

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ESCOLA DE ENGENHARIA

Curso de Pós-Graduação em Engenharia Elétrica - CPGEE

**APLICAÇÃO DA TRANSFORMADA WAVELET PACKET
NA ANÁLISE E CLASSIFICAÇÃO DE SINAIS DE VOZES
PATOLÓGICAS**

ADRIANE PARRAGA

Dissertação para obtenção do título de Mestre em Engenharia Elétrica

Porto Alegre

2002

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ESCOLA DE ENGENHARIA

Curso de Pós-Graduação em Engenharia Elétrica - CPGEE

**APLICAÇÃO DA TRANSFORMADA WAVELET PACKET
NA ANÁLISE E CLASSIFICAÇÃO DE SINAIS DE VOZES
PATOLÓGICAS**

ADRIANE PARRAGA

Engenheiro Eletricista

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Elétrica - CPGEE, como parte dos requisitos para a obtenção do título de Mestre em Engenharia. Área de concentração: Automação e Instrumentação Eletro-Eletrônica. Desenvolvida no Laboratório de Processamento de Sinais e Imagem do Departamento de Engenharia Elétrica da Universidade Federal do Rio Grande do Sul.

Porto Alegre

2002

APLICAÇÃO DA TRANSFORMADA WAVELET PACKET NA ANÁLISE E CLASSIFICAÇÃO DE SINAIS DE VOZES PATOLÓGICAS

ADRIANE PARRAGA

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

Orientador: _____

Prof. Adalberto Schuck Jr., UFRGS

Dr. 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. John Oersted Wisbeck, UFRGS

Dr. pela Universidade Federal de Santa Catarina – UFSC – 2001

Prof. Altamiro Amadeu Susin, UFRGS

Dr. pelo Institut National Polytechnique de Grenoble – INPG – 1981

Coordenador do CPGEE: _____

Prof. Dr. Altamiro Amadeu Susin

Porto Alegre, março de 2002.

ÍNDICE

ÍNDICE	iv
ÍNDICE DE FIGURAS	vi
ÍNDICE DE TABELAS	x
TABELA DE ABREVIÇÕES	xii
RESUMO	xiv
ABSTRACT	xv
1 INTRODUÇÃO	1
2 REVISÃO DA LITERATURA E ESTADO DA ARTE	4
2.1 <i>Introdução</i>	4
2.2 <i>Classificação de Sinais de Vozes Patológica</i>	4
2.3 <i>Aplicações da Transformada Wavelet</i>	11
2.4 <i>Considerações Finais</i>	24
3 FUNDAMENTOS TEÓRICOS	25
3.1 <i>Introdução</i>	25
3.2 <i>Fisiologia da Voz e Patologias</i>	25
3.3 <i>Transformada Wavelet Packet</i>	44
4 MATERIAIS E MÉTODOS	70
4.1 <i>Introdução</i>	70
4.2 <i>Banco de Vozes Patológicas</i>	70
4.3 <i>Teste da Ferramenta Wavelet Packet</i>	72
4.4 <i>Aplicação da TWP nas Vozes Patológicas</i>	78
5 RESULTADOS	80

5.1	<i>Introdução</i>	80
5.2	<i>Resultados da Best Basis dos Sinais Sintéticos</i>	80
5.3	<i>Resultados da Best Basis no Banco de Vozes</i>	83
6	DISCUSSÕES E CONCLUSÕES	96
6.1	<i>Introdução</i>	96
6.2	<i>Análise dos Resultados e Conclusões</i>	96
6.3	<i>Sugestões para Trabalhos Futuros</i>	98
	REFERÊNCIAS	100
	APÊNDICES	106
	APÊNDICE A	106
	APÊNDICE B	110

ÍNDICE DE FIGURAS

3-1 Diagrama esquemático que mostra as partes do sistema respiratório e a laringe.....	26
3-2 Partes que compõem a Laringe	27
3-3 Representação das partes do sistema articulatório.....	29
3-4 a) Representação esquemática do perfil do sistema vocal de cada vogal b) formas de ondas típicas no tempo c) espectro correspondente para cada vogal	30
3-5 Espectro do som da laringe.....	31
3-6 Representação do modelo fonte-filtro para vogais	32
3-7 Decomposição Wavelet Discreta multiresolução.....	51
3-8 Processo de decimação	52
3-9 Decomposição em sub-bandas Wavelet Discreta.....	53
3-10 Representação dos espaços vetoriais para m variando de 0 a 2.....	54
3-11 Diagrama de Blocos da Transformada Wavelet Discreta.....	58
3-12 Diagrama de Blocos da Transformada Wavelet Packet.....	59
3-13 Em cada nível de dilatação σ^s tem-se as bases correspondentes aos subespaços Λ_f	61
3-14 Árvore Wavelet Packet a) análise b) Síntese	61
3-15 Bases Wavelets Packets Haar-Walsh até o 3º nível (σ^3) de decomposição.....	62
3-16 Coeficientes Wavelet Packet na ordem natural ou Paley.....	63
3-17 Coeficientes Wavelet Packet em ordem crescente de freqüência nominal.....	64
3-18 Transformada Wavelet Discreta em cinza.....	64
3-19 Base Wavelet Packet para escolha de um nível fixo (σ^2) em cinza.....	65
3-20 Outra possível base Wavelet Packet em cinza.....	65
3-21 Um estágio de decomposição do sinal.....	67
3-22 Diagrama da árvore de completa obtida pela decomposição.....	68
3-23 Diagrama de um nó pai e seus nós filhos para o cálculo da função custo.....	68
3-24 Ilustração do princípio de seleção <i>Best Basis</i>	69
4-1 Modelo fonte-filtro da geração da vogal sintética.....	73
4-2 Vogal sintética /a/ para a) Homens e $f_0=120\text{Hz}$ e b) Mulheres $f_0=220\text{Hz}$	74

4-3 Método de inserção de <i>Jitter</i> no trem de impulsos.....	76
4-4 Método de inserção de <i>Shimmer</i> no trem de impulsos.....	76
4-5 Árvore binária Wavelet Packet com o 5° nível de decomposição.....	77
A.1 <i>Best Basis</i> do Homem com a função <i>coif2</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	110
A.2 <i>Best Basis</i> do Homem com a função <i>coif5</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	111
A.3 <i>Best Basis</i> do Homem com a função <i>db10</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	112
A.4 <i>Best Basis</i> do Homem com a função <i>db6</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	113
A.5 <i>Best Basis</i> do Homem com a função <i>haar</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	114
A.6 <i>Best Basis</i> do Homem com a função <i>sym5</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	115
A.7 <i>Best Basis</i> do Homem com a função <i>sym8</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	116
A.8 <i>Best Basis</i> da Mulher com a função <i>coif2</i> para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$	117
A.9 <i>Best Basis</i> da Mulher com a função <i>coif5</i> para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$	118
A.10 <i>Best Basis</i> da Mulher com a função <i>db10</i> para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$	119
A.11 <i>Best Basis</i> da Mulher com a função <i>db6</i> para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$	120
A.12 <i>Best Basis</i> da Mulher com a função <i>haar</i> para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$	121
A.13 <i>Best Basis</i> da Mulher com a função <i>sym5</i> para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$	122
A.14 <i>Best Basis</i> da Mulher com a função <i>sym8</i> para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$	123
A.15 <i>Best Basis</i> Casal com a função <i>coif2</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	124

A.16 <i>Best Basis</i> Casal com a função <i>coif2</i> para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$	125
A.17 <i>Best Basis</i> Casal com a função <i>coif5</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	126
A.18 <i>Best Basis</i> Casal com a função <i>coif5</i> para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$	127
A.19 <i>Best Basis</i> Casal com a função <i>db10</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	128
A.20 <i>Best Basis</i> Casal com a função <i>db10</i> para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$	129
A.21 <i>Best Basis</i> Casal com a função <i>db6</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	130
A.22 <i>Best Basis</i> Casal com a função <i>db6</i> para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$	131
A.23 <i>Best Basis</i> Casal com a função <i>haar</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	132
A.24 <i>Best Basis</i> Casal com a função <i>haar</i> para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$	133
A.25 <i>Best Basis</i> Casal com a função <i>sym5</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	134
A.26 <i>Best Basis</i> Casal com a função <i>sym5</i> para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$	135
A.27 <i>Best Basis</i> Casal com a função <i>sym8</i> para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$	136
A.28 <i>Best Basis</i> Casal com a função <i>sym8</i> para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$	137
A.29 <i>Best Basis</i> LPC patológicos com a função <i>coif2</i> para a) normais b) laringite c) leucoplasia d) disfonia.....	138
A.30 <i>Best Basis</i> LPC patológicos com a função <i>coif5</i> para a) normais b) laringite c) leucoplasia d) disfonia.....	139
A.31 <i>Best Basis</i> LPC patológicos com a função <i>db10</i> para a) normais b) laringite c) leucoplasia d) disfonia.....	140
A.32 <i>Best Basis</i> LPC patológicos com a função <i>db6</i> para a) normais b) laringite c) leucoplasia d) disfonia.....	141

A.33 <i>Best Basis</i> LPC patológicos com a função haar para a) normais b) laringite c) leucoplasia d) disfonia.....	142
A.34 <i>Best Basis</i> LPC patológicos com a função sym5 para a) normais b) laringite c) leucoplasia d) disfonia.....	143
A.35 <i>Best Basis</i> LPC patológicos com a função sym8 para a) normais b) laringite c) leucoplasia d) disfonia.....	144
A.36 <i>Best Basis</i> com função sym5 para LPC Homem com <i>Jitter</i> de a) 0,5% b) 1% c) 2% d) 5% e) 10%.....	145
A.37 <i>Best Basis</i> com função sym5 para LPC Homem com <i>Shimmer</i> de a) 5% b) 10% c) 15% d) 20% e) 25%.....	146
A.38 <i>Best Basis</i> com função sym5 para LPC Mulher com <i>Jitter</i> de a) 0,5% b) 1% c) 2% d) 5% e) 10%.....	147
A.39 <i>Best Basis</i> com função sym5 para LPC Homem com <i>Shimmer</i> de a) 5% b) 10% c) 15% d) 20% e) 25%.....	148

ÍNDICE DE TABELAS

3-1 Principais Sintomas.....	34
3-2 Sinais Perceptuais de distúrbios da voz.....	35
3-3 Principais sinais acústicos.....	37
3-4 Principais Sinais Fisiológicos.....	37
4-1 Lista de frequências fundamentais para sinais sintéticos de homem e mulher.....	75
4-2 Lista de frequências para sinais sintéticos para casal.....	75
4-3 Grupos de patologias.....	78
4-4 Grupos de patologias para o banco de vozes descrito em Apêndice A.....	79
5-1 Número de tipos diferentes de <i>Best Basis</i> ou de árvores com a função custo Entropia	81
5-2 Número de tipos diferentes de <i>Best Basis</i> ou de árvores com a função custo LogEnergia.....	82
5-3 Número de tipos diferentes de Best Basis ou de árvores com a função custo Norma 2.....	82
5-4 Diferentes tipos de árvores de cada função Wavelet Packet para os sinais sintéticos gerados a partir do LPC de quatro patologias.....	83
5-5 Diferentes tipos de árvores da função sym5 para os sinais sintéticos gerados a partir do LPC de homens e mulheres com variações de <i>Jitter</i> e <i>Shimmer</i>	83
5-6 Presença dos nós das <i>Best Basis</i> do Banco de vozes para sinais Normais com a função Sym5	84
5-7 Presença dos nós das <i>Best Basis</i> do Banco de vozes para sinais Patológicos com a função Sym5	85
5-8 Continuação Tabela 5-7.....	86
5-9 Continuação Tabela 5-8.....	87
5-10 Continuação Tabela 5-9.....	88
5-11 Classificação de Normais versus Patológicos para a função sym5 e entropia.....	88
5-12 Presença dos nós das <i>Best Basis</i> do Banco de vozes para sinais Normais com a função Coif2	89
5-13 Presença dos nós das <i>Best Basis</i> do Banco de vozes para sinais patológicos com a função Coif2	90

5-14 Continuação Tabela 5-13.....	91
5-15 Continuação Tabela 5-14.....	92
5-16 Continuação Tabela 5-15.....	93
5-17 Classificação normais <i>versus</i> patológicos para a função <i>coif2</i> e entropia.....	93
5-18 Classificação normais <i>versus</i> patológicos para a função <i>sym5</i> e Norma 2.....	94
5-19 Número de ocorrências dos pares de nós do último nível da árvore Wavelet Packet para os grupos de patologias.....	95

