu a taxa de 13,11%, 4,92%, 4,92%, 6,56%, 8,20%, 13,11% para os grupos 1 ao 6 respectivamente. Comparando estas taxas de erro com as encontradas neste trabalho (12,5%, 4,68%, 12,5%, 0%, 3,12%, 10,93%) obteve-se agora uma melhora de desempenho. Mas o número de neurônios ocultos também é consideravelmente maior que o encontrado em (CROVATO, 2004) que era de 6, 8, 6, 6, 8, 4 contra 20, 25, 10, 5, 25, 20 neste trabalho, para os grupos 1 ao 6 respectivamente.

Quanto maior o número de neurônios, a ANN possui mais parâmetros ajustáveis no treinamento possibilitando se ajustar melhor ao espaço das entradas, este é um dos motivos pelo qual o erro de classificação diminui neste trabalho.

As taxas de acerto encontradas neste trabalho são compatíveis com as encontradas na literatura, como é o caso de 62,33% na discriminação de patológicos e normais em (ROSA, 1998); 95,1% em (HADJITODOROV, 2000); 91% em (MARTINEZ, 2000); 54,79% (ROSA, 2000); 91% em (JO, 2001); 67,2% em (PARRAGA, 2002); 84% em (SCHUCK, 2003); 86,89% em (CROVATO, 2004) e 89,07% neste trabalho.

Pacientes com sulco vocálico (PAC5, PAC14, PAC74) não pertencem a nenhuma das 6 classificações propostas por (ROSA, 1998). Assim também os pacientes com algum tipo de pós-operatório são dificilmente classificados em algum tipo de doenças, pois na maioria dos casos após um curto tempo da operação, o paciente recupera sua voz. (PAC25, PAC11, PAC06, PAC10, PAC17). Todos estes pacientes oferecem algum tipo de dificuldade na classificação. Um novo grupo chamado "Nenhum Grupo de Doenças" foi criado (Grupo 7), por não pertencer a nenhum dos 6 anteriores. Nunca se utilizou nenhum paciente do Grupo 7 para treinar alguma ANN especialista.

Ao avaliar os pacientes do Grupo 7 pelo sistema, as classificações realizadas podem estar incorretas porque não existe nenhuma ANN especialista neste grupo. Os resultados mostram que os pacientes PAC25 e PAC11, ambos senhores com pós-operatório (laringectomia), foram identificados pelo sistema, como pertencendo ao grupo da laringite crônica (Grupo 1). O paciente PAC14, uma mulher com pós-operatório de cisto e em tratamento fonoterápico, possuía sulco vocálico; e foi identificado pelo sistema como pertencendo ao grupo dos pacientes com voz normal (Grupo 6). Os demais indivíduos do Grupo 7 foram classificados corretamente como não pertencendo a nenhum grupo das ANN.

O sistema proposto se mostrou capaz de classificar as vozes do banco de vozes nos respectivos grupos, com um alto grau de sucesso, 100% no melhor caso e 87.5% no pior caso.

Erros de generalização grandes (como os da Tabela 14) e erros de classificação pequenos (como os da Tabela 9) mostram que estas redes neurais acertam bastante o grupo de patologias dos pacientes apresentados, mas não estão muito “convictas” disso. Isso pode ser observado pelo valor “ y_i ” (da Equação 76), no Apêndice D. Sempre que valores “ y_i ” são próximos a 0.9 significa que a rede foi apresentada a um paciente que ela considera “com muita certeza” pertence ao grupo de sua “especialidade”, pois apresenta saída semelhante à prevista para um paciente que pertence ao grupo de sua “especialidade” durante a fase de treinamento.

Segundo a Tabela 14, quase todos os especialistas (com exceção da ANN do grupo 4 e grupo 5), têm erros de generalização elevados, mas mesmo assim acertam o grupo de patologias.

Uma questão ainda pendente é a formação de um conjunto de teste que contenha pacientes de todos os grupos de patologias para avaliar melhor a performance do sistema. O próprio conjunto de treinamento é limitado em número, a sugestão para um futuro trabalho é ampliar o mesmo.

Também é possível continuar a linha de pesquisa testando outras funções custo no algoritmo BBA, como a sugerida no artigo de (YEN, 2002) avaliando o desempenho do sistema. Também é possível investigar o uso de ANN do tipo Funções de Base Radial (RBF), ou até aplicar outros métodos estatísticos para melhorar a performance.

Uma outra sugestão que fica é uma solução num contexto mais amplo, a elaboração de um software que juntamente com um microcomputador com placa de som, seja capaz de realizar aquisição de voz de pacientes suspeitos de doença nas pregas vocais, armazenar este sinal junto com os dados do paciente, em forma de base de dados (disponível para análise *off-line*) além de possuir as características descritas em (KENT, 1993) sobre este tipo de software (possibilidade de visualizar espectro, estimativas dos formantes via LPC, espectrograma, etc), e finalmente, que seja capaz de realizar uma avaliação quantitativa sobre as possibilidades do paciente estar ou não com algum grupo de disfunções. Para isso este sistema poderia utilizar a técnica apresentada neste trabalho, e desta forma ajudar ao diagnóstico.

7 REFERÊNCIAS

ACCARDO, A. P.; ENZO, M. An algorithm for the automatic differentiation between the speech of normals and patients with Friedreich's ataxia based on the short time fractal dimension. **Computers in Biology and Medicine**. Elmsford: Elsevier, v. 28, n. 1, p. 75-89, jan. 1998.

AKAY, M. **Time Frequency and Wavelets in Biomedical Signal Processing**. Dartmouth College: IEEE Press, 1998. 774 p. ISBN 0-7803-1147-7.

BAKEN, R. J. **Clinical Measurement of Speech and Voice**. Needham Heights, MA: Allyn and Bacon, a division of Simon and Schuster, Inc., 1987. ISBN: 1-565-93869-0.

BALMELLI, L.; MOJSILOVIC, A. Wavelet domain features for textures description, classification and replicability analysis. In: CONFERENCE ON IMAGE PROCESSING (ICIP), **Proceedings ...**[S. l.]: IEEE, v. 4, p. 440-444, out. 1999.

BENDAT, J. S.; PIERSOL, A. G. **Random Data: Analysis and measurement procedures**. 2.ed. John Wiley & Sons, Inc., 1986. 574p. ISBN: 0-471-31773-0.

BENTLEY, P. M.; McDONNELL. Wavelet transforms: an introduction. **Electronics & Communication Engineering Journal**. New York, v. 6, n. 4, p. 175-186, ago. 1994.

BENEDILTSSON, J. A. et al. Parallel consensual neural networks. **Transactions on Neural Networks**. New York. IEEE, v. 8, n. 1, p 54-64, jan.1997.

BISHOP, C. M. **Neural Networks for Pattern Recognition**. Oxford University Press, 1997. ISBN: 0198538642.

BOONE, D. R.; McFARLANE S. C. **A voz e a terapia vocal**. 5. ed., Porto Alegre: Artes médicas, 1994. ISBN: 85-7307-036-6.

BURRUS, C. S.; GOPINATH A., GUO H. **Introduction to Wavelets and Wavelet Transforms: a primer**. Prentice-Hall, 1998. ISBN: 01-3489-600-9.

CHANG, S. et al. Speech feature extracted from adaptive wavelet for speech recognition. **Electronics Letters**. [S.l.], v. 34, n. 23, p. 2211-2213, 12 nov. 1998.

CHILDERS, D. G.; BAE K. S. Detection of laryngeal function using speech and electroglottographic data. **IEEE Transactions on Biomedical Engineering**. New York: IEEE, v. 39, n. 1, p. 19-25, jan.1992.

COIFMAN, R. R.; WICKERHAUSER, M. V. Entropy-based algorithms for best basis selection. **Transactions on Information Theory**. New York: IEEE v. 38, n.2, p.1713-1716, mar. 1992.