TABELA DE ABREVIACOES

	Significado em Ingls	Sugesto em Portugus
AC	Acoustic Emission	Emisso Acstica
AMDF	Average Magnitude Difference Function	
ANN	Artificial Neural Network	Redes Neurais Artificiais
AR	Autoregressive	Autoregressivo
ARS	Automatic Response System	Sistema de Resposta Automtica
AS	Autocorrelation Sequence	Seqncia de Autocorrelao
CART	Classification And Regression Trees	
CEP	Cepstrum	Cepstrum
CVQPM	Cross-Validated Based On Posterior Probabilities	
CWT	Continuous Wavelet Transform	Transformada Wavelet Contnua
DAT	Digital Audio Tape	
DP	Discriminant Pursuit	
DyWT	Dyadic Wavelet Transform	Transformada Wavelet Didica
DWT	Discrete Wavelet Transform	Transformada Wavelet Discreta
ECG	Eletrocardigraphy	
EEG	Electroencephalography	Eletroencefalografia
EGG	Electroglottography	Eletroglotografia
FCM	Fuzzy Clustering Method	Mtodo de Clusterizao Fuzzy
FI	Inverse Filter	Filtro Inverso
FWP	Fast Wavelet Packet	
HMM	Hidden Markov Model	Modelo Oculto de Markov

k-NN	k-Nearest Neighbour	
LDA	Linear Discriminant Analysis	Análise Discriminante Linear
LDB	Local Discriminant Basis	
LPC	Linear Predictive Coding	
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 Motora
NNE	Normalized Noise Energy	Energia Normalizada de Ruído
PCNN	Parallel Consensual Neural Network	
PDM	Prototype Distribution Map	Protótipos de Mapas de Distribuição
QMF	Quadrature Mirror Filters	Filtros de Quadratura de Espelho
QF	Quadrature Filters	Filtros de Quadratura
RMS	Root Mean Square	Raiz Média Quadrada
SAR	Synthetic Aperture Radar	
SBC	Sub-Banda Base Cepstral	
SIFT	Simple Inverse Filter Tracking	
SNR	Signal Noise Ratio	Relação Sinal/Ruído
SOM	Self-Organizing Map	
STFR	Short Time Fourier Transform	
TWD	Discrete Wavelet Transform	Transformada Wavelet Discreta
TWP	Wavelet Packet Transform	Transformada Wavelet Packet
TF	Fourier Transform	Transformada de Fourier
FIR	Finite Impulse Response	Filtro de Resposta Finita ao Impulso

RESUMO

O exame para o diagnóstico de doenças da laringe é usualmente realizado através da videolaringoscopia e videoestroboscopia. A maioria das doenças na laringe provoca mudanças na voz do paciente. Diversos índices têm sido propostos para avaliar quantitativamente a qualidade da voz. Também foram propostos vários métodos para classificação automática de patologias da laringe utilizando apenas a voz do paciente. Este trabalho apresenta a aplicação da Transformada Wavelet Packet e do algoritmo *Best Basis* [COI92] para a classificação automática de vozes em patológicas ou normais.

Os resultados obtidos mostraram que é possível classificar a voz utilizando esta Transformada. Tem-se como principal conclusão que um classificador linear pode ser obtido ao se empregar a Transformada Wavelet Packet como extrator de características. O classificador é linear baseado na existência ou não de nós na decomposição da Transformada Wavelet Packet. A função Wavelet que apresentou os melhores resultados foi a symlet 5 e a melhor função custo foi a entropia. Este classificador linear separa vozes normais de vozes patológicas com um erro de classificação de 23,07% para falsos positivos e de 14,58% para falsos negativos.

ABSTRACT

The diagnosis of laryngeal diseases is usually realized by videolaryngoscopy and videostroboscopy. Most laryngeal diseases cause changes in the patient's voice. Several indices have been proposed to assess the quality of the voice in a quantitative manner. Several methods were proposed for automatic classification of laryngeal pathologies using only the patient's voice. This work presents the application of the Wavelet Packet Transform and the Best Basis algorithm [COI92] for automatic classification of voices into pathological or normal.

The obtained results have shown that it is possible to classify voices using this Transform. The main conclusion is that a linear classifier can be obtained using the Wavelet Packet Transform for feature extraction. The classifier is linear based on the presence or not of the nodes on the decomposition of the Wavelet Packet Transform. The best results have been obtained with the Wavelet basis symlet 5 and the entropy as the cost function. This linear classifier distinguished normal voices from pathological ones with a classification error of the 23,07% for false positive and the 14,58% for false negative.

1 INTRODUÇÃO

A maioria das doenças na laringe provoca mudanças na voz do paciente. O primeiro teste tradicionalmente feito pelos otorrinolaringologistas é a audição da voz do paciente, seguido pela videolaringoscopia direta e videoestroboscopia. A videolaringoscopia direta é um exame realizado pelo médico com o objetivo de visualizar a laringe utilizando uma microcâmera. A videoestroboscopia permite a visualização do comportamento vibratório das pregas vocais [COL96]. Estes são procedimentos desconfortáveis e como tal, durante o exame podem causar dor, náusea e ansiedade, entre outros inconvenientes, aos pacientes. Essas técnicas visuais resultam em uma avaliação qualitativa, de resultados difíceis de serem quantificados, e necessitam do conhecimento e da experiência do avaliador.

Diversas técnicas alternativas de auxílio ao diagnóstico têm sido propostas utilizando somente a voz do paciente. Foram propostos diversos índices quantitativos para diagnosticar e acompanhar a evolução de tratamentos de doenças da laringe [BAK87]. Vários métodos de diagnóstico automático utilizando a voz do paciente têm sido pesquisados, tais como filtragem inversa [CHI92], redes neurais [MAR00], [RIT01], [JO01], medidas acústicas da voz [FRO98], entre outras. Estas técnicas, em relação aos métodos tradicionais, apresentam como vantagens:

1. Propiciar um exame menos desconfortável ao Paciente;
2. Fornecer uma avaliação quantitativa, padronizar os critérios de avaliação da qualidade da voz (aspereza, rugosidade, sopro, etc...);
3. Possibilitar o desenvolvimento de sistemas automáticos de auxílio ao diagnóstico por computador com baixo custo.

Muitos sinais na natureza possuem características não estacionárias, tais como os sinais de voz [AKA98]. A análise do sinal utilizando-se a Transformada de Fourier (TF) é mais adequada para análise de sinais estocásticos estacionários. Neste tipo de análise a informação de tempo é perdida e apenas a informação de frequência está presente. Gabor adaptou a Transformada de Fourier para analisar apenas uma pequena seção do sinal no tempo, uma técnica chamada de *windowing* (janelamento) 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 [RIO91]. A STFT possui

um compromisso entre tempo e frequência do 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. Por outro lado, esta janela resulta em um número reduzido de amostras usadas no cálculo da Transformada de Fourier, na qual fornece também um número reduzido de frequências discretas que podem ser representadas no domínio frequência. Uma vez determinada a resolução temporal 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 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 de um determinado sinal [RIO91]. A Transformada Wavelet surgiu há poucos anos como uma alternativa a esta Transformada para análise tempo-frequência. Uma maneira eficiente de aplicar a Transformada Wavelet Discreta é através de filtros, técnica desenvolvida por Mallat (1989), onde se tem a decomposição da Wavelet implementando a análise Multiresolução. A Transformada Wavelet Packet, proposta por [COI92] é uma generalização do conceito da Transformada Wavelet Discreta.

A Transformada Wavelet Packet tem sido utilizada nos últimos anos em processamento de sinais permitindo melhorar a classificação de sinais. Esta técnica é utilizada como uma ferramenta para extrair características relevantes de sinais de uma ou duas dimensões, melhorando a performance dos classificadores ao extrair características relativas através da decomposição dos sinais em diferentes bandas de frequências. Aplicações desta Transformada em sinais de naturezas diferentes como sinais de vibrações de máquinas [PAY96], [YEN00], sinais de eletrocardiogramas [AKA98], texturas de imagens [BAL99], [SHI00], entre outras aplicações, têm mostrado a abrangência de sua utilização como pré-processador para classificadores.

Neste trabalho é proposta a aplicação da Transformada Wavelet Packet e do algoritmo *Best Basis* [COI92] para a classificação automática de patologias da voz. Este algoritmo foi originalmente proposto para compressão de dados, um problema matematicamente análogo à extração de características [WU96], o que justifica seu uso

nesta aplicação. Este método utiliza a voz do paciente, apresentando as mesmas vantagens (1,2 e 3) citadas anteriormente em relação aos métodos tradicionais . Sendo assim, o objetivo desta dissertação é estudar, testar e analisar a aplicabilidade da Transformada Wavelet Packet como um classificador de vozes patológicas ou normais. Para isso, o trabalho é organizado como segue.

No capítulo 2 estão relatadas diversas aplicações da Transformada Wavelet como extrator de características para classificadores de sinais de uma ou de duas dimensões. Neste capítulo também são apresentadas diferentes técnicas propostas para a classificação e avaliação de patologias da laringe.

No capítulo 3 encontram-se os fundamentos teóricos deste trabalho. Na seção 3.2 são apresentadas a fisiologia da voz e doenças que ocorrem na laringe; o modelo fonte-filtro para a produção de vogais também é apresentado. Na seção 3.3 encontra-se a teoria relativa à Transformada Wavelet. Para fins didáticos, são discutidas as Transformadas Wavelet Contínua, Discreta e Packet, concluindo com o algoritmo *Best Basis*.

No capítulo 4 encontram-se descritos os materiais e métodos que foram utilizados neste trabalho, incluindo a aquisição dos sinais de vozes, a digitalização e o banco de dados obtido dos pacientes; os programas para as análises com a Wavelet Packet, a geração dos sinais sintéticos, as bases e as funções custo utilizadas.

No capítulo 5 são mostrados os resultados obtidos aplicando a Transformada Wavelet Packet a sinais sintéticos e sinais do banco de vozes previamente diagnosticados.

Finalmente, é apresentada no capítulo 6 uma discussão dos resultados obtidos. Tem-se como principal conclusão que o classificador linear pode ser utilizado ao se empregar a Transformada Wavelet Packet com a função symlet 5 e a entropia como função custo. Este classificador separa voz normal de patológica com 23,07% de falsos positivos e 14,58% de falsos negativos. São também apresentadas sugestões para trabalhos futuros e possíveis caminhos para a continuidade desta pesquisa.

2 REVISÃO DA LITERATURA E ESTADO DA ARTE

2.1 INTRODUÇÃO

A Revisão de Literatura está dividida em duas partes. Na primeira parte, seção 2.2, são relatados trabalhos que têm o objetivo de análise e classificação de vozes com patologias na laringe, utilizando diferentes técnicas e ferramentas, até os dias atuais.

A segunda parte, seção 2.3, apresenta diversas aplicações da Transformada Wavelet (Contínua, Discreta e a Packet) com o objetivo de classificação de sinais de uma e duas dimensões. Também são relatadas algumas aplicações da Transformada Wavelet em sinais de vozes patológicas. Estes trabalhos apresentados mostram a abrangência desta Transformada em diferentes áreas, com sinais de naturezas diferentes, descrevendo as técnicas e os resultados obtidos.

Em ambas as seções, a revisão de literatura será apresentada em ordem cronológica. Na seção 2.4 são feitas considerações a respeito da revisão bibliográfica apresentada neste capítulo, contextualizando o presente trabalho nesta área de pesquisa.

Até o presente momento, não se encontrou na literatura a aplicação da Transformada Wavelet Packet para a classificação de patologias das cordas vocais, o que indica o caráter original deste trabalho.

2.2 CLASSIFICAÇÃO DE SINAIS DE VOZES PATOLÓGICA

No artigo [RON75], 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 [DEL80] apresentaram uma técnica automática de classificação de disfunção da laringe, utilizando os pólos e zeros de um filtro inverso (FI) digital. A informação para a classificação está contida na posição dos zeros do polinômio do filtro inverso. 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:

1. Casos de performance de filtro (como *Jitter*, *Shimmer*, entre outros);
2. 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 [GUO89], 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 de acordo. Os resultados deste artigo mostraram que este método é capaz de reconhecer características acústicas de vozes anormais. 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.

Em [CHI92] 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:

1. Medida de distorção espectral usando métodos de *Pitch* com *Linear Predicting Code* (LPC) e quantização vetorial;
2. Análise do sinal de eletroglotografia (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 [ROS98] 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 em [ACC98], 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 Friedreich. Ataxia de Friedreich 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 fractais de dimensão 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 [FRO98], os autores propuseram uma medida acústica para soproidade, 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 do artigo [ROS99] utilizaram técnicas de processamento de sinais para extrair características acústicas da voz, e estas foram utilizadas em redes neurais para determinar as condições glotais 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. Após as transformações, as medidas acústicas foram fornecidas à entrada da rede neural. Cada neurônio de saída representa uma condição das cordas vocais (agrupadas em 6 grupos diferentes). A rede neural 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 [HAD00], os autores propuseram um método de classificação de vozes normais e de vozes patológicas com doenças na laringe. 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 dois protótipos de mapas de distribuição (PDM – *Prototype Distribution Map*), desenvolvido pelos autores. A fdp dos vetores de entrada de vozes desconhecidas como patológicas ou normais, também é modelada pelo PDM. Para os testes foram utilizadas 400 vozes, 100 normais e 300 pacientes com patologias na laringe. Os falantes pronunciaram a vogal sustentada /a/ para ser analisada. 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 [MAR00], 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. A técnica que teve o pior desempenho foi a FFT, segundo os autores, 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 [ROS00] 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, como antes visto, 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%.