COIFMAN, R. R.; WICKERHAUSER, M. V. Experiments with adapted wavelet de-noising for medical signals and images. In: **TIME FREQUENCY AND WAVELETS IN BIOMEDICAL SIGNAL PROCESSING**, ser. **IEEE Press Series in Biomedical Engineering**. New York: M. Akay, Ed., 1998, n.12, p. 323-346. 1998.

COLTON, R. H.; CASPER, J. K. **Compreendendo os Problemas de Voz**. Porto Alegre: Artes médicas, 1996. ISBN 85-7307-106-0.

CROVATO, C. D. P.; SCHUCK Jr., A. Classificação de sinais de voz utilizando a transformada wavelet packet e redes neurais artificiais. In: **CONGRESSO LATINOAMERICANO DE ENGENHARIA BIOMEDICA**. set. 2004, João Pessoa, PB. **Proceedings...**, João Pessoa, PB: IFMBE, 2004, p. 1027-1030, v.5, n.1, ISBN: 85-987390104.

DAUBECHIES, I. The wavelet transform, time-frequency localization and signal analysis. **IEEE Transaction on. Information Theory**. New York, v. 36, 1990.

DAUBECHIES, I. **Ten Lectures on Wavelets**. Soc. for Industrial & Applied Math., 1992. 375p. ISBN: 0-89871-274-2.

DELLER Jr., J. R.; ANDERSON, D. J. Automatic classification of laryngeal dysfunction using the roots of the digital inverse filter. **IEEE Trans. on Biom. Eng.**, New York, v. BME 27, p.714-721, 1980.

DELLER Jr., P. R.; PROAKIS, J. G.; HANSEN, J. H. L. **Discrete-Time Processing of Speech Signals**. New York, USA. Macmillan Publishing Company, 1993. ISBN: 023283017.

De YAO, MOHAMMED, R. ASIMI-SADJADI. A study of effects of sonar bandwidth for underwater target classification. **IEEE Journal of Oceanic Engineering**. New York, v. 27, n. 3, jul. 2002.

ENGLEHART, K. et al. Classification of the myoelectric signal using time-frequency based representations. **Medical Engineering & Physics**. [S.l.], Elsevier, v. 21, n. 6-7, p. 431-438, jul. 1999.

KAVANAGH, K. ENT-USA Ear, Nose & Throat, web page. Disponível em: <<http://ent-usa.com>>. Acesso em: 10 dez. 2004.

FREEMAN, J. A.; SKAPURA, D. M. **Neural Networks: Algorithms, applications and programming techniques**. Addison-Wesley Publishing Company Inc. jul. 1992. ISBN:0-201-51376-5.

FRITZELL, B. Inverse filtering. **Journal of Voice**. New York: Raven Press. Ltd. v. 6 n. 2 p. 111-114, 1992.

FROHLICH, M.; MICHAELIS, D.; WERNER, S. H. Acoustic 'breathiness measures' in the description of pathologic voices. In: International Conference on Acoustics Speech and Signal Processing, 1998. [S.l.], **Proceedings...**New York: IEEE. 1998, v. 2, p.937-940.

FUNAHASHI K. On the approximate realization of continuous mapping by neural networks. **Neural Networks**. Oxford, UK: Elsevier Science Ltd., v. 2, p.183-192, 1998. ISSN: 08936080.

GARCIA, C.; ZIKOS, G.; TZIRITAS, G. Wavelet packet analysis for face recognition. **Image and Vision Computing**, [S.l.] v. 18, n. 4, p. 289-297, mar. 2000.

GHORAYEB, B.Y., Disponível em <<http://www.ghorayeb.com>>. Acesso em: 20 abr. 2004.

GUOXIN, K. et al. A method of spectra-analysis for abnormal voice signals: engineering in medicine and biology society, images of the twenty-first century. In: Annual International Conference of the IEEE Engineering, 1989, [S.l.]. **Proceedings...New York, IEEE**, 1989.

HADJITODOROV, S.; BOYANOV, B.; TESTON, B. Laryngeal pathology detection by means of class-specific neural maps. **IEEE Transactions on Information Technology in Biomedicine**. [S.l.], v. 4 n. 1, p. 68-73, mar. 2000.

HAYKIN, S. **Redes Neurais: Princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. ISBN: 8-5730-77-18-2

HINARO, M. **Videostroboscopic Examination of the Larynx**. San Diego, California: Singular Publishing Group, Inc., 1993. ISBN: 1-879105-52-7.

HONG, J.P. et al. Heart sound recognition through analysis of wavelet transform and neural network. **IEIC Trans. Inf. & Syst.** [S.l.], v. E86-D. n. 6, jun. 2003.

HSIN; HSI-CHIN. Texture segmentation using modulated wavelet transform. **IEEE Transactions on Image Processing**. [S.l.], v. 9, n. 7, p. 1299-1302, jul. 2000.

HUANG, Q. et al. Underwater target classification using wavelets packets and neural networks. **IEEE Trans. On Neural Network**. [S.l.], v. 11, n. 3, mai. 2000.

INTRATOR, N. et al. Wavelet feature extraction for discrimination tasks. In: CANADIAN WORKSHOP ON INFORMATION THEORY, 1997. **Proceedings...Toronto**, jun. 1997.

JIANG, Q.; GOH, S. S.; LIN, Z. Local discriminant time-frequency atoms for signal classification. **Elsevier: Signal Processing**. [S.l.], v. 72, n. 1, p. 47-52, 4 jan. 1999.

JO, C. et al. Screening of pathological voice from ARS using neural networks. In: INTERNATIONAL WORKSHOP ON MODELS AND ANALYSIS OF VOCAL EMISSIONS FOR BIOMEDICAL APPLICATIONS, 2., 2001, Firenze-Italy, **Proceedings...** Firenze-Italy, set. 2001.

KADAMBE, S.; BOUDREAUX-BARTELS, G. F. Application of the wavelet transform for pitch detection of speech signals. **IEEE Trans. on Inf. Theory**. [S.l.], v. 38, p. 917-924, mar. 1992.

KADAMBE, S.; ROBIN, M.; BOUDREAUX-BARTELS, G. F. Wavelet transform-based QRS complex detector. **IEEE Transactions on Biomedical Engineering**. New York, v. 46, n. 7, Jul. 1999.

KARRAKCHOU, M., KUNT, M. From continuous wavelet transform to wavelet packet: application to the estimation of pulmonary microvascular pressure. **In: IEEE Metin Akay**, 1998. p. 367-387, ISBN 0-7803-1147-7.

KENT, R. D.; READ C. **The Acoustics Analysis of Speech**. San Diego: Singular Publishing Group Inc. CA.1992. ISBN: 1-879105-43-8.

KENT, R. D. Vocal track acoustics. **Journal Of Voice**. New York: Raven Press. 1993.

KESHAHA, N.; MOURA, J. M. F. Matching wavelet packets to gaussian random processes. **IEEE Transactions on Signal Processing**. New York, v. 47, n. 6, p. 1604-1614,1999.

KOHONEN, T. **Self-Organization and Associative Memory** . Berlin: Springer Verlag, 1984. ISBN: 3-540-18314

KOSKO, B. **Neural Networks For Signal Processing**. California: Prentice-Hall International Editions, 1992. ISBN: 0-13-614694-5.

KOVÁCS, Z. L. **Redes Neurais Artificiais: fundamentos e aplicações**.2. ed. rev. São Paulo: Collegium Congitio, 1996. ISBN: 85-86396-02-8.

KUHL, I. **Manual Prático de Laringologia**. Porto Alegre: Editora da Universidade, 1982. ISBN: 85-7025-062-2.

LIANG, H.; NARTIMO, I. A feature extraction algorithm based on wavelet packet decomposition for heart sound signals. In: TIME-FREQUENCY AND TIME-SCALE ANALYSIS INTERNATIONAL SYMPOSIUM, [S.l.], **Proceedings...** New York: **IEEE** p. 93 -96, 1998.

MACKENZIE, D. et al. Wavelets, seeing the forest and the trees, the path from research to human benefit project. **National Academy of Sciences**. Washington, USA, dec. 2001. Disponível em: < www.beyonddiscovery.org>. Acesso em: ago. 2004.

MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. [S.l.], v. 11, p.674-693, jul. 1989.

MALLAT, S. G.; ZHANG, Z. Matching pursuits with time-frequency dictionaries. **IEEE Transactions on Signal Processing**. [S.l.], v. 41, n. 12, p. 3397-3415, dec. 1993.

MALLET, Y. et al. Classification using adaptive wavelets for feature extraction. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. [S.l.], v. 19, n. 10, p. 1058-1066, oct. 1997.

MANFREDI, C. et al. A comparative analysis of fundamental frequency estimation methods with application to pathological voices. **Medical Engineering & Physics**. [S.l.], v. 22, n. 2, p. 135-147, mar. 2000.

MANFREDI, C. Adaptive noise energy estimation in pathological speech signals. **IEEE Transactions on Biomedical Engineering**. [S.l.], v. 47, n. 11, p. 1538-1543, nov. 2000.

MARTINEZ, C. E.; RUFINER, H. L. Acoustic analysis of speech for detection of laryngeal pathologies. In: INFORMATION TECHNOLOGY APPLICATION IN BIOMEDICINE, 2000. **Proceedings...**[S.l.]: IEEE EMBS, 2000.

MEYER, F. G.; AVERBUCH, A. Z., STROMBERG, J. Fast adaptive wavelet packet image compression. **IEEE Transactions on Image Processing**. [S.l.], v. 9, n. 5, p. 1-9, mai. 2000.

MINSKY, M.; PAPERT, A. **Perceptrons**: an introduction to computational geometry. Cambridge, MA: MIT Press, expanded edition, 1988/1969. ISBN: 0-262631-11-3

NIEDZIELSKA, G.; GLIJER E.; NIEDZIELSKI, A. Acoustic analysis of voice in children with noduli vocales. **Elsevier International Journal of Pediatric Otorhinolaryngology**. [S.l.], v. 60, n. 2, p 119-122, 20 ago. 2001.

OBAIDAT, M. S.; BRODZIK A.; SADOON, B. A performance evaluation study of four wavelet algorithms for the Pitch period estimation of speech signals. **Information Sciences**. [S.l.], v. 112, n. 1-4, p. 213-221, dec. 1998.

OBAIDAT, M. et al. Estimation of pitch period of speech signal using a new dyadic wavelet algorithm. **Information Sciences**. [S.l.], v. 119, n. 1-2, p. 21-39, 1 oct. 1999.

OPPENHEIM, V. A.; SCHAFER W. R. **Discrete-time signal processing**. NY: Prentice- Hall signal processing series, 1989. ISBN: 0-13-754920-2.

PAYA, B. A.; ESAT, I. I.; BADI, M. N. M. Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. **Mechanical Systems and Signal Processing**. [S.l.], v. 11, n.5 p.751-765, set. 1997.

PARRAGA A. **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. 2002, 110 p. Dissertação (Mestre em Engenharia). Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

RAMCHANDRAN, K.; VETTERLI, M. Best wavelet packet bases in a rate distortion sense. **IEEE Transactions on Image Processing**, [S.l.], v. 2, n. 2, p. 160-175, abr. 1993.

RIOUL, O.; VETTERLI, M. Wavelet and signal processing. **IEEE Signal Processing Magazine**. v. 8 n. 4, out. 1991.

RITCHINGS, R.T.; McGUILLION, M.; MOORE C.J. Pathological voice quality assesment using artificial neural networks. In: INTERNATIONAL WORKSHOP ON MODELS AND ANALYSIS OF VOCAL EMISSIONS FOR BIOMEDICAL APPLICATIONS. 2. Firenze-Italy, **Proceedings...**IEEE: Firenze-Italy, Set. 2001.

ROCHA, B., Disponível em <<http://www.corpohumano.com.br>>. Acesso em: 20 abr. 2004.

RONTAL, E.; RONTAL, M.;ROLNICK, M. I. Objective evaluation of vocal pathology using voice spectrography. **Annals of Otology, Rhinologi and Laryngology**. [S.l.] v.84, p.662-671. 1975.

ROSA, M. de O.; PEREIRA, C. J; CARVALHO, A. Evaluation of neural classifiers using statistic methods for identification of laryngeal pathologies. In: BRAZILIAN SYMPOSIUM ON NEURAL NETWORKS, **Proceedings...** [S. l.], 1998.

ROSA, M. de O.; GRELLER, M.; CARVALHO, A. Signal processing and statistical procedures to identify laryngeal pathologies. In: IEEE ELECTRONICS, CIRCUITS AND SYSTEMS SOCIETY. 6. 1999, **Proceedings...** [S. l.], 1999.

ROSA, M. de O.; PEREIRA, M. J. C.; GRELLET, M. Adaptive estimation of residue signal for voice pathology diagnosis. **IEEE Transactions on Biomedical Engineering**. [S.l.], v. 47, n. 1, p. 96-104, jan. 2000.

RUMELHART, D.E., HINTON, G.E., WILLIAMS, R.J. Learning internal representations by error propagation. In: PARALLEL DISTRIBUTED PROCESSING: explorations in the Microstructure of Cognition. Cambridge, MA: The MIT Press. 1986, v. 1, p. 318-362. ISBN:0-262-68053.

SAITO, N.; COIFMAN, R. R. Local discriminant bases, In: A. F. LAINE AND A. M. UNSER, ed., Mathematical Imaging: **Wavelet Applications in Signal and Image Processing**. Proc. SPIE, [S.l.], v. 2303, jul. 1994.

SARIKAYA, R.; HANSEN, J. High resolution speech feature parametrization for monophone-based stressed speech recognition. **IEEE Transactions on Processing Letters**, [S.l.], v. 7, n. 7, p. 182-185, jul. 2000.

SCHUCK Jr., A. **Análise e processamento de sinais de voz disfônica através da transformada wavelet discreta**. 1998. 143 p. Tese (Doutor em engenharia). Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais -PPGEM, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1998.

SCHUCK Jr., A. , GUIMARÃES, L.V., WISBECK, O. Dysphonic Voice Classification using Waveltes Packets Transform and Artificial Neural Networks. In: ANNUAL INTERNATIONAL CONFERENCE IEEE EMBS, 25. 2003, Cancún, MX, **Proceedings...** Cancún, MX, 2003 p. 2958-2961.

SHENG, Y. **Wavelet Transform**. The Transforms and Applications Handbook: Boca Raton: CRC Press in cooperation with IEEE press, 1996.

SHIRAZI, M. N.; NODA, H.; TAKAO, N. Texture classification based on Markov modeling in wavelet feature space. **Image and Vision Computing**, [S.l.], v. 18, n. 12, p. 967-973, set. 2000.

SIMPSON, D. M. An introduction to the discrete orthogonal wavelet transform. **Revista Brasileira de Engenharia**. v. 9 n. 1, p. 57-81, 1993. Caderno de engenharia Biomédica.

SOSTARIC, A.; ZAZULA, D.; DONCARLI, C. Feature extraction using wavelet packets. In: B. ZAJC (ed.), **ELECTROTECHNICAL AND COMPUTER SCIENCE CONFERENCE ERK**, 6. 1997, Portoroz, Slovenia, Ljubljana. **Proceedings...IEEE: Region 8, Slovenian**, 1997, p. A/159-162. Set. 1997.

TASWELL, C. Satisfying search algorithms for selecting near-best bases in adaptive tree-structured wavelet transforms, **IEEE Transactions on Signal Processing**, [S.l.], v. 44, n.10, p. 2423-2438, out. 1996.

TASWELL, C. Near-Best Basis Selection Algorithms With Non-Additive Information Cost Functions. in: M. G. Amin, ed., **INTERNATIONAL SYMPOSIUM ON TIME-FREQUENCY AND TIME-SCALE ANALYSIS**, 1994, Philadelphia, PA, **Proceedings...IEEE-SP**. p. 13-16, out. 1994.