Em [RIT01] os autores descreveram um protótipo de um sistema para a avaliação objetiva da qualidade da voz em pacientes em recuperação de vários estágios de câncer da 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 anormal e 6 = mais anormal). O sistema extrai os principais parâmetros, como a impedância dos sinais (Eletroglotografia, EEG) e estes são usados para treinar e testar as redes neurais (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 *Multilayer Perceptron* (MLP) com sete saídas. O algoritmo de treinamento utilizado foi o *back-propagation*. 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.

No artigo [JO01], 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, usando rede pública de telefone, 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 a taxa de classificação obtida utilizando vozes que foram gravadas diretamente em *Digital Audio Tape* (DAT). Foram utilizados setenta e seis casos de patologias e trinta e cinco 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 observado 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 [NIE01], 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 aproximados 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.

2.3 APLICAÇÕES DA TRANSFORMADA WAVELET

Mallat, em [MAL89], descreveu o modelo matemático para o cálculo e interpretação do conceito de representação multiresolução 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).

Kadambe e Boudreaux-Bartels em [KAD92] descreveram a utilização da Transformada Wavelet Diádica (DyWT) como um detector de *Pitch*. A informação do período do *Pitch* é usada em várias aplicações como:

1. Identificação e verificação de falantes;
2. Análise e síntese de fala;
3. Diagnósticos de doenças de voz.

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 exista 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:

1. Não assume estacionariedade dentro da janela de análise;
2. Estima o período *Pitch* precisamente, com erro relativo máximo de 2% para $\text{SNR} \geq -18\text{dB}$;

3. É apropriado para uma maior variação de períodos de *Pitch*;
4. É computacionalmente simples;
5. Exibe resultados superiores aos métodos clássicos de detecção de *Pitch*.

Coifman e Wickerhauser em [COI92], propuseram um método eficiente de compressão de sinais de som e imagem. Este método utiliza uma biblioteca de bases ortogonais, pré-definidas, e um funcional que mede a aproximação da base com um determinado sinal ou uma família de sinais. As funções pré-definidas são wavelet-packets e funções trigonométricas localizadas no domínio tempo e frequência, e possuem a importante propriedade de controle na localização do plano tempo-frequência. 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, ou seja, todas as expansões do sinal nesta nova biblioteca conservam energia, por isso é possível compará-las. 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*, mas 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. Esta técnica será apresentada em mais detalhes no capítulo 3.

No artigo [RAM93], Ramchandran e Vetterli propuseram um algoritmo rápido com o objetivo de encontrar a melhor base (*Best Basis*) na estrutura de decomposição em árvore. Este algoritmo utiliza a Transformada Wavelet Packet (TWP) para compressão de sinais, empregando uma estrutura que inclui taxa de distorção.

O algoritmo é desenvolvido para “podar” as árvores completas da TWP, que significa toda a biblioteca de bases admissíveis da Wavelet Packet, na qual uma sub-árvore é a melhor base segundo alguma função custo, como por exemplo 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 TWP, esta usando o conjunto de filtros Wavelet Daubechies D4. As aplicações usando a TWP apresentaram resultados favoráveis em relação ao JPEG.

Em [SAI94], Saito e Coifman descreveram uma extensão do método *Best Basis* [COI92]. 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 base que são bem localizadas no plano tempo-frequência. Este método foi testado em dois conjuntos de sinais sintéticos onde são extraídas características (coeficientes da expansão dos sinais nas funções bases). Estas características são fornecidas a classificadores, e 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 [TAS96] propôs uma técnica para a procura da *Best Basis* da Transformada Wavelet Packet (TWP), diferente do algoritmo apresentado por Coifman e Wickerhauser [COI92]. Esta técnica faz uma procura na estrutura de árvore da TWP, 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*), que segundo o autor, pode ser utilizado em certas situações. 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. Os autores explicaram que resultados melhores significam ter o mesmo nível de performance, porém com menos esforço computacional.

Em [WU96], os autores descreveram o uso da Wavelet Packet 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 comprimidos, cada decomposição possui determinada informação do sinal em diferentes células tempo-frequência. Aparentemente, segundo os autores, nem todas as decomposições Wavelets contém informação relevante. Assim, o sinal é reconstruído eliminando ou tornando zero as decomposições em Wavelets Packets que contenham pouca informação. Desta forma, o processo de extração de característica é chamado de compressão sendo 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 características Packets selecionadas, quatro critérios foram propostos. O método proposto é testado em dois exemplos práticos:

1. Monitoramento da trepidação em torneamento;
2. Desgaste da ferramenta em perfurar.

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

Em [PAY97] 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 da rede neural, e a função Wavelet usada foi a Daubechies 4. Esta combinação teve sucesso em distinguir diferentes tipos de falhas que ocorreram no modelo mecânico (*drive-line*) descrito na artigo. Este modelo foi testado em condições normais, depois falhas foram inseridas individualmente na caixa de câmbio do motor, e por fim, todas as falhas em conjunto. A taxa de classificação foi de 96% no conjunto de teste. 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 [SOS97], os autores utilizaram a Transformada Wavelet Packet como uma ferramenta para extração de características para fins de classificação de sinais. Para a seleção da *Best Basis* foi usado o algoritmo desenvolvido por Saito e Coifman [SAI94], 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 Potencial*). Para classificar sete tipos (classes) de MUAP, foram utilizados os seguintes classificadores:

1. Linear Discriminant Analysis (LDA);
2. Classification and Regression Trees (CART);
3. k-Nearest Neighbour (k-NN);
4. 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 [MAL97] propuseram um algoritmo para a classificação de dados espectrais minerais, através da Transformada Wavelet Discreta (TWD). A TWD 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 chamado Wavelet adaptativa. O critério de escolha é baseado na probabilidade *a posteriori* chamado *cross-validated based on posterior probabilities* (CVQPM). Os autores aplicaram o algoritmo em um conjunto de dados simulados e em sinais mineralógicos. Nos dois exemplos apresentados o classificador utilizado foi Bayesiano linear. Os resultados da classificação utilizando a Transformada Wavelet adaptativa foram comparados as Wavelets Daubechies e Coiflet, e os índices da classificação exata do grupo de teste foram 97%, 99% e 93%, respectivamente.

No artigo [BEN97], 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 sensoriamento 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 da rede neural. Os experimentos relatados afirmam que o sistema proposto supera o existente, aumentando a exatidão na classificação dos dados de teste.

Os autores do artigo [INT97] utilizaram a combinação de coeficientes da Wavelet Packet e rede neural para classificar classes de sinais. Os tipos de sinais usados neste trabalho são sons de mamíferos marinhos, diferenciando porco do mar, cachalote (semelhante à baleia, mas com dentes) e ruído. Outra aplicação dos autores é na análise de vibração de aviões que requer medidas de frequência e amortecimento de modos de vibração crítica como uma função da velocidade do voo. O objetivo desta combinação é reduzir a complexidade da Rede Neural, aplicando a Transformada Wavelet ao sinal

original e selecionando os maiores coeficientes Wavelet como sinal de entrada da Rede. 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%.

Um exemplo de aplicações da Transformada Wavelet Packet (TWP) em sistemas flexíveis de manufatura encontra-se no artigo [XIA98]. 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 TWP 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 a 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 Transformada Wavelet Packet (TWP) 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 são usados como características para a classificação das condições do sistema. A TWP 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 TWP e FCM foi de 90,25% para o conjunto de teste.

Os autores do artigo [CHA98] descreveram uma nova técnica de extração de características de sinais de voz, chamada 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 a melhor decomposição da árvore, a função custo utilizado foi a entropia de Shannon. Os resultados experimentais foram comparados com a classificação dos coeficientes cepstrum. Os resultados mostraram que o reconhecimento da fala independente do locutor, foram 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.

O autor em [SCH98] 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 soproiedade, 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 um índice existente para o cálculo da soproiedade. Segundo o autor, este novo método serve como auxílio ao diagnóstico assim como para o acompanhamento dos resultados obtidos por um tratamento.

Nos artigos [OBA98] e [OBA99], os autores propuseram um algoritmo baseado na Transformada Wavelet diádica (DyTW) para detectar o período de *Pitch*. Segundo os autores, as técnicas tradicionais tal 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 [ENG99] propuseram as representações tempo-frequência como extrator de características para a classificação de padrões de sinais eletromiográficos (MES – *Myoelectric Signal*). Segundo os autores, as representações tempo-frequência lineares e discretas são preferíveis às transformações quadráticas e contínuas.

Segundo os autores, os sinais eletromiográficos coletados na superfície da pele fornecem informações sobre as atividades neuromusculares e têm sido muito usados para diagnósticos clínicos. Um vetor de características pode então ser objeto para alguma forma de reconhecimento de padrão onde se atribui algum estado. 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-frequência lineares:

1. Características no domínio tempo (TD);
2. Short Time Fourier Transform (STFR);
3. Transformada Wavelet (WT);
4. Transformada Wavelet Packet (TWP).

Os classificadores:

1. Análise discriminante linear (LDA);
2. Perceptron multicamadas (MLP).

Após os autores terem testado todas as combinações das etapas de classificação citadas acima, os resultados apresentaram uma melhora na performance nesta progressão:

1. TD
2. STFT
3. WT
4. TWP

E a TWP forneceu um erro médio de classificação de 6,25%.

Em [KES99] 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:

1. Markov de 1ª ordem e;
2. Banda limitada.

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

No artigo [JIA99] os autores propuseram três métodos de extração de características para fins de classificação de sinais, selecionando átomos no plano tempo-frequência. Os átomos selecionados pelos métodos propostos extraem 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:

1. Forma de ondas triangulares e;
2. 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 [SAI94], e o terceiro é uma combinação dos dois métodos. As menores taxas de erro de classificação para os dois sinais foram utilizando 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 [BAL99] 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 é *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 [MAN00a], alguns métodos de estimação de *Pitch* são comparados e, apontados as vantagens e desvantagens destes métodos. As técnicas utilizadas e descritas neste artigo são:

1. Autoregressivo (AR);
2. *Average Magnitude Difference Function* (AMDF);
3. *Simple Inverse Filter Tracking* (SIFT);
4. *Cepstrum* (CEP);
5. 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, os métodos SIFT e WT apresentaram os melhores resultados.

Outra aplicação de Wavelet em voz encontra-se no artigo [MAN00b], onde a autora propôs a estimação da freqüência 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 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 freqüência fundamental foram utilizados:

1. CWT + AMDF
2. 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 [SAR00], os autores investigaram a utilização de Parâmetros Wavelet Packet (WPP) e parâmetros da *Sub-Banda Base Cepstral* (SBC) 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:

1. *Areneutral* (natural);
2. *Angry* (brabo);
3. *Loud* (alto);
4. *Lombard effect* (efeito “lombard” - situação onde o falante modifica sua fala para aumentar ou melhorar a qualidade de comunicação).

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 [SHI00] é apresentada uma metodologia de classificação de texturas de imagem baseado no modelamento estocástico de textura no domínio frequência. Os autores mostram que diferentes texturas possuem respostas em frequência em bandas intermediárias [CHA93], o que justifica o uso da Transformada Wavelet Packet na tentativa de caracterizar e classificar texturas. Os autores propuseram um esquema de classificação de textura baseado no modelo de Markov no espaço Wavelet. Em relação à Transformada Wavelet padrão (TW), a Transformada Wavelet Packet 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 [HSI00], os autores propuseram a implementação da Transformada Wavelet Packet (TWP) 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 [COI92]. 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 TWP modulada, pois estas preservam as propriedades desejadas.

Em [GAR00] é apresentado um novo método de reconhecimento de faces humanas com iluminação constante. Este sistema é baseado na análise da decomposição Wavelet Packet 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 Wavelet Packet 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 Transformada Wavelet Packet é consideravelmente menor.

No artigo [MEY00], é desenvolvido um algoritmo numérico rápido para Wavelet Packet (WP), 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 frequentemente quantizados para zero.

Foi implementado o algoritmo para compressão chamado *Fast Wavelet Packet* (FWP) e testado em quatro imagens com variações de texturas. As imagens utilizadas são 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: o algoritmo seleciona uma base que preserva as oscilações de textura e o algoritmo é quatro vezes mais rápido.

No artigo [YEN00] é introduzida uma alternativa para extrair informações tempo-frequência das assinaturas de vibrações de máquinas usando a Transformada Wavelet Packet (TWP). O objetivo do trabalho desenvolvido pelos autores é verificar a viabilidade da TWP 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 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 rede neural, que possui a habilidade de capturar a não-linearidade dos dados de entrada.

Os resultados dos testes com TWP foram comparados com a transformada de Fourier, ambos alcançaram excelentes resultados na classificação. Porém, quando os

sinais foram corrompidos artificialmente por diferentes ruídos, simulando um ambiente ruidoso real, a TWP apresentou resultados promissores em relação a Fourier.

2.4 CONSIDERAÇÕES FINAIS

Analisando este capítulo, percebe-se que a Transformada Wavelet Packet tem sido uma ferramenta bastante utilizada no processamento de sinais em particular auxiliando na tarefa de classificação de sinais.

Não se encontrou nenhuma aplicação desta Transformada na tentativa de classificar sinais patológicos ou normais conforme a seção 2.2. Porém, como visto na revisão de literatura, a Transformada Wavelet Packet tem sido muito utilizada no processo de classificação, justificando seu uso para distinguir sinais de vozes patológicas ou normais.

3 FUNDAMENTOS TEÓRICOS

3.1 INTRODUÇÃO

Os fundamentos teóricos estão divididos em duas partes.

Na seção 3.2 deste , será apresentado um breve resumo da fisiologia da voz, assim como das doenças relacionadas ao sistema fonador, mais especificamente na laringe.

Na seção 3.3 será mostrada a teoria da Transformada Wavelet Packet. Porém, para fins didáticos, será apresentada antes uma introdução da Transformada Wavelet Contínua, depois a Transformada Wavelet Discreta, para finalmente se chegar na Transformada Wavelet Packet. Desta forma, a teoria Transformada Wavelet é abordada.

3.2 FISIOLOGIA DA VOZ E PATOLOGIAS

3.2.1 Fisiologia Da Voz

O aparelho da fala pode ser dividido em três subsistemas anatômicos principais [KEN92]:

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

3.2.1 a) Subsistema respiratório

O subsistema respiratório é formado pelos pulmões, traquéia, diafragma e brônquios, conforme mostra a Fig. 3-1. Este sistema, além de fornecer ventilação para o apoio a vida, também produz energia aerodinâmica para a fala. O ato da fala necessita do funcionamento do aparelho respiratório, que fornece a corrente de ar indispensável à produção da maior parte dos sons da linguagem humana. O som é produzido pela regulação da pressão de ar subglotal (pressão abaixo das cordas vocais) gerada pelo subsistema respiratório. 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 articularatório para a formação dos sons.

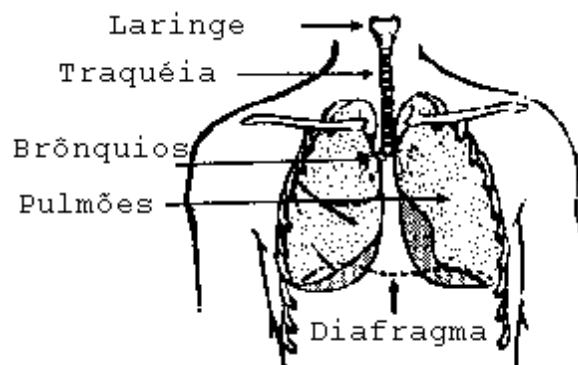


Figura 3. 1 Diagrama esquemático que mostra as partes do sistema respiratório e a laringe (extraído [KEN92]).

3.2.1 b) Subsistema da laringe

A Laringe é constituída por um conjunto arquitetônico de cartilagens unidas por ligamentos, membranas e músculos. Esta arquitetura é cilíndrica, especialmente na sua parte inferior, formada à custa da cartilagem cricóide.

A disposição das principais cartilagens, como a tiróide, cricóide e aritnóides, dão à laringe o formato de uma caixa aérea cilíndrica.

A Laringe é um órgão localizado acima da traquéia, conforme mostrado na Fig 3-1. Na Fig. 3-2 a) tem-se a vista lateral da laringe, composta de suas principais cartilagens:

- A cricóide: constitui sua base, tendo a forma de um anel colocado horizontalmente sobre a traquéia;
- A tireóide: encontra-se na sua parte anterior e superior, com a forma de um escudo ou livro aberto.
- As duas aritenóides, pequenas pirâmides situadas na parte posterior da cricóide. Esta é praticamente imóvel. As outras cartilagens movimentam-se em relação a ela, graças aos diferentes músculos da laringe.

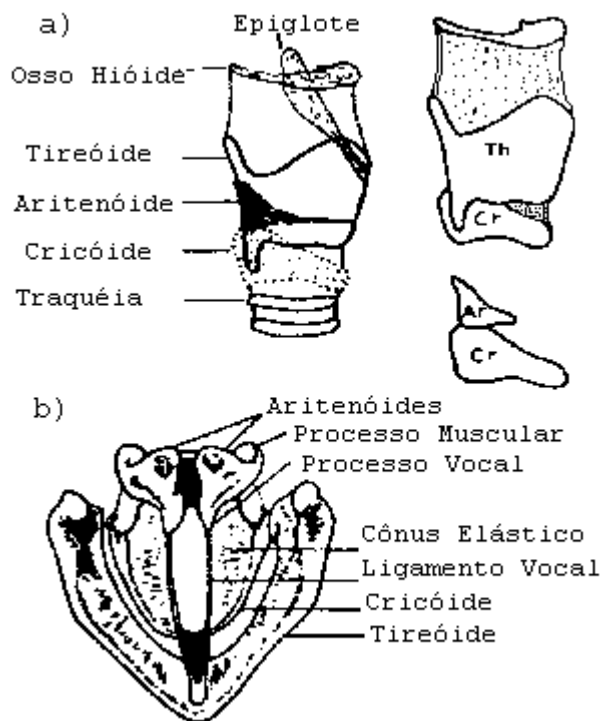


Figura 3. 2 Partes que compõem a Laringe. Em a) tem-se a vista lateral e em b) tem-se a vista superior da Laringe (extraído [KEN92]).

Entre a parte interna da base das aritenóides e da tireóide, estão as cordas vocais. A denominação cordas vocais se refere, na verdade, a dois pares de lábios, simetricamente 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. O afastamento e a aproximação das cordas vocais, provenientes dos movimentos das aritenóides e dos músculos que as dirigem, são responsáveis pela abertura ou fechamento total ou parcial da glote.

As vogais são sons produzidos com a vibração das cordas vocais (sendo o ar uma fonte de energia). No ato da fala, na fase de expiração, um conjunto de músculos aduz 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 supraglótico 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 sessenta a trezentos Hz . 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 [DEL93].

Este subsistema é o responsável pela produção dos sons de vogais (sons vocálicos), através da vibração das cordas vocais.

3.2.1 c) 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 Fig. 3-3. 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 glotais, atuando como cavidade ressonante variável que altera o conteúdo harmônico dos sons glotais.

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 permitido 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 [KEN92].

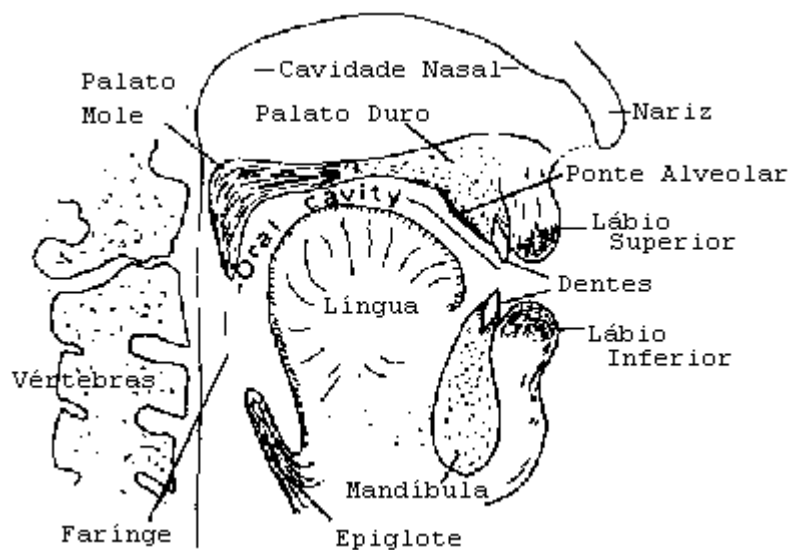


Figura 3. 3 Representação das partes do sistema articulatório (extraído do [KEN92]).

3.2.2 Modelos acústicos da Produção da fala e alguns aspectos espectrais

Há dois tipos elementares de sons na fala: sons vozeados e não-vozeados. Os sons vozeados são sons produzidos pelo ar atravessando a glote ou a abertura das cordas vocais. O som produzido pela laringe é chamado de vozeado ou fonação. Ao contrário, os sons não-vozeados são gerados por uma constrição em algum ponto ao longo do sistema vocal. Ao se forçar o ar através da constrição produz-se turbulência [DEL93]. No entanto, neste trabalho serão tratados apenas os sons vozeados.

A característica espectral da fala é variante no tempo (ou não-estacionária), pois o sistema físico muda no tempo.

No espectro das vogais existem regiões de pico (ou de ênfase) que são chamados ressonantes. Estes picos ressonantes são conseqüências do subsistema articulatório que forma diversas cavidades acústicas no sistema vocal. A localização dos picos ressonantes no domínio frequência, depende das dimensões físicas e da forma do sistema vocal. Cada sistema vocal é caracterizado por um conjunto frequências ressonantes. Do ponto de vista de modelagem do sistema, o subsistema articulatório determina as propriedades do filtro da fala. Como estes picos ressonantes “formam” o espectro de uma forma geral, cientistas da fala referem-se a eles como formantes. Os Formantes do espectro são referidos como F_1 , F_2 , F_3 ..., começando pela menor

freqüência. Em princípio, há infinitos formantes, mas na prática, encontra-se de três a cinco após a conversão analógica/digital apropriada do sinal [DEL93].

As vogais podem ser distinguidas pela localização das formantes no espectro. Na Fig. 3-4, encontram-se 5 vogais /a/, /e/, /i/, /o/ e /u/. Na Fig. 3-4 a) corresponde a representação esquemática do perfil do sistema vocal de cada vogal b) as formas de ondas típicas no tempo e c) o espectro correspondente para cada vogal.

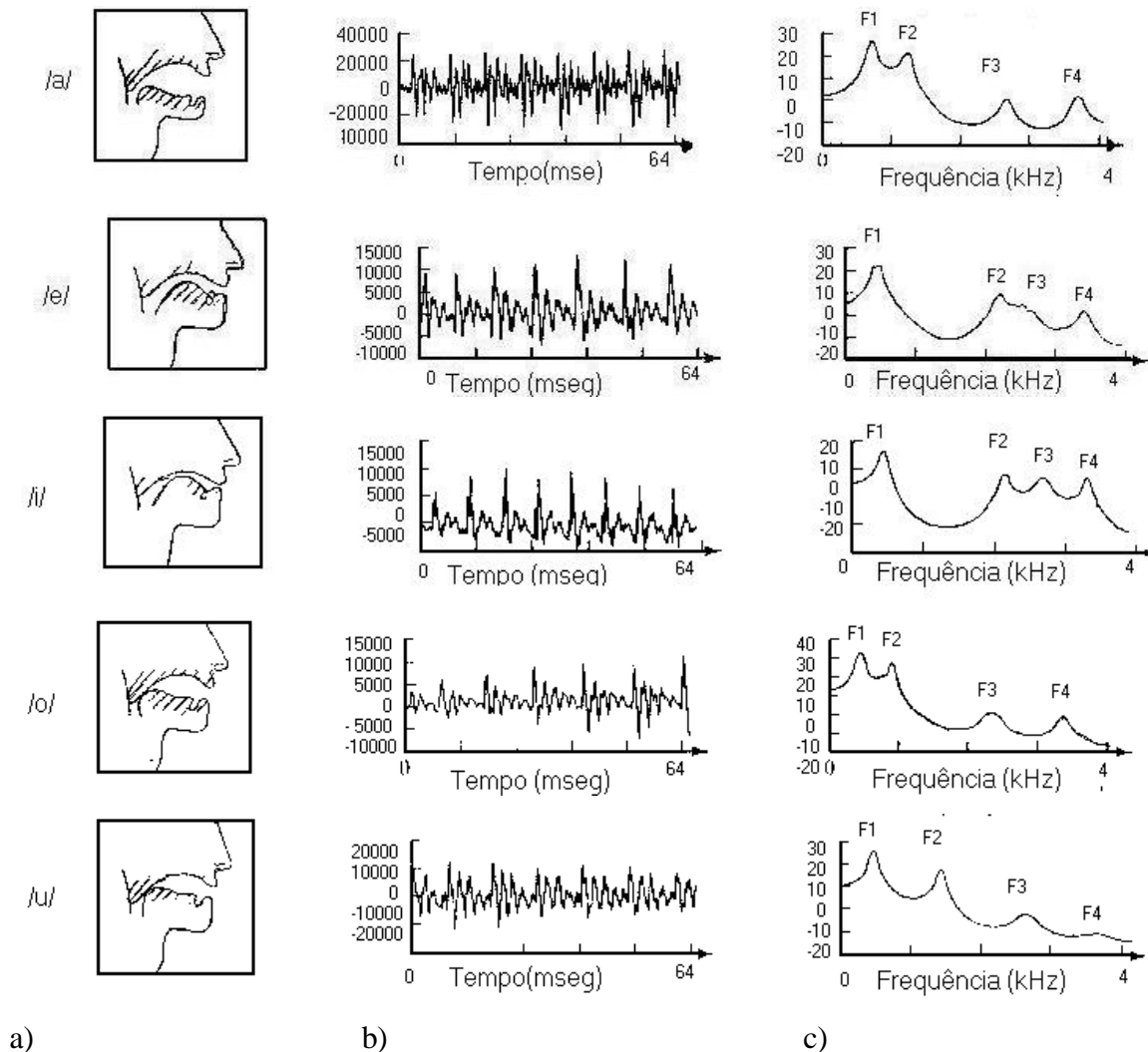


Figura 3-4 a) Representação esquemática do perfil do sistema vocal de cada vogal b) formas de ondas típicas no tempo c) espectro correspondente para cada vogal [DEL93].