UPMC University of Pittsburg Medical Center. Web Page. Disponível em: <<http://voicedisorders.upmc.com>> acesso em: Abri. 2004.

WERBOS, P.J. Backpropagation through time: what it does and how do it. **Proceedings of the IEEE**. [S.l.], v. 78, n. 10, out. 1990.

WICKERHAUSER, M. V. **Adapted Wavelet Packet Analysis from Theory to software**. Wellesley, MA: AK Peters, Ltd., 9 mai. 1994. ISBN: 1-56881-041-5.

WIDROW, B.; LEHR, M. A. 30 Years of adaptive neural networks: perceptron, madaline, and backpropagation. **Proceedings of the IEEE**. NY, v. 78, n.9, p. 1415-1442 set. 1990.

WILAMOWSKI, B.M. Neural network architectures and learning. **IEEE International Conference on Industrial Technology ICIT**. Maribor, Solenia: IEEE, 2003.

VAN De WOUWER, G.; SCHEUNDERS, P.; VAN DYCK D. Wavelet-FILVQ classifier for speech analysis. **Vision Lab**. Belgium. RUCA (University of Antwerp). 1997.

WU, Y.; DU, R. Feature extraction and assessment using wavelet packets for monitoring of machining processes. **Mechanical Systems and Signal Processing**. [S.l.], v. 10 n. 1 p.29-53, jan. 1996.

XIAOLI, L.; ZHEJUN Y. Tool wear monitoring with wavelet packet transform-fuzzy clustering method. **Wear**. v. 219, n 2, p. 145-154, set. 1998.

YEN, G. G.; KUO-CHUNG, L. Wavelet packet feature extraction for vibration monitoring. **IEEE Transactions on Industrial Electronics**. [S.l.], v. 47, n.3, p. 650-667, 2000.

8 APÊNDICES

8.1 APÊNDICE A

No	Pacientes	Descrição	Patologia
1	PAC01	Professora de canto, voz normal	Micronódulo nas pregas vocais
2	PAC02	Cantor nativista, vibrato na emissão, voz normal	Pregas normais
3	PAC04	Rapaz cantor, voz muito rouca	Laringite crônica, possível uso de drogas
4	PAC05	Mulher, voz soprosa	Sulco vocálico
5	PAC06	Mulher, voz soprosa.	Edema pós-cirúrgico
6	PAC08	Rapaz pré-cirúrgico	Voz supra-glótica (ventricular)
7	PAC09	Paciente, 25 dias sem falar	Pós-laringite por abuso vocal
8	PAC10	Paciente de 78 anos, voz quase normal	Paralisia unilateral bem compensada
9	PAC11	Paciente, pós-operatório de 6 anos, feita neocorda	Laringectomia fronto-lateral
10	PAC13	Voz mal-compensada	Paralisia uni-latéral
11	PAC14	Paciente pós-operatória de cisto, em tratamento fonoterápico	Sulco vocálico
12	PAC15	Fumante	Leucoplasia+laringite crônica
13	PAC16		Rinite e Sinusite alérgica
14	PAC17	Paciente de 66 anos, pós operatório+radioterapia	CA na laringe, lesão prega vocal direita
15	PAC18	Paciente fonastênica, cansaço na voz	Pregas normais
16	PAC19	Estudante fonoaudiologia, voz normal	Pregas normais
17	PAC21	Paciente de 16 anos, voz rouca por secreção na traquéia	Cisto pequeno na prega vocal direita
18	PAC22	Paciente, pós-operatório de 3 semanas	Leucoplasia
19	PAC23	Paciente fumante, voz rouca e grave	Edema de Reike
20	PAC24	Paciente, voz muito soprosa e rouca	Cisto, uma prega vocal inchada e sem ondas
21	PAC25	Paciente, pós operatório 45 laringectomia parcial	Laringectomia fronto-latéral
22	PAC26	Voz limpa, tratamento fonoterápico	Disfonia funcional+micro-nódulo
23	PAC27	Paciente, voz limpa mas levemente soprosa, tratamento fonoterápico	Disfonia funcional
24	PAC28	Voz limpa, tratamento fonoterápico	Disfonia funcional
25	PAC29	Voz limpa, fina e levemente rouca, tratamento fonoterápico	Disfonia funcional
26	PAC30	Paciente, voz grossa e rouca	Disfonia funcional+laringite
27	PAC31	Voz alterada	Paralisia prega vocal direita
28	PAC32	Voz grossa e rouca, fumante	Edema de Reike+laringite crônica
29	PAC34	Voz com sibilção	Papilomas
30	PAC35	Paciente com voz rouca e grossa	Edema de Reike-pregas frouxas
31	PAC36		Edema de Reike+formações polipóides
32	PAC37	Voz levemente soprosa, pós-cirúrgica de papilôma recorrente	Pregas vocais limpas
33	PAC41	Paciente meia idade, voz rouca	Neoplasia na laringe
34	PAC42	Paciente pós-cirúrgico de muito tempo, revisão. Voz muito rouca	Laringectomia fronto-latéral
35	PAC43	Paciente 59 anos	Paralisia unilateral prega esquerda
36	PAC44	Voz normal	Pregas vocais normais
37	PAC45	Voz normal	Pregas vocais normais

No	Pacientes	Descrição	Patologia
38	PAC46	Voz muito soprosa, em fonoterapia	Paralisia prega vocal direita
39	PAC47	Voz normal	Pregas vocais normais
40	PAC48	Menina com voz normal, Angela Maria	Pregas vocais normais
41	PAC49	Senhor com voz normal	Pregas vocais normais
42	PAC50		Micronódulo nas pregas vocais
43	PAC51		Laringite + Uso Aerolyn
44	PAC53	Moça, voz rouca e soprosa	Em teste
45	PAC54	Paciente com voz normal (talvez alteração ressonância)	Linfonodo no pescoço
46	PAC55	Cantor nativista, Padrão Fry, voz normal	Pregas vocais normais
47	PAC56	Paciente, voz normal, língua presa, jitter e shimmer visualmente baixos	Prega vocal direita vascularizada
48	PAC57	Voz muito ruim	Laringectomia fronto-latéral
49	PAC58	Paciente, com quebra de fonação, voz rouca	Granuloma abaixo prega vocal direita
50	PAC59	Cantor nativista, voz rouca e cansada.	Leucoplasia+ Irritação
51	PAC60	Voz rouca	Laringectomia fronto-latéral
52	PAC61	Paciente, voz normal, nervosa	Pregas vocais normais
53	PAC62	Fumante, pouco tempo de fonação	Leucoplasia+ Laringite crônica
54	PAC63	Voz horrível	Leucoplasia+ Laringite crônica
55	PAC64	Estudante fonoaudióloga, voz normal, nervosa	Pregas vocais normais
56	PAC66	Paciente idoso	Presbifonia
57	PAC67	Ex-fumante	Secreção nas pregas
58	PAC68	Voz normal	Pregas vocais normais
59	PAC69	Rapaz com diplofonia	Pólipo grande na prega vocal
60	PAC70	Voz normal, pós-cirúrgico de tireóide à Laser	Pregas vocais limpas
61	PAC71	Criança	Nódulo bilateral
62	PAC72	Paciente pós-operatória de 3 dias	Pregas vocais normais
63	PAC73	Paciente, voz áspera	Ceratose+Laringite crônica
64	PAC74	Adolescente, voz levemente soprosa.	Fenda Posterior aumentada

8.2 APÊNDICE B

O limiar de classificação adotado para considerar o indivíduo como pertencente ao grupo de patologias foi de “0”, ou seja, se a saída for ‘+’ o indivíduo pertence ao grupo de patologias, se for ‘-’ não pertence.