Em geral, o formante F_1 varia muito com a altura da língua (acima ou abaixo da boca) ou a abertura da mandíbula e F_2 varia com a posição da língua (frente ou atrás). Para extrair a freqüência fundamental de um sinal de voz se aplica um filtro passa-baixa

para eliminar a influência dos formantes. O filtro deve ser capaz de excluir a menor frequência do subsistema articulatório (formante F_1) mantendo, porém, a frequência fundamental. A vogal que apresenta a maior frequência F_1 é a vogal /a/. Esta vogal é muito usada no exame visual da garganta, pois a abertura da mandíbula e a posição da língua permitem uma boa visão a olho nu ou com o auxílio de algum dispositivo ótico [KEN92].

A fonte de energia acústica das vogais é tipicamente gerada pela vibração das cordas vocais. A forma de onda do som da laringe idealizada é uma série de pulsos triangulares espaçados a um período fundamental T_0 . O espectro da laringe (ou a sua Transformada de Fourier) produzido pela vibração do sistema vocal é um espectro de harmônicas no qual as componentes diminuem a uma taxa de 12dB/oitava, como mostra a Fig. 3-5. O resultado é chamado de espectro de linha, ou espectro na qual a distribuição de energia está na forma de linhas. O espectro de energia da voz pode ser idealizado como um espectro de linha na qual cada linha individual é um múltiplo inteiro da frequência fundamental (f_0). A média da frequência fundamental dos homens é de 120Hz e para a mulher é de 220Hz. Porém, o ser humano pode produzir frequências maiores ou menores que a média. Praticamente falando, o espectro mostra que a energia da voz é predominantemente de componentes harmônicas de baixa-frequência. As relações entre as amplitudes das harmônicas variam muito das configurações da laringe e de cada falante, incluindo certas patologias. Variações na frequência fundamental apresentam alterações no espectro do som laringeal (ou glotal).

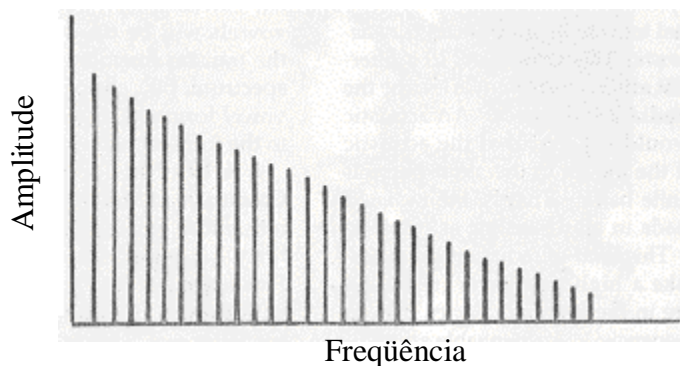


Figura 3-5 Espectro do som da laringe.

O subsistema articulatório é o responsável pelas modificações do som glotal, atuando como uma cavidade ressonante variável, atenuando certas frequências e amplificando outras do sinal glotal gerado pela fonte [KEN92].

Do ponto de vista de modelagem o subsistema articulatório pode ser representado por um filtro, com uma determinada função de transferência que altera o espectro do som glotal.

O modelo fonte-filtro para a produção de vogais é ilustrado na Fig. 3-6 e a análise no domínio freqüência é dada pela equação:

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

Onde:

(f) - freqüência em Hertz;

P(f) - é o espectro de pressão emitido pelo som;

U(f) - espectro do volume da glote;

T(f) - função de transferência do sistema vocal e;

R(f) - é a característica de radiação (interface boca/ meio exterior).

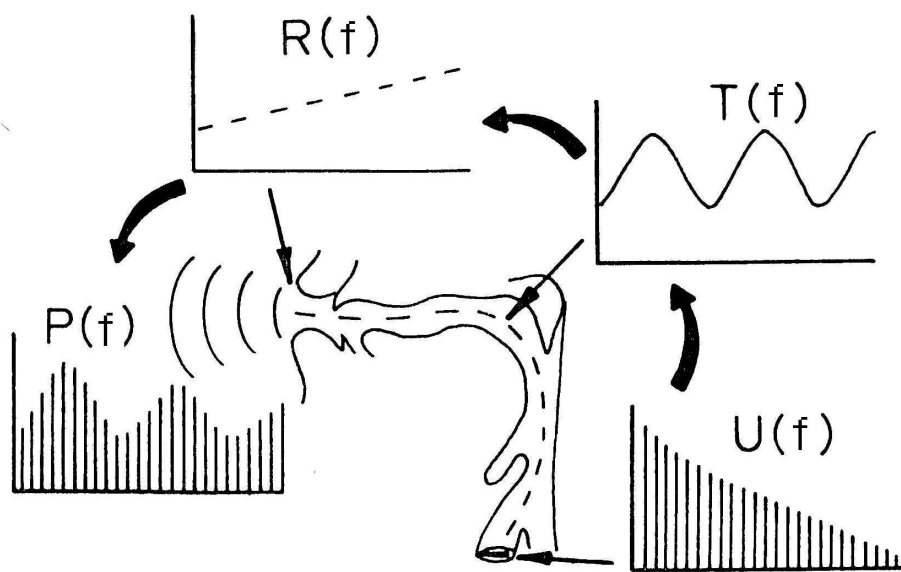


Figura 3-6 Representação do modelo fonte-filtro para vogais (extraído do KEN92).

3.2.3 Patologias

A laringe tem funções biológicas reflexivas que protegem a via aérea e mantém a vida. Essas funções básicas determinam os limites da voz e podem ocasionalmente afetar seu funcionamento. A laringe também fornece o sinal acústico para a fala, capacidade singularmente humana. O canto, a exposição dramática, o riso, o choro preenchem necessidades humanas adicionais através da voz. 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 [COL96].

As pregas vocais, as pregas vestibulares, as pregas ariepiglóticas e outras estruturas dentro da laringe servem para proteger a via aérea de substâncias estranhas. 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 [COL96].

Distúrbios da Voz

Os distúrbios da voz resultam de estruturas ou do mal funcionamento dessas em algum lugar no trato vocal.

Quando a voz muda de alguma forma negativa, diz-se que ela está perturbada ou disfônica. Tais mudanças possuem muitos nomes comuns, tais como: rouquidão, rudeza, aspereza, estridência, entre outras. O termo mais genérico para estas disfunções é a disфонia, que significa qualquer alteração na vocalização normal [BOO94].

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 [COL96].

Na tabela 3-1, 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 [COL96].

Tabela 3-1 Principais Sintomas (extraído de [COL96]).

Rouquidão	Este sintoma reflete uma vibração aperiódica das pregas vocais. Outros termos também utilizados pelos pacientes: voz “rouca”, ”áspera” ou ”raspada”
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 frequência	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. Este sintoma é relatado por jovens adolescentes que utilizam uma frequência inapropriadamente aguda como voz habitual ao invés da voz masculina típica de frequência 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 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.

Sinais de distúrbios vocais são características da voz que podem ser observadas ou testadas. Os sinais representam um inventário de características vocais embasadas em exames, observações e medições [COL96].

Os sinais podem ser:

- a. Perceptuais;
- b. Acústicos;
- c. Fisiológicos.

3.2.3 a) Sinais Perceptuais:

Os sinais perceptuais de distúrbios vocais são as características da voz de um indivíduo que são percebidas pelo ouvinte/observador. Os principais sinais perceptuais estão na tabela 3-2 [COL96].

Tabela 3-2 Sinais Perceptuais de distúrbios da voz.

Sinais Perceptuais de Distúrbios Vocais
Frequência
Frequência inapropriada
Quebras de frequência
Intensidade
Variações de intensidade (suave, excessiva ou descontrolada)
Qualidade
Rouquidão
Soprosidade
Tensão
Tremor
Tensão/compressão
Quebra de soprosidade
Diplofonia
Comportamento não-fonatório
Estridor
Pigarro
Afonia
Consistente
episódica

3.2.3 b) Sinais Acústicos:

A voz é produzida por movimentos das pregas vocais interrompendo o fluxo de ar egressivo. Os movimentos das pregas são controlados pelas características biomecânicas das próprias pregas, pela magnitude da pressão de ar abaixo das cordas e por seu controle neural. A patologia pode afetar tais movimentos interferindo em quaisquer destas variáveis.

O movimento das pregas vocais resulta em interrupção periódica do fluxo de ar em velocidade apropriada a percepção do som. A acústica é o estudo do som e a acústica vocal pode suprir importantes informações referentes ao movimento das pregas vocais. Há uma grande quantidade de correspondência entre a fisiologia e a acústica e muito pode ser inferido sobre a fisiologia com base em análise acústica. Ademais, os parâmetros acústicos são provavelmente os mais fáceis de registrar e analisar objetivamente.

Há muitos sinais acústicos (tabela 3-3) que podem ser associados a qualquer patologia. O *Jitter* (perturbação de frequência) é um dos índices que reflete anomalias das pregas vocais e pode ser mais fácil de medir. Porém, todos os sinais acústicos refletem algum aspecto da patologia de interesse subjacente.

A frequência de vibração das pregas vocais é referida como a frequência fundamental, que em geral é definida como frequência fundamental média, ou, no caso de fala, frequência fundamental da fala.

A patologia pode afetar a frequência de vibração, com o resultado de que homens e mulheres produzirão uma frequência excessivamente elevada ou baixa, ou seja, fora dos valores médios para o sexo.

A perturbação se refere à irregularidade da vibração das pregas vocais, denominada *Jitter* vocal. Os falantes normais apresentam uma pequena quantidade de perturbação, que pode representar uma variação de massa, tensão, atividade muscular ou atividade neural das pregas vocais.

A perturbação ou *Jitter* pode ser expressa de uma variedade de modos. Em [BAK87] estão descritas diversas técnicas para o cálculo de *Jitter*.

Do mesmo modo que para a frequência fundamental, é possível que a amplitude do som da prega vocal varie de um ciclo para o seguinte. Esta característica é denominada perturbação de amplitude ou *Shimmer* [COL96].

Tabela 3-3 Principais sinais acústicos.

Sinais Acústicos

Frequência fundamental
Frequência fundamental média da fala
Variabilidade da frequência
Extensão fonatória
Perturbação
Amplitude
Nível médio de pressão sonora
Variabilidade de amplitude
Extensão dinâmica
Ruído espectral
Tempo de início e fim de sonorização
Início e fim da emissão
Tempo de fonação
Quebras de sonoridade
Quebras de frequência
Acústica normal

3.2.3 c) Sinais Fisiológicos:

Os sinais fisiológicos que podem ser afetados por patologia incluem características aerodinâmicas (fluxo de ar e pressão), comportamento vibratório (área de contato, forma de onda) e atividade muscular. Estes estão resumidos na tabela 3-4 [COL96].

Tabela 3-4 Principais Sinais Fisiológicos.

Sinais Fisiológicos

Aerodinâmica
Fluxo de ar aumentado ou reduzido
Pressão de ar aumentada ou reduzida
Variabilidade do fluxo de ar
Comportamento vibratório
Atividade Muscular

A seguir, será apresentado um resumo de algumas patologias mais comuns que ocorrem na laringe.

Disfonia Funcional

Algumas das vozes mais perturbadas que ouvimos podem não ter causas orgânicas ou físicas. Os pacientes podem aproximar as pregas de uma maneira frouxa, produzindo soprosidade ou de forma apertada, produzindo sintomas de aspereza ou constrição. Problemas vocais que não resultam de uma patologia orgânica podem ser denominadas de disfonia funcional, um termo que transmite muito pouco ao clínico de voz além da importante implicação de que não há uma patologia estrutural presente. Uma disfonia de causa funcional não soa, necessariamente, diferente de uma orgânica. Algumas das vozes mais ásperas são produzidas por pessoas cujas laringes não demonstram qualquer patologia. Por outro lado, problemas orgânicos sérios, como câncer em fase inicial pode não produzir qualquer alteração da voz.

A qualidade da voz é, muitas vezes, o problema principal na disfonia funcional. As pregas vocais podem aproximar-se de uma maneira deficiente e produzir alterações na qualidade da voz; palavras como rouca, áspera, estridente e soprosa são aplicadas a estas dimensões de qualidade de voz da disfonia funcional. Aspereza é muitas vezes, o produto da superaproximação das pregas, quando elas aproximam-se agudamente, produzindo o que é também percebido como ataque vocal brusco. A aspereza também indica uma voz que requer muito esforço e força para ser produzida. O problema oposto pode ser a soprosidade, produzida por pregas que podem aproximar-se com demasiada frouxidão. A frouxidão de aproximação geralmente produz um escape de ar excessivo, percebido como soprosidade.

Nódulos Vocais

Nódulos nas cordas vocais são as lesões benignas mais comuns das pregas vocais, tanto em crianças como em adultos. 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. À medida que os nódulos bilaterais aproximam-se um do outro na vocalização, há uma fenda glótica aberta de cada lado do

nódulo. Esta fenda glótica aberta resulta numa falta de adução completa das pregas vocais. Esta aproximação incorreta leva a soprosidade na voz e ao desperdício de ar. Também a massa aumentada dos nódulos nas pregas vocais contribui para uma altura de voz mais grave e maior periodicidade (julgada como rouquidão) [BOO94]. 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 [COL96].

Pólipos vocais

Os pólipos vocais comumente ocorrem na mesma região das pregas vocais como os nódulos, e são mais freqüentes 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 disфонia severa. A prega vocal normal vibra em uma freqüência, 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 [BOO94].

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. 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 [COL96].

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 frequentemente associado ao fumo. Dos sinais acústicos, a frequência fundamental da fonação é mais baixa do que esperada para o sexo e a idade do paciente. Pode haver frequência aumentada e perturbação da amplitude, bem como a presença de ruído espectral devido à rouquidão [COL96].

Edema de Reinke é o edema bilateral da camada subepitelial 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 com um edema, pobre em vasos o que modifica completamente o aspecto anatômico da região glótica. O edema de Reinke é limitado às cordas vocais e atinge até a região aritnóidea, onde alcança maior espessura. O sintoma clássico do Edema de Reinke é a disfonia, produzida por cordas vocais mais flácidas [KUH82].

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. Frequência e perturbação de amplitude acima do normal são esperadas, assim como o ruído espectral aumentado. O principal sintoma vocal é a rouquidão [COL96].

Laringectomia

Uma das funções primárias da laringe humana é prevenir a aspiração de alimentos e líquidos na via aérea. A laringe contém três regiões da 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 [BOO94].

Câncer

Câncer ou carcinoma no sistema vocal é uma doença ameaçadora da vida 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: supraglótica, envolvendo estruturas como as pregas ventricular e ariepiglótica, a epiglote, as cartilagens aritenóideas e as paredes da hipofaringe; glótica, da comissura anterior até as extremidades do processo vocal das aritenóideas; e subglótica, 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.

Leucoplasia

Leucoplasia são sinais esbranquiçados que são lesões de acréscimo à membrana superficial da mucosa e, muitas vezes, estendem-se por baixo da superfície para o espaço subepitelial. 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 soproidade, intensidade reduzida e disфонia geral. O tratamento da leucoplasia é médico-cirúrgico e a terapia vocal apenas contribui para desenvolver a melhor voz possível [COL96].

Paralisia

Lesões que afetam o nervo vago em alguma parte ao longo do seu curso, desde a base até a laringe. Estas lesões resultam 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 NLR, 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 soproidade e a rouquidão. Ocasionalmente, a diploфонia pode estar presente. A paralisia bilateral do tipo adutor causará soproidade severa 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 [BOO94], [COL96].

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 [COL96]. Pacientes com granuloma podem experimentar disfonia severa, caracterizada por rouquidão, soprosidade e a necessidade de limpar a garganta freqüentemente [BOO94].

3.3 TRANSFORMADA WAVELET PACKET

3.3.1 Introdução

Muitos sinais na natureza, particularmente os biológicos, possuem características não estacionárias, tais como os sinais de voz, imagens, eletrocardiograma (ECG), entre outros [AKA98]. A análise do sinal utilizando-se a Transformada de Fourier (TF) é a mais adequada para análise de sinais estacionários [BEN86]. Ao aplicar-se a TF a um sinal, a informação de tempo é perdida e apenas a informação de frequência está presente. A TF não se aplica a sinais não estacionários e é definida como:

$$X(f) = F[x(t)] = \int_{-\infty}^{\infty} x(t) \cdot e^{-j \cdot 2\pi \cdot f \cdot t} dt \quad (3.1)$$

Na análise por Fourier não existe a preocupação em se saber o instante em que as componentes de frequência do sinal atuam. É tacitamente assumido que as componentes em frequência do sinal não se alteram com o tempo, já que a Transformada de Fourier só se aplica para sinais estacionários.

Gabor adaptou a Transformada de Fourier para analisar apenas uma pequena seção do sinal no tempo, uma técnica chamada de *windowing* (janelamento) 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 [RIO91], conforme a equação (3.2).

$$STFT(\tau, f) = \int x(t) g(t - \tau) e^{-j2\pi f t} dt \quad (3.2)$$

Esta função pode ser descrita como a Transformada de Fourier do sinal $x(t)$, previamente ‘janelada’ por uma função $g(t)$ ao redor do tempo τ . Este método assume estacionariedade do sinal dentro desta janela limitada $g(t)$.

Entretanto, 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 um trecho estreito (janela estreita) tem associado no espectro uma função com lóbulo principal largo, o que reduz a resolução entre frequências contínuas.

Dada uma janela $g(t)$ e a sua Transformada de Fourier $G(f)$, a largura de banda (ou resolução espectral) Δf é definida como [RIO91], [SHE96], [AKA98]:

$$\Delta f^2 = \frac{\int f^2 |G(f)|^2 df}{\int |G(f)|^2 df} \quad (3.3)$$

Da mesma forma, a resolução temporal na análise STFT, é dada por Δt :

$$\Delta t^2 = \frac{\int t^2 |g(t)|^2 dt}{\int |g(t)|^2 dt} \quad (3.4)$$

Note-se que os denominadores das equações (3.3) e (3.4) são a energia da janela $g(t)$.

A resolução no tempo e em frequência não pode ser arbitrariamente pequena, e seu produto fica limitado pela inequação de Heisenberg [RIO91]:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \quad (3.5)$$

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 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 de um determinado sinal [RIO91].

A Transformada Wavelet surgiu há poucos anos como uma alternativa a STFT para análise tempo-frequência. Na seção 3.3.2 será apresentada a Transformada Wavelet Contínua.

Para entender a nomenclatura utilizada neste capítulo, serão feitas algumas considerações preliminares:

Seja então o espaço de todas as funções reais quadraticamente integráveis, conforme a equação (3.6).

$$\int_{-\infty}^{\infty} |f(t)|^2 dt < \infty \quad (3.6)$$

O espaço formado por estas funções é denominado de espaço de Hilbert, notado por $L^2(\mathfrak{R})$.

O produto interno entre duas funções f e g , ambas pertencentes ao espaço $L^2(\mathfrak{R})$, é definido por [WIC94]:

$$\langle f, g \rangle_{L^2(\mathfrak{R})} \stackrel{def}{=} \int f(t)g(t)dt \quad (3.7)$$

E a integral do quadrado da função:

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \langle f, f \rangle_{L^2} \quad (3.8)$$

A norma no espaço $L^2(\mathfrak{R})$ pode ser definida como o produto interno conforme a equação:

$$\|f\|_{L^2} \stackrel{def}{=} \langle f, f \rangle^{1/2} \quad (3.9)$$

A coleção de funções u mensuráveis ($\|u\|_p < \infty$) é chamada de espaço de Lebesgue denotado por $L^p(\mathfrak{R})$. O espaço de Lebesgue para funções discretas é representado por $l^p(Z)$.

As normas L^p , denotada por $\| \cdot \|_p$ para $1 \leq p < \infty$, são definidas para funções u mensuráveis:

$$\|u\|_p = \left[\int |u(t)|^p dt \right]^{1/p} \quad (3.10)$$

Sejam 2 subespaços vetoriais ortogonais $A_{2n,k}$ e $A_{2n+1,k}$. É mostrado em [WIC94] que $A_{n,k+1}$ é a soma direta dos dois subespaços vetoriais, conforme a equação (3.11):

$$A_{n,k+1} = A_{2n,k} \oplus A_{2n+1,k}. \quad (3.11)$$

3.3.2 Transformada Wavelet Contínua – CWT

A análise de Wavelet foi proposta com o intuito de 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 multiresolução [RIO91]. Esta análise é análoga a um banco de filtros, e a resolução temporal deve aumentar com a frequência central da análise dos filtros. Por isso, foi proposto que Δf seja proporcional a f , de forma que:

$$\frac{\Delta f}{f} = c \quad (3.12)$$

Onde c é uma constante igual ao inverso do fator de qualidade Q do filtro.

A análise do banco de filtros é composta por filtros passa-banda, com fator de qualidade Q constante [RIO91].

Quando a relação da equação (3.12) é satisfeita, Δf e Δt mudam com a frequência central do filtro, satisfazendo a inequação de *Heisenberg*, porém a resolução no tempo se torna excelente para altas frequências, enquanto que se tem ótima resolução em frequência nas frequências mais baixas [RIO91].

A análise por Wavelet usa uma janela temporal pequena onde se deseja mais informações de alta frequência, e janelas temporais maiores onde se deseja informações de baixa frequência.

Na Transformada Wavelet a noção de escala é introduzida como uma alternativa para a frequência, gerando então a chamada representação tempo-escala. O sinal é então representado em um plano tempo-escala, assim como o STFT representa o sinal no plano tempo-frequência [RIO91].

A Transformada Wavelet Contínua (CWT) é definida a partir de um espaço de funções ortonormais, denominado baixas Wavelets, as quais formam uma base de

funções da Transformada Wavelet. Assim a CWT é definida como o produto interno na forma:

$$CWT_{f(b,a)} = \langle f(t), \psi_{a,b}(t) \rangle = \int f(t) \cdot \psi_{a,b}(t) \cdot dt \quad (3.13)$$

onde $\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$ e a é o fator de escala e b a variável de

deslocamento.

A função $\psi_{a,b}(t)$ ($a=1$ e $b=0$) é chamada de *Wavelet* básica ou *Wavelet* mãe. Variando-se apropriadamente os termos a e b construi-se a base de funções da Transformada. A CWT mede a similaridade entre o sinal e a função básica $\psi_{a,b}(t)$. As Wavelets $\psi_{a,b}(t)$ são versões escalonadas pelo parâmetro a e transladadas pelo parâmetro b da função básica $\psi(t)$ ou *Wavelet* mãe.

A constante $\frac{1}{\sqrt{|a|}}$ é usado para normalização da energia da função Wavelet em diferentes escalas tal que:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1, \text{ para qualquer nível de dilatação} \quad (3.14)$$

A Transformada Wavelet Contínua Inversa é obtida pela equação (3.15):

$$f(t) = \frac{1}{c_h} \iint CWT(a,b) \psi_{a,b}(t) \frac{da \cdot db}{a^2} \quad (3.15)$$

Da equação (3.15), c_h é uma constante que depende de $\psi(t)$. A equação (3.16) é chamada de condição de admissibilidade, que garante a existência da Transformada Inversa e satisfaz a Resolução de Identidade (mais detalhes em[SHE96]):

$$c_h = \int \frac{|H(w)|^2}{|w|} dw < +\infty \quad (3.16)$$

onde $H(w)$ é a Transformada de Fourier da função Wavelet $\psi(t)$ [SHE96].

A condição de Admissibilidade implica que a Transformada de Fourier da Wavelet deve ser zero na componente de frequência $w=0$:

$$|H(w)|^2 \Big|_{w=0} = 0 \quad (3.17)$$

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

$$\int \psi(t) dt = 0 \quad (3.18)$$

As Wavelets mães podem assumir diversas formas. A escolha da Wavelet mãe é um fator importante, mas a melhor escolha para uma determinada aplicação em particular é uma área ainda em pesquisa [BEN94].