Especialista Grupo 1			Especialista Grupo 2			Especialista Grupo 3		
Paciente	Saída	Desej.	Paciente	Saída	Desej.	Paciente	Saída	Desej.
PAC01	-0.8232	-	PAC01	-0.9747	-	PAC01	-0.6038	-
PAC02	-0.7305	-	PAC02	-0.8747	-	PAC02	-0.5288	-
PAC04	-0.5096	+	PAC04	-0.9488	-	PAC04	-0.5460	-
PAC06	-0.6555	-	PAC06	-0.6340	-	PAC06	-0.4531	-
PAC09	-0.1879	+	PAC09	-0.9922	-	PAC09	-0.6753	-
PAC10	-0.6895	-	PAC10	-0.5056	-	PAC10	-0.4706	+
PAC13	-0.9703	-	PAC13	-0.9642	-	PAC13	-0.0691	+
PAC15	0.9834	+	PAC15	0.9471	+	PAC15	-0.7942	-
PAC18	-0.6014	-	PAC18	-0.8336	-	PAC18	-0.5292	-
PAC19	-0.6896	-	PAC19	-0.8241	-	PAC19	-0.3843	-
PAC21	-0.4972	-	PAC21	-0.8518	-	PAC21	-0.7726	-
PAC22	-0.9102	-	PAC22	0.5961	+	PAC22	-0.5310	-
PAC23	-0.6796	-	PAC23	-0.7771	-	PAC23	-0.5192	-
PAC24	-0.7403	-	PAC24	-0.9815	-	PAC24	-0.6486	-
PAC26	-0.7638	-	PAC26	-0.7465	-	PAC26	-0.4175	+
PAC27	-0.7677	-	PAC27	-0.9078	-	PAC27	-0.3050	+
PAC28	-0.4534	-	PAC28	-0.9659	-	PAC28	-0.5156	+
PAC29	-0.4338	-	PAC29	-0.9910	-	PAC29	-0.6967	+
PAC30	0.1869	+	PAC30	-0.8914	-	PAC30	-0.5780	+
PAC31	-0.9544	-	PAC31	-0.9545	-	PAC31	-0.5030	+
PAC32	0.4169	+	PAC32	-0.8338	-	PAC32	-0.7058	-
PAC34	-0.9283	+	PAC34	-0.8742	-	PAC34	-0.6977	-
PAC35	-0.2732	-	PAC35	-0.9943	-	PAC35	-0.6449	-
PAC36	-0.5620	-	PAC36	-0.5136	-	PAC36	-0.4763	-
PAC41	-0.1023	-	PAC41	0.6814	+	PAC41	-0.5646	-
PAC43	-0.5893	-	PAC43	-0.9449	-	PAC43	-0.5571	+
PAC44	-0.8394	-	PAC44	-0.8556	-	PAC44	-0.4531	-
PAC45	-0.7086	-	PAC45	-0.9916	-	PAC45	-0.5480	-
PAC46	-0.9533	-	PAC46	-0.8622	-	PAC46	-0.0216	+
PAC47	-0.7803	-	PAC47	-0.9687	-	PAC47	-0.6152	-
PAC48	-0.6478	-	PAC48	-0.9756	-	PAC48	-0.4722	-
PAC49	-0.5091	-	PAC49	-0.6972	-	PAC49	-0.5983	-
PAC50	-0.7029	-	PAC50	-0.6898	-	PAC50	-0.5095	-
PAC51	0.3628	+	PAC51	-0.9988	-	PAC51	-0.6571	-
PAC53	0.4449	+	PAC53	-0.8370	-	PAC53	-0.7790	-
PAC55	-0.9097	-	PAC55	-0.9326	-	PAC55	-0.5211	-
PAC56	0.7711	+	PAC56	-0.9976	-	PAC56	-0.8257	-
PAC58	-0.4067	-	PAC58	-0.9934	+	PAC58	-0.5681	-
PAC59	-0.6270	-	PAC59	-0.0247	+	PAC59	-0.4725	-
PAC61	-0.6021	-	PAC61	-0.9458	-	PAC61	-0.5537	-

Especialista Grupo 1			Especialista Grupo 2			Especialista Grupo 3		
Paciente	Saída	Desej.	Paciente	Saída	Desej.	Paciente	Saída	Desej.
PAC62	-0.5572	+	PAC62	0.5450	+	PAC62	-0.5721	-
PAC63	-0.3644	+	PAC63	-0.9198	+	PAC63	-0.5158	-
PAC64	-0.2967	-	PAC64	-0.9914	-	PAC64	-0.6444	-
PAC66	-0.8228	+	PAC66	-0.8732	-	PAC66	-0.4705	-
PAC67	0.5150	+	PAC67	-0.9928	-	PAC67	-0.6404	-
PAC68	0.0536	-	PAC68	-0.9988	-	PAC68	-0.6026	-
PAC69	-0.7798	-	PAC69	-0.9362	-	PAC69	-0.5367	-
PAC70	-0.7182	-	PAC70	-0.6576	-	PAC70	-0.5520	-
PAC71	-0.7512	-	PAC71	-0.9766	-	PAC71	-0.4992	-
PAC73	-0.0024	+	PAC73	-0.9409	-	PAC73	-0.6756	-

Especialista Grupo 4			Especialista Grupo 5			Especialista Grupo 6		
Paciente	Saída	Desej.	Paciente	Saída	Desej.	Paciente	Saída	Desej.
PAC01	-0.9909	-	PAC01	0.5765	+	PAC01	-0.6779	-
PAC02	-0.7630	-	PAC02	-0.5495	-	PAC02	0.0942	+
PAC04	-0.6678	-	PAC04	-0.8933	-	PAC04	-0.5078	-
PAC06	0.6646	+	PAC06	-0.9834	-	PAC06	-0.9486	-
PAC09	-0.9979	-	PAC09	-0.5665	-	PAC09	-0.4316	-
PAC10	-0.7498	-	PAC10	-0.9778	-	PAC10	-0.2941	-
PAC13	-0.9849	-	PAC13	-0.8550	-	PAC13	-1	-
PAC15	-0.9643	-	PAC15	-0.9926	-	PAC15	-0.9961	-
PAC18	-0.9760	-	PAC18	-0.8379	-	PAC18	0.0831	+
PAC19	-0.9932	-	PAC19	-0.8987	-	PAC19	0.4244	+
PAC21	-0.8618	-	PAC21	0.9313	+	PAC21	-0.9996	-
PAC22	-0.9753	-	PAC22	-0.9381	-	PAC22	-0.8125	-
PAC23	0.1896	+	PAC23	-0.6679	-	PAC23	-0.4044	-
PAC24	-0.8937	-	PAC24	0.6996	+	PAC24	-0.9981	-
PAC26	-0.9972	-	PAC26	0.5891	+	PAC26	-0.3515	-
PAC27	-0.9887	-	PAC27	-0.9423	-	PAC27	-0.8091	-
PAC28	-0.9847	-	PAC28	-0.9073	-	PAC28	-0.8829	-
PAC29	-0.9964	-	PAC29	-0.9841	-	PAC29	-0.5218	-
PAC30	-0.8537	-	PAC30	-0.9343	-	PAC30	-0.9519	-
PAC31	-0.9987	-	PAC31	-0.6084	-	PAC31	-0.9731	-
PAC32	0.7913	+	PAC32	-0.8155	-	PAC32	-0.9716	-