3.3.3 Transformada Wavelet Discreta - DWT

O cálculo dos coeficientes de todas as possíveis escalas requer um enorme esforço computacional e gera informação redundante. A Transformada Wavelet Contínua de um sinal de uma dimensão gera uma representação de duas dimensões no espaço tempo-escala.

Uma alternativa à CWT é a utilização de parâmetros de escalonamento e translação discretos. A esta análise dá-se o nome de *Discrete Wavelet Transform* (DWT) ou Transformada Wavelet Discreta [SHE96].

A Transformada Wavelet Discreta de uma função $f(t)$ é definida como:

$$DWT(m, p) = d_{m,p} = \int_{-\infty}^{\infty} f(t) \cdot \Psi_{m,p} dt \quad (3.19)$$

onde $\Psi_{m,p}$ forma uma base de funções Wavelets, criadas a partir da função Wavelet mãe $\psi_{a,b}(t)$, mediante parâmetros de escala e translação discretos (m e p , respectivamente). Assim, $\psi_{a,b}(t)$ é dado por:

$$\Psi_{m,p} = \frac{1}{\sqrt{a_0^m}} \Psi\left(\frac{t}{a_0^m} - p \cdot b_0\right) = \frac{1}{\sqrt{a_0^m}} \Psi\left(\frac{t - p \cdot a_0^m \cdot b_0}{a_0^m}\right) \quad (3.20)$$

onde a_0 e b_0 são constantes e m e p pertencem ao conjunto dos inteiros.

Os parâmetros de dilatação e translação são m e p , respectivamente. E as constantes a_0 e b_0 dão a variação da dilatação e o passo da translação, respectivamente.

Uma escolha feita para as constantes a_0 e b_0 são $a_0 = 2$ e $b_0 = 1$. Desta forma, temos a Wavelet diádica, amostrada a uma taxa 2^m . Também se escolhe que as componentes $\Psi_{m,p}$ do conjunto de funções Wavelets sejam funções ortogonais e reais. A função $\Psi_{m,p}$ será ortogonal se, e somente se:

$$\langle \Psi_{m,p}, \Psi_{m,k} \rangle = \delta(p-k) = \begin{cases} 1, & p = k; \\ 0, & p \neq k. \end{cases} \quad (3.21)$$

Então se tem que:

$$\Psi_{m,p} = \frac{1}{\sqrt{2^m}} \Psi\left(\frac{t - 2^m \cdot p}{2^m}\right); \quad m, p \in \text{inteiros}. \quad (3.22)$$

Desta forma, tem-se uma escala de dilatação como uma potência de dois ($a_m = 2^m$), e passos de translação de um passo da escala de dilatação ($b_p = 2^m \cdot p = a_m \cdot p$).

A reconstrução do sinal $f(t)$ através dos coeficientes Wavelets exige uma escolha muito especial de $\psi(t)$. Daubechies [DAU90] provou que a condição necessária e suficiente para a reconstrução do sinal dos coeficientes Wavelets é que a energia, somatório do módulo ao quadrado de DWT(m,p), deve estar entre dois limites positivos:

$$A \|f\|^2 \leq \sum_{m,p} |\langle f, \Psi_{m,p} \rangle|^2 \leq B \|f\|^2 \quad (3.23)$$

onde $\|f\|^2$ é a energia de $f(t)$, $A > 0$ e $B < \infty$.

Se a condição da equação (3.23) é satisfeita, $\psi_{m,p}(t)$ é um *frame*, e A e B são os limites deste *frame* [SHE96]. Se $A=B$, a energia da Transformada Wavelet é proporcional a energia do sinal e as Wavelets se comportam como bases ortogonais.

A reconstrução do sinal é dada pela equação (3.24) [DAU90]:

$$f(t) = \frac{2}{A+B} \sum_m \sum_p c_{m,p} \psi_{m,p}(t) \quad (3.24)$$

Uma maneira eficiente de aplicar esta Transformada é através de filtros, técnica desenvolvida por Mallat [MAL89], onde se tem a decomposição da Wavelet implementando a análise Multiresolução. A análise multiresolução é uma técnica que permite analisar sinais em múltiplas bandas de frequências [SHE96].

A Transformada Wavelet multiresolução pode ser considerada como um filtro passa-baixa e outro passa-alta. A função Wavelet $\psi(t)$ está relacionada a um filtro passa-alta, a qual produz os coeficientes de detalhes da decomposição Wavelet. Na análise multiresolução, há uma função adicional que está relacionada ao filtro passa-baixa. Esta é chamada de função de escalonamento $\phi(t)$ (*scaling functions*), e está associada com os coeficientes de aproximação da decomposição da Wavelet. O processo de filtragem de decomposição Wavelet Discreta multiresolução está na Fig. 3-7.

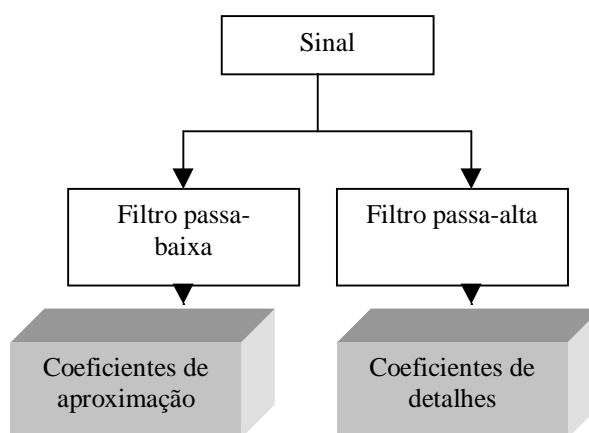


Figura 3-7 Decomposição Wavelet Discreta multiresolução.

Mallat, em seu artigo [MAL89], analisou a decomposição multidimensional de imagens utilizando os filtros passa-baixa e passa-alta. Dos coeficientes do filtro passa-

baixa, tem-se uma aproximação da imagem original, originando os coeficientes de aproximação. Do filtro passa-alta surgem os coeficientes da Wavelet para cada detalhe da imagem, como as bordas da imagem, originando os coeficientes de detalhes. Para muitos sinais, inclusive a voz humana, as componentes de baixa frequência são aquelas que contêm maior informação. Já nas componentes de alta frequência da voz, são encontrados detalhes que, se removidos, alteram o som mas ainda é possível saber o que está sendo falado (não há perda de inteligibilidade). A remoção das componentes de baixa frequência, contudo, afeta drasticamente a inteligibilidade.

Os filtros digitais utilizados na DWT multiresolução, chamados *Quadrature Filter* (QF), tem como característica filtrar e decimar o sinal, respectivamente. Decimação é uma operação que diminui a taxa de amostragem do sinal [OPP89]. Em outras palavras, significa eliminar dados ou pontos intercalados. Decimar um sinal pelo fator 2, significa que a cada 2 pontos, um é eliminado (sub-amostrado por um fator de 2), obtendo-se metade do sinal original. No processo de decimação, aumenta-se o tamanho do período de amostragem do sinal, e conseqüentemente, diminui-se a taxa de amostragem.

O processo de decimação está representado na Fig. 3-8. Tem-se no primeiro bloco o sinal original com 1000 amostras. Após o sinal passar pelos filtros passa-baixa e passa-alta, sofrem decimação e dão origem aos coeficientes Wavelet, os cA (coeficientes de aproximação) e os cD (coeficientes de detalhes), cada um com metade da dimensão do sinal original.

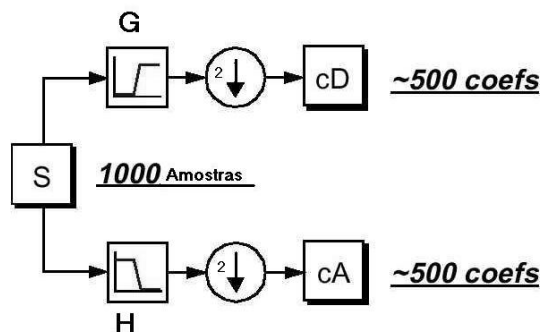


Figura 3-8 Processo de filtragem e decimação (H filtro passa-baixa e G filtro passa-alta).

A implementação da DWT pode ser introduzida como uma decomposição em sub-bandas, ou seja, o processo de decomposição é iterativo com sucessivas decomposições nos coeficientes de aproximação. Desta forma, obtém-se a 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 Fig. 3-9:

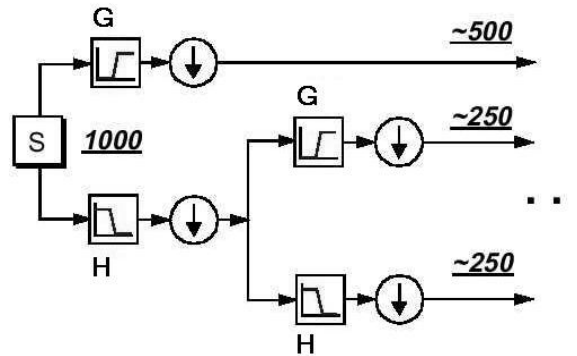


Figura 3-9 Decomposição em sub-bandas Wavelet Discreta.

A função de escalonamento $\phi(t)$ é contínua e geralmente real. Esta função não satisfaz a condição de admissibilidade, ou seja:

$$\int \Phi(t) dt = 1 \quad (3.25)$$

A função de escalonamento $\phi(t)$ é dilatada pelo fator de escala diádica:

$$\phi_{m,p} = 2^{-m/2} \phi(2^{-m} t - p) \quad (3.26)$$

Os coeficientes de aproximação $c_{m,p}$ de uma função $f(t)$ são dados por:

$$c_{m,p} = \langle f, \Phi_{m,p} \rangle \quad (3.27)$$

onde m indica o nível de dilatação.

A projeção de um sinal $f(t)$ na base ortogonal $\phi(t)$ forma um subespaço V_0 que é uma aproximação “borrada” de $f(t)$ [SHE96]. A função dilatada $\phi(t/2)$ gera um subespaço V_1 de dilatação $m=1$, sendo que $V_1 \supset V_0$. A base deste subespaço é $\phi_{1,p}(t)$. Este processo pode ser repetido sucessivamente, formando subespaços para diferentes valores de dilatação m . A dilatação dos espaços formados é 2^m e $V_m \supset V_{m-1} \supset \dots V_0$.

Uma representação gráfica destes espaços está na Fig. 3-10.

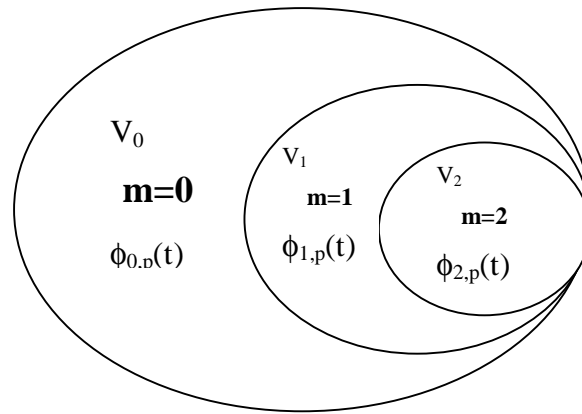


Figura 3-10 Representação dos espaços vetoriais para m variando de 0 a 2.

A diferença dos espaços ($V_0 - V_1$) é o subespaço W_1 , cuja base é $\Psi_{m,p}$. As funções $\Psi_{m,p}$ e $\Phi_{m,p}$ devem ser ortogonais entre si, como também devem ser ortogonais em relação a elas próprias em relação a translação e dilatação discretas. Assim, conforme [SIM93], tem-se que:

$$\begin{aligned} \langle \Phi_{m,p}, \Phi_{m,k} \rangle &= \delta(p-k) \\ \langle \Psi_{m,p}, \Psi_{m,k} \rangle &= \delta(p-k) \quad e \\ \langle \Phi_{m,p}, \Psi_{m,k} \rangle &= 0 \end{aligned} \quad (3.28)$$

onde

$$\delta(p-k) = \begin{cases} 1, & p = k; \\ 0, & p \neq k. \end{cases} \quad (3.29)$$

A decomposição da DWT em sub-bandas pode ser implementada eficientemente como mostra a Fig. 3-9. A filtragem é feita utilizando o par de filtros passa-baixa e passa-alta, chamados de Filtros de Quadratura (*Quadrature filter* - QF). Através do par de filtro de Quadratura é possível obter uma cobertura aproximada do espectro de frequências [BEN94].