Especialista Grupo 4			Especialista Grupo 5			Especialista Grupo 6		
Paciente	Saída	Desej.	Paciente	Saída	Desej.	Paciente	Saída	Desej.
PAC34	-0.9161	-	PAC34	0.4268	-	PAC34	-1	-
PAC35	0.6177	+	PAC35	-0.9302	-	PAC35	-0.7673	-
PAC36	0.3598	+	PAC36	-0.7631	-	PAC36	-0.6675	-
PAC41	-0.8537	-	PAC41	-0.9569	-	PAC41	-0.9168	-
PAC43	-0.8966	-	PAC43	-0.9485	-	PAC43	-0.5676	-
PAC44	-0.2929	-	PAC44	-0.8333	-	PAC44	0.3117	+
PAC45	-0.9081	-	PAC45	-0.7618	-	PAC45	0.3335	+
PAC46	-0.8493	-	PAC46	-0.9032	-	PAC46	-1	-
PAC47	-0.9760	-	PAC47	-0.6499	-	PAC47	0.6620	+
PAC48	-0.9425	-	PAC48	-0.9747	-	PAC48	-0.9271	+
PAC49	-0.7835	-	PAC49	-0.7190	-	PAC49	-0.2505	+
PAC50	-0.7847	-	PAC50	-0.6425	+	PAC50	-0.6543	-
PAC51	-0.9822	-	PAC51	-0.9723	-	PAC51	-0.9210	-
PAC53	-0.9968	-	PAC53	-0.9877	-	PAC53	-0.8788	-
PAC55	-0.9962	-	PAC55	-0.4856	-	PAC55	-0.1250	+
PAC56	-0.9831	-	PAC56	-0.9976	-	PAC56	-0.7869	-
PAC58	-0.9811	-	PAC58	-0.9692	-	PAC58	-0.0809	-
PAC59	-0.7745	-	PAC59	-0.7924	-	PAC59	-0.7264	-
PAC61	-0.9474	-	PAC61	-0.4308	-	PAC61	-0.4066	+
PAC62	-0.7692	-	PAC62	-0.7927	-	PAC62	-0.9395	-
PAC63	-0.9711	-	PAC63	-0.8443	-	PAC63	0.1170	-
PAC64	-0.9868	-	PAC64	-0.8366	-	PAC64	-0.1093	+
PAC66	-0.8482	-	PAC66	-0.4114	-	PAC66	-0.1325	-
PAC67	-0.8213	-	PAC67	-0.9785	-	PAC67	-0.5405	-
PAC68	-0.9904	-	PAC68	-0.9790	-	PAC68	0.0829	+
PAC69	-0.9406	-	PAC69	-0.3482	+	PAC69	0.0004	-
PAC70	-0.9967	-	PAC70	-0.6640	-	PAC70	0.5360	+
PAC71	-0.9663	-	PAC71	0.1519	+	PAC71	-0.0214	-
PAC73	-0.8383	-	PAC73	-0.8979	-	PAC73	-0.5921	-

8.3 APÊNDICE C

Neurônios na camada oculta=5, Especialista do Grupo 1						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	41,39	60,00	25,33	53,14	26,48	41,26792
Erro de Validação(%)	78,36	66,90	63,93	79,16	81,00	73,86982
Épocas	31500	190000	108500	79000	298000	141400

Neurônios na camada oculta=10, Especialista do Grupo 1						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	46,11	66,71	46,37	81,10	24,09	52,876
Erro de Validação(%)	72,39	75,57	78,17	84,84	66,86	75,566
Épocas	23500	4500	37000	3500	335000	80700

Neurônios na camada oculta=15, Especialista do Grupo 1						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	48,9	54,15	80	80,78	19,06	56,578
Erro de Validação(%)	64,99	78,84	82,1	78,41	58,12	72,492
Épocas	15000	20500	2000	2000	389000	85700

Neurônios na camada oculta=20, Especialista do Grupo 1						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	53	58	84	78	73	69,2
Erro de Validação(%)	72	72	81	74	22	64,2
Épocas	25000	20000	1500	500	421000	93600

Neurônios na camada oculta=25, Especialista do Grupo 1						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	60	65	85	83,9244	25	63,78488
Erro de Validação(%)	80	80	68	70,8673	74	74,57346
Épocas	3000	1000	2000	1000,00	360000	73400

Neurônios na camada oculta=30, Especialista do Grupo 1						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	65	66	90	83	11,87	63,174
Erro de Validação(%)	65	74	73,4	82	73,44	73,568
Épocas	20000	11000	1000	1000	500000	106600

Neurônios na camada oculta=5, Especialista do Grupo 2						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	43,38	50,47	57,11	37,02	41,43	45,882
Erro de Validação(%)	37,27	51,14	49,47	65,42	48,91	50,442
Épocas	500	9000	8500	500	500	3800

Neurônios na camada oculta=10, Especialista do Grupo 2						
---	--	--	--	--	--	--

Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	54,36	31,06	52,41	20,05	9,64	33,504
Erro de Validação(%)	59,02	45,88	44,29	51,05	26,11	45,27
Épocas	4000	100000	5500	156500	500000	153200

Neurônios na camada oculta=15, Especialista do Grupo 2

Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	54,41	50,48	56,96	14,61	9,72	37,236
Erro de Validação(%)	51,25	43,88	44,89	46,52	32,99	43,906
Épocas	3500	8000	7000	239500	500000	151600

Neurônios na camada oculta=20, Especialista do Grupo 2

Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	55,81	27,76	60,86	17,47	9,26	34,232
Erro de Validação(%)	57,83	46,59	56,48	49,41	29,75	48,012
Épocas	2000	99000	2000	19400	500000	124480

Neurônios na camada oculta=25, Especialista do Grupo 2

Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	55,49	16,17	58,56	11,85	26,62	33,738
Erro de Validação(%)	52,66	31,02	44,47	45,22	9	36,474
Épocas	2000	169000	2500	313500	500000	197400

Neurônios na camada oculta=30, Especialista do Grupo 2

Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	62,8	45,31	68,78	17,44	41,34	47,134
Erro de Validação(%)	57,49	27,17	58,92	48,53	51,49	48,72
Épocas	1500	500	1500	189000	500	38600

Neurônios na camada oculta=5, Especialista do Grupo 3

Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	58,58	50,36	53,58	51,46	34,04	49,604
Erro de Validação(%)	62,84	94,75	47,7	42,72	45,3	58,662
Épocas	21500	6500	500	45500	165500	47900

Neurônios na camada oculta=10, Especialista do Grupo 3

Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	58,28	53,69	52,42	58,41	57,82	56,124
Erro de Validação(%)	61,48	80,31	59,29	34,73	37,44	54,65
Épocas	10500	1500	13500	500	500	5300

Neurônios na camada oculta=15, Especialista do Grupo 3						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	57,76	34,1	35,49	50,45	59,78	47,516
Erro de Validação(%)	64,86	82,22	61,19	45,2	46,3	59,954
Épocas	2500	500	80500	46000	13000	28500

Neurônios na camada oculta=20, Especialista do Grupo 3						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	63,86	42,39	58,27	53,2	62,48	56,04
Erro de Validação(%)	72,15	82,87	60,09	45,33	48	61,688
Épocas	1500	500	5500	30000	6000	8700

Neurônios na camada oculta=25, Especialista do Grupo 3						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	58,08	42,03	75	51,73	60,1	57,388
Erro de Validação(%)	58,67	87,6	68,96	41,7	48,73	61,132
Épocas	6000	3500	1500	32000	12500	11100

Neurônios na camada oculta=30, Especialista do Grupo 3						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	70,97	49,81	63,27	53,53	59,71	59,458
Erro de Validação(%)	68,26	76,73	67,68	42,98	47,37	60,604
Épocas	500	3500	6500	22000	12000	8900

Neurônios na camada oculta=5, Especialista do Grupo 4						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	32,3	5,99	26,19	4,3842	7,13	15,19884
Erro de Validação(%)	32,58	24,31	28,81	27,78	15,54	25,804
Épocas	15000	500000	500	500000	500000	303100

Neurônios na camada oculta=10, Especialista do Grupo 4						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	4,1	6,68	34,77	35,47	7,87	17,778
Erro de Validação(%)	29,9	21,14	33,25	37,25	19,04	28,116
Épocas	500000	500000	5500	7500	500000	302600

Neurônios na camada oculta=15, Especialista do Grupo 4						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	35,11	42,38	43,1	34,45	33,94	37,796
Erro de Validação(%)	33,85	38,29	41,03	36,28	31,22	36,134
Épocas	6500	6500	800	8500	10500	6560

Neurônios na camada oculta=20, Especialista do Grupo 4						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	35,11	40,99	30,6	35,63	6,33	29,732
Erro de Validação(%)	33,85	32,48	24,72	41,49	15,51	29,61
Épocas	6500	4500	500	7000	500000	103700

Neurônios na camada oculta=25, Especialista do Grupo 4						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	4,5	34,26	37,75	35,36	35,08	29,39
Erro de Validação(%)	28,08	33,12	37,44	35,86	39,01	34,702
Épocas	500000	3500	7000	4000	5000	103900

Neurônios na camada oculta=30, Especialista do Grupo 4						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	35,03	4,98	34,76	34,55	32,34	28,332
Erro de Validação(%)	37	23,08	40,12	33,11	32,26	33,114
Épocas	5500	500000	6000	4500	6000	104400

Neurônios na camada oculta=5, Especialista do Grupo 5						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	28,78	30,57	23,2	45,56	44,41	34,504
Erro de Validação(%)	27,04	48,28	32,82	43,59	55,13	41,372
Épocas	500	6000	98000	19500	6500	26100

Neurônios na camada oculta=10, Especialista do Grupo 5						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	8,261	16,22	44,03	46,51	41,03	31,2102
Erro de Validação(%)	39,59	45,66	47,1	42,4	48,65	44,68
Épocas	334500	151000	6000	18500	8000	103600

Neurônios na camada oculta=15, Especialista do Grupo 5						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	35,75	8	21,28	50,64	42,57	31,648
Erro de Validação(%)	38,1	40,35	32,52	42,04	53,78	41,358
Épocas	5000	464000	132500	4000	4000	121900

Neurônios na camada oculta=20, Especialista do Grupo 5						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	6,58	15,38	42,92	46,48	31,42	28,556
Erro de Validação(%)	20,96	44,67	47,68	43,97	43,87	40,23
Épocas	403500	245500	3500	12000	500	133000

Neurônios na camada oculta=25, Especialista do Grupo 5						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	5,09	12,21	23,76	46,21	41,27	25,708
Erro de Validação(%)	31,03	43,25	23,8	35	55,33	37,682
Épocas	500000	222000	500	1500	3000	145400

Neurônios na camada oculta=30, Especialista do Grupo 5						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	4,81	16,99	44,34	47,6	43,07	31,362
Erro de Validação(%)	41,1	41,98	35,95	43,64	58,96	44,326
Épocas	500000	124000	2000	10000	2000	127600

Neurônios na camada oculta=5, Especialista do Grupo 6						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	69,95	53,46	6,45	29,23	44,83	40,784
Erro de Validação(%)	70,91	67,76	49,71	69,82	64,49	64,538
Épocas	3000	46500	500000	159500	109500	163700

Neurônios na camada oculta=10, Especialista do Grupo 6						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	69,95	53,46	23,23	70,6515	48,84	53,2263
Erro de Validação(%)	70,91	67,76	51,41	78,2	62,97	66,25
Épocas	3000	46500	194000	3000	80000	65300

Neurônios na camada oculta=15, Especialista do Grupo 6						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	69,72	24,26	21,7	32,03	45,86	38,714
Erro de Validação(%)	78,09	64,92	51,87	70,96	66,19	66,406
Épocas	500	157500	211000	118500	79000	113300

Neurônios na camada oculta=20, Especialista do Grupo 6						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	72,5	64,83	17,17	37,11	46,59	47,64
Erro de Validação(%)	73,84	25,9	53,69	78,22	64,14	59,158
Épocas	2000	500	260500	94500	80500	87600

Neurônios na camada oculta=25, Especialista do Grupo 6						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	16,08	54,17	17,63	73,5	47,2	41,716
Erro de Validação(%)	70,18	66,95	53,04	78,55	64,32	66,608
Épocas	313500	41500	228500	1500	77500	132500

Neurônios na camada oculta=30, Especialista do Grupo 6						
Configuração	A	B	C	D	E	Média
Erro de Treinamento(%)	73,32	78,53	6,68	73,57	44,72	55,364
Erro de Validação(%)	75,32	68,84	53,15	75,49	65,13	67,586
Épocas	1500	2500	500000	1500	108500	122800

8.4 APÊNDICE D

O limiar de classificação adotado para considerar o indivíduo como pertencente ao grupo de patologias foi de “0”, ou seja, se a saída for ‘+’ o indivíduo pertence ao grupo de patologias, se for ‘-’ não pertence.

Paciente:	PAC 01	PAC 02	PAC 04	PAC 06	PAC 09	PAC 10	PAC 13	PAC 15
------------------	--------	--------	--------	--------	--------	--------	--------	--------

Especialista	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’
1	-0,8232	-0,7305	-0,5096	-0,6556	-0,1879	-0,6895	-0,9703	0,9834
2	-0,9747	-0,8747	-0,9488	-0,834	-0,9922	-0,5056	-0,9642	0,9471
3	-0,6148	-0,7332	-0,9305	-0,6221	-0,66	-0,7339	0,7764	-0,7471
4	-0,9909	-0,763	-0,6678	0,6646	-0,9979	-0,7498	-0,9849	-0,9643
5	0,5765	-0,5495	-0,8933	-0,9834	-0,5665	-0,9778	-0,855	-0,9926
6	-0,6779	0,0942	-0,5078	-0,9486	-0,4316	-0,2941	-1	-0,9961

Paciente:	PAC 18	PAC 19	PAC 21	PAC 22	PAC 23	PAC 24	PAC 26	PAC 27
------------------	--------	--------	--------	--------	--------	--------	--------	--------

Especialista	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’
1	-0,6014	-0,6896	-0,4972	-0,9102	-0,6796	-0,7403	-0,7638	-0,7677
2	-0,8336	-0,8241	-0,8518	0,5961	-0,7771	-0,9815	-0,7465	-0,9078
3	-0,6691	-0,4401	-0,4252	-0,6526	-0,7207	-0,1122	-0,7456	0,0552
4	-0,976	-0,9932	-0,8618	-0,9753	0,1896	-0,8937	-0,9972	-0,9887
5	-0,8379	-0,8987	0,9313	-0,9381	-0,6679	0,6996	0,5891	-0,9423
6	0,0831	0,4244	-0,9996	-0,8125	-0,4044	-0,9981	-0,3515	-0,8091

Paciente:	PAC 28	PAC 29	PAC 30	PAC 31	PAC 32	PAC 34	PAC 35	PAC 36
------------------	--------	--------	--------	--------	--------	--------	--------	--------

Especialista	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’
1	-0,4534	-0,4338	0,1869	-0,9544	0,4169	-0,9283	-0,2732	-0,562
2	-0,9659	-0,991	-0,8914	-0,9545	-0,8338	-0,8742	-0,9943	-0,9136
3	-0,6194	-0,7596	-0,6299	-0,0415	-0,6685	0,5814	-0,7071	-0,6547
4	-0,9847	-0,9964	-0,8537	-0,9987	0,7913	-0,9161	0,6177	0,3598
5	-0,9073	-0,9841	-0,9343	-0,6084	-0,8155	0,4268	-0,9302	-0,7631
6	-0,8829	-0,5218	-0,9519	-0,9731	-0,9716	-1	-0,7673	-0,6675

Paciente:	PAC 41	PAC 43	PAC 44	PAC 45	PAC 46	PAC 47	PAC 48	PAC 49
------------------	--------	--------	--------	--------	--------	--------	--------	--------

Especialista	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’	‘y _i ’
1	-0,1023	-0,5893	-0,8394	-0,7086	-0,9533	-0,7803	-0,6478	-0,5091
2	0,6814	-0,9449	-0,8556	-0,9916	-0,8622	-0,9687	-0,9756	-0,6972
3	-0,6511	-0,6414	-0,7172	-0,7115	0,5974	-0,7171	-0,4761	-0,7402
4	-0,8537	-0,8966	-0,2929	-0,9081	-0,8493	-0,976	-0,9425	-0,7835
5	-0,9569	-0,9485	-0,8333	-0,9618	-0,9032	-0,6499	-0,9747	-0,719
6	-0,9168	-0,5676	0,3117	0,3335	-1	0,662	-0,9271	-0,2505

Paciente:	PAC 50	PAC 51	PAC 53	PAC 55	PAC 56	PAC 58	PAC 59	PAC 61
------------------	--------	--------	--------	--------	--------	--------	--------	--------

Especialista	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '
1	-0,7029	0,3628	0,4449	-0,9097	0,7711	-0,4067	-0,627	-0,6021
2	-0,6898	-0,9988	-0,837	-0,9326	-0,9976	-0,9934	-0,0247	-0,9458
3	-0,7167	-0,6311	-0,7375	-0,7849	-0,7894	-0,6809	-0,7025	-0,7456
4	-0,7847	-0,9822	-0,9968	-0,9962	-0,9831	-0,9811	-0,7745	-0,9474
5	-0,6425	-0,9723	-0,9877	-0,4856	-0,9976	-0,9692	-0,7924	-0,4308
6	-0,6543	-0,921	-0,8788	-0,125	-0,7869	-0,0809	-0,7264	-0,4066

Paciente:	PAC 62	PAC 63	PAC 64	PAC 66	PAC 67	PAC 68	PAC 69	PAC 70
------------------	--------	--------	--------	--------	--------	--------	--------	--------

Especialista	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '	'y _i '
1	-0,5572	-0,3644	-0,2967	-0,8228	0,515	0,0536	-0,7798	-0,7182
2	0,545	-0,9198	-0,9914	-0,8732	-0,9928	-0,9988	-0,9362	-0,6576
3	-0,6575	-0,679	-0,7267	-0,7368	-0,6505	-0,63	-0,7075	-0,7514
4	-0,7692	-0,9711	-0,9868	-0,8482	-0,8213	-0,9904	-0,9406	-0,9967
5	-0,7927	-0,8443	-0,8366	-0,4114	-0,9785	-0,979	-0,3482	-0,664
6	-0,9395	0,117	-0,1093	-0,1325	-0,5405	0,0829	0,0004	0,536

Paciente:	PAC 71	PAC 73
------------------	--------	--------

Especialista	'y _i '	'y _i '
1	-0,7512	-0,0024
2	-0,9766	-0,9409
3	-0,6674	-0,7005
4	-0,9663	-0,8383
5	0,1519	-0,8979
6	-0,0214	-0,5921

8.5 APÊNDICE E

Erros de validação, Grupo 1

Paciente PAC)	Erro de validação G1 (%)
1	34.4116
2	15.8525
4	135.6063
6	6.8625
9	123.2759
10	38.2375
13	0.0010
15	0
18	60.1362
19	32.3708
21	57.5836
22	22.4655
23	21.9263
24	6.3547
26	25.0795
27	0.2703
28	39.6365
29	44.5554
30	104.7532
31	0.0707
32	23.5432
35	52.4166
36	51.9543
41	67.7300
43	0.8526
44	12.2225
45	41.4546
46	44.3569
47	16.1061
48	48.0496
49	39.4607
50	7.1052
51	104.0190
53	99.7029
55	0.9120
56	120.2538
58	51.2266
59	11.1211
61	42.0655
62	40.1098
63	159.3881
64	22.0577
66	17.0004
67	62.0197
68	36.9073
69	2.0694
70	39.2688
71	14.8819
73	141.4024
MÉDIA →	43.6553

Erros de validação, Grupo 2

Paciente (PAC)	Erro de validaçãoG2(%)
1	0.0099
2	10.9058
4	16.1882
6	15.4936
9	0.0528
10	19.7646
13	136.8140
15	134.7116
18	35.7392
19	25.3680
21	48.5469
22	164.1774
23	24.7926
24	0.2746
26	28.1358
27	26.7232
28	12.5399
29	2.0844
30	4.6991
31	21.4212
32	7.4855
35	0.0430
36	33.0586
41	133.6052
43	3.7059
44	3.0969
45	1.1957
46	111.2630
47	1.0031
48	0.0197
49	31.0456
50	21.0273
51	11.7174
53	27.1357
55	0.4181
56	12.8870
58	147.0426
59	115.4363
61	0.1407
62	100.6990
63	164.7310
64	0.2888
66	2.5119
67	0.3407
68	27.8249
69	29.2405
70	41.3819
71	0.2967
73	4.1665
MÉDIA →	35.94

Erros de validação, Grupo 3.

Paciente (PAC)	Erro de validação(%)
1	22.4110
2	8.5393
4	28.6347
6	23.6170
9	25.5063
10	162.0579
13	14.1157
15	20.5998
18	37.1087
19	57.6169
21	61.6749
22	34.6256
23	14.4850
24	42.5332
26	147.3132
27	161.6466
28	158.5565
29	180.6726
30	135.4778
31	156.6256
32	0.0419
35	1.8379
36	0.0850
41	8.8946
43	150.4825
44	9.9870
45	2.2309
46	184.8449
47	42.4354
48	31.7316
49	11.8243
50	24.2619
51	38.4018
53	70.8094
55	7.2026
56	11.1860
58	22.7995
59	33.4849
61	3.7754
62	34.6965
63	32.7346
64	33.0133
66	23.6254
67	36.8140
68	21.3861
69	23.9888
70	0.7905
71	23.6554
73	0.2349
MÉDIA →	48.59

Erros de validação, Grupo 4.

Paciente (PAC)	Erro de validação(%)
1	13.0117
2	33.4401
4	20.2023
6	126.5737
9	0.1812
10	26.1846
13	1.5380
15	40.9639
18	0.5992
19	9.6866
21	7.6405
22	26.0461
23	180.6916
24	30.5296
26	19.4216
27	0.2372
28	0.5389
29	25.2449
30	6.6902
31	0.0263
32	173.1283
35	158.8427
36	104.5296
41	42.9018
43	5.4346
44	27.9671
45	19.2103
46	12.6655
47	0.7683
48	9.6542
49	14.5683
50	23.4465
51	0.2111
53	0.0297
55	21.1662
56	19.2614
58	0.1080
59	12.7135
61	2.2947
62	22.9776
63	1.2345
64	11.0644
66	4.3756
67	26.2606
68	0.0651
69	26.7407
70	0.0692
71	21.7596
73	24.0105
MÉDIA →	27.6920

Erros de validação, Grupo 5.

Paciente (PAC)	Erro de validação(%)
1	150.2089
2	37.8669
4	1.8418
6	0
9	23.2054
10	0.0002
13	31.4189
15	0.6513
18	0.4380
19	0.0001
21	0.1417
22	0.0006
23	30.9124
24	150.3834
26	156.4121
27	0.0530
28	7.3026
29	0.0035
30	0.1274
31	1.9347
32	29.5960
35	27.3341
36	35.4158
41	0.3049
43	0.6229
44	1.8474
45	4.7613
46	85.7247
47	23.9720
48	29.0903
49	34.0973
50	168.1660
51	0.0994
53	8.4875
55	19.6631
56	0
58	0.7537
59	0.8319
61	37.5812
62	22.8258
63	1.0526
64	3.5547
66	28.9866
67	0.2976
68	1.7614
69	169.3140
70	25.3286
71	175.2670
73	35.2087
MÉDIA →	31.9357

Erros de validação, Grupo 6.

Paciente (PAC)	Erro de validação(%)
1	39.4803
2	95.2001
4	22.6322
6	37.4277
9	7.5684
10	51.1785
13	0
15	0
18	139.8510
19	132.1729
21	0
22	0.0004
23	61.4251
24	0
26	64.9468
27	0.2871
28	0.0616
29	47.6084
30	0.0001
31	0.0001
32	21.4680
35	4.4168
36	52.6983
41	0.0010
43	52.1208
44	27.7329
45	109.3138
46	0
47	0.3029
48	48.9575
49	143.6316
50	5.7856
51	0.0001
53	38.3737
55	147.6254
56	38.5455
58	56.5097
59	59.5258
61	19.8326
62	0.0006
63	37.6641
64	135.5257
66	48.4098
67	50.6933
68	94.8212
69	53.3311
70	84.2678
71	61.2730
73	30.3622
MÉDIA →	43.3272