A função de escalonamento e as Wavelet estão associados a estes filtros QF através da equação de 2 escalas [SIM93], conforme as equações abaixo:

$$h(p-2k) = \langle \phi_{m-1,p}, \phi_{m,k} \rangle = \frac{1}{\sqrt{2}} \int \phi(t-p) \phi\left(\frac{t}{2}-k\right) dt \quad (3.30)$$

e

$$g(p-2k) = \langle \phi_{m-1,p}, \psi_{m,k} \rangle = \frac{1}{\sqrt{2}} \int \phi(t-p) \psi\left(\frac{t}{2}-k\right) dt \quad (3.31)$$

Desta forma, pode-se calcular os coeficientes de um nível de resolução para outro utilizando filtros digitais (QF), e desenvolver algoritmos rápidos recursivamente [SCH98]. Estas relações estão nas equações (3.30) e (3.31), e estas são constantes em qualquer nível de dilatação, ou seja, não é necessário recalculá-las para cada novo nível de dilatação.

$$c_{m,k} = \sum_p c_{m-1,p} h(p-2k) \quad (3.32)$$

$$d_{m,k} = \sum_p d_{m-1,p} g(p-2k) \quad (3.33)$$

As equações de duas escalas são fundamentais na análise multiresolução [SHE96].

Para exemplificar a função de escalonamento e a função Wavelet associada é apresentada a função Wavelet Haar, que tem sua função de escalonamento definida por:

$$\Phi(t) = \begin{cases} 1, & 0 \leq t < 1; \\ 0, & t \geq 1 \text{ e } t < 0. \end{cases} \quad (3.34)$$

E sua função Wavelet definida por:

$$\Psi(t) = \begin{cases} 1, & 0 \leq t < 1/2; \\ -1, & 1/2 \leq t < 1; \\ 0, & \text{qualquer outro } t. \end{cases} \quad (3.35)$$

3.3.4 Transformada Wavelet Packet

A Transformada Wavelet Packet é uma generalização do conceito da Transformada Wavelet Discreta, na qual a resolução tempo-freqüência pode ser escolhida de acordo com o sinal. Isto é feito dentro dos limites do princípio de incerteza de Heisenberg [KAR98].

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 Wavelet Discreta, o sinal é dividido em coeficientes de aproximação e detalhes, como visto na Fig. 3-8. Apenas os coeficientes de aproximação são novamente divididos em coeficientes de aproximação e detalhes, chamado de segundo nível de decomposição. E assim pode ser feito n vezes, onde 2^n é o tamanho do sinal amostrado, conforme a Fig. 3-9. 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 [COI98].

O cálculo numérico dos coeficientes da Wavelet Packet é a seqüência do produto interno de um sinal $x = x(t)$ em $L^2(\mathfrak{R})$ com as funções Wavelet Packet $w_{sfp}(t)$:

$$\lambda_{sf}(p) \stackrel{def}{=} \langle x, w_{sfp} \rangle = \int_{\mathfrak{R}} x(t) \cdot 2^{-s/2} w_f(2^{-s}t - p) dt \quad (3.36)$$

onde $\lambda_{sf}(p)$ são os coeficientes da expansão de x nas funções Wavelet Packet, $w_{sfp}(t)$ são as funções Wavelet Packet básicas, s é o índice de escala (ou dilatação), p é o índice de posição (ou translação), e f é o índice de freqüência (ou também denominado por n de oscilações, dependendo da literatura); $s, p \in \mathbb{Z}$, e $f \geq 0$.

Ao contrário da Wavelet Discreta, que é função de duas variáveis (escalonamento e translação), a Wavelet Packet é uma função de três variáveis incluindo a variável f que corresponde a um índice de freqüência da Wavelet [WIC94].

Seja H e G o par conjugado de filtros de quadratura (*Quadrature Filters* QF).

A biblioteca de funções Wavelet Packet é construída com um algoritmo recursivo, e as funções $w_f(t)$ são definidas, considerando a escala fixa [WIC94], conforme as equações (3.37) e (3.38) .

$$w_{2f}(t) = \sqrt{2} \sum_{p \in \mathbb{Z}} h(p) \cdot w_f(2t - p) \quad (3.37)$$

$$w_{2f+1}(t) = \sqrt{2} \sum_{p \in \mathbb{Z}} g(p) \cdot w_f(2t - p) \text{ para } f=0,1,2,3\dots \quad (3.38)$$

onde $h(p)$ e $g(p)$ são os Filtros de Quadratura H e G (operadores que convoluem e decimam).

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 [WIC94]. Ao contrário da DWT, a Transformada Wavelet Packet utiliza mais filtros G em cada nível de dilatação.

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 (3.39):

$$w_{sfp}(t) \stackrel{def}{=} 2^{-s/2} w_f(2^{-s}t - p) \quad (3.39)$$

A decomposição Wavelet Packet pode ser implementada aplicando os filtros H e G na seqüência de coeficientes λ_{sf} da equação (3.36), satisfazendo a relação recursiva [WIC94]:

$$\lambda_{s+1,2f}(p) = \sum_{j \in \mathbb{Z}} h(j) \lambda_{sf}(2p - j) = H \lambda_{sf}(p) \quad (3.40)$$

$$\lambda_{s+1,2f+1}(p) = \sum_{j \in \mathbb{Z}} g(j) \lambda_{sf}(2p - j) = G \lambda_{sf}(p) \quad (3.41)$$

As provas das equações (3.40) e (3.41) encontram-se em [WIC94].

Com as equações (3.40) e (3.41) não é necessário recalcular o produto interno de todo o sinal, apenas aplicar os filtros H e G nos coeficientes do nível de dilatação anterior.

Se forem aplicados sucessivamente um certo número de filtros H seguidos de um único filtro G ao sinal x , tem-se a Transformada Wavelet Discreta conforme a equação (3.42) e o esquema da Fig. 3-11.

$$\int_R x(t) \cdot 2^{-s/2} w_1(2^{-s}t - p) dt = GH^{s-1}x(k) \quad (3.42)$$

onde $x(t)$ é o sinal a ser analisado no domínio contínuo, $x(k)$ é a versão discretizada deste sinal no primeiro nível de resolução, e $k \in \mathbb{Z}$.

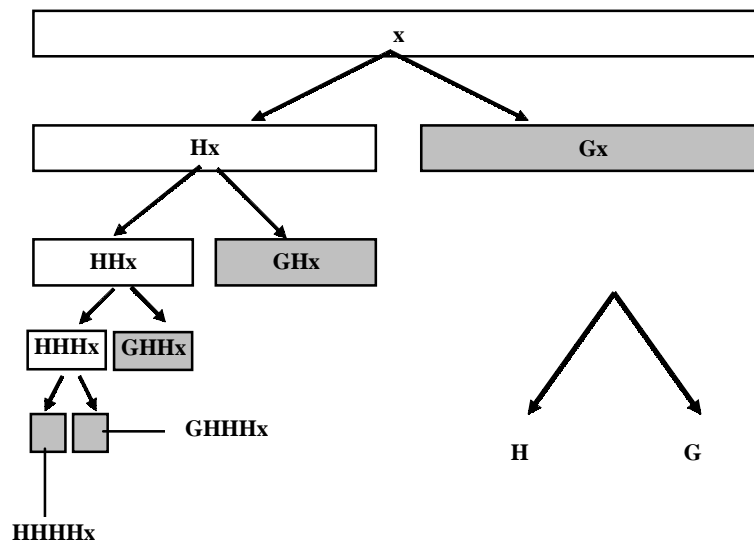


Figura 3-11 Diagrama de Blocos da Transformada Wavelet Discreta (H é o filtro passa-baixa e G o filtro passa-alta).

A decomposição Wavelet Packet é construída com um algoritmo Wavelet recursivo, usando mais filtros G, conforme o esquema da Fig. 3-12.

Aplicações recursivas de H e G das equações (3.40) e (3.41) representam o produto interno do sinal x com outras Wavelets Packets, como no exemplo da equação (3.43).

$$\int_{\mathbb{R}} x(t) \cdot 2^{-3/2} w_5(2^{-3}t - p) dt = G.H.G.x(k) \quad (3.43)$$

Na equação (3.43) o índice de escala é 3 e o índice de frequência é 5.

As áreas retangulares em cinza na Fig. 3-12 constituem um conjunto de bases particular para a representação do sinal x . Este conjunto é uma representação mínima na qual a seqüência x pode ser exatamente reconstruída. Existem outros subconjuntos de bases disponíveis que representam a decomposição completa do sinal original em subbandas [COI98].

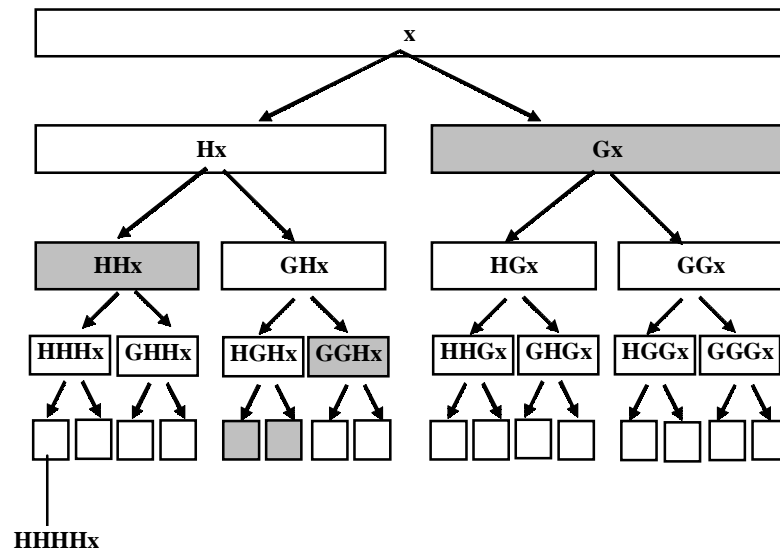


Figura 3-12 Diagrama de Blocos da Transformada Wavelet Packet (H é o filtro passa-baixa e G o filtro passa-alta).

Seja Λ_f um espaço vetorial de translações inteiras de Wavelet Packet w_f para escalas unitárias [WIC94]:

$$\Lambda_f \stackrel{def}{=} \left\{ x(t) = \sum_p \lambda(p) w_f(t-p) \right\} \subset L^2(\mathfrak{R}) \quad (3.44)$$

Como as translações estão relacionadas com os filtros QF, isto implica que:

$$\Lambda_{2f} = H\Lambda_f \quad \text{e} \quad \Lambda_{2f+1} = G\Lambda_f \quad (3.45)$$

Para $f=0$ tem-se $\Lambda_0 = V_0$, é o espaço de aproximação definido pelo filtro passa-baixa H para escala unitária.

Existe uma base natural para cada um destes espaços, ou seja, o conjunto de funções da equação (3.46) forma uma base para o espaço Λ_f :

$$w_f(t-p): p \in \mathbb{Z} \quad (3.46)$$

Se H e G são filtros ortogonais QF, então o conjunto de funções da equação (3.46) é uma base ortonormal para o espaço Λ_f com escala fixa.

Seja σ^s a representação da escala s (ou nível de dilatação s) do espaço Λ_f .

Por definição:

$$\sigma^0 \Lambda_0 = \sigma^1 \Lambda_0 + \sigma^1 \Lambda_1 = \dots = \sigma^s \Lambda_0 + \sigma^s \Lambda_1 + \dots + \sigma^s \Lambda_{2^s-1} \quad (3.47)$$

Note que para $f=0$ e $s=0$, tem-se o sinal original que está sendo analisado. A notação λ_{00} ou λ também é usada para representar o sinal no espaço $\sigma^0 \Lambda_0$.

De uma forma geral:

$$\Lambda_f = \sigma^s \Lambda_{2^s f} + \dots + \sigma^s \Lambda_{2^s(f+1)-1} \quad (3.48)$$

para todo f, $s \geq 0$ e σ^s representa o nível de dilatação s.

A partir da equação (3.36) é calculado o produto interno de $x(k)$ com a Wavelet Packet em qualquer subespaço $\sigma^s \Lambda_f$, para $s > 0$ e $0 \leq f < 2^s$, aplicando os operadores H e G um total de s vezes [WIC94].

As funções Wavelet Packet $\{w_{spf} : p \in Z\}$ são uma base para o espaço $\sigma^s \Lambda_f$ [WIC94].

Uma maneira simples de visualizar os espaços vetoriais $\sigma^s \Lambda_f$ é através de um diagrama de blocos da Transformada Wavelet Packet representando os subespaços conforme a Fig. 3.13.

A representação gráfica de árvore binária da Transformada Wavelet Packet, chamada de Análise Wavelet Packet, encontra-se na Fig. 3-14 a), onde λ_{sf} são os coeficientes de decomposição. Na Fig. 3-14 b) estão representados os espaços $\sigma^s \Lambda_f$ da Wavelet Packet associados aos coeficientes λ_{sf} necessários para reconstruir o sinal, chamado de Síntese Wavelet Packet. Na raiz das árvores, tem-se o sinal original λ_{00} e seu espaço $\sigma^0 \Lambda_0$ associado.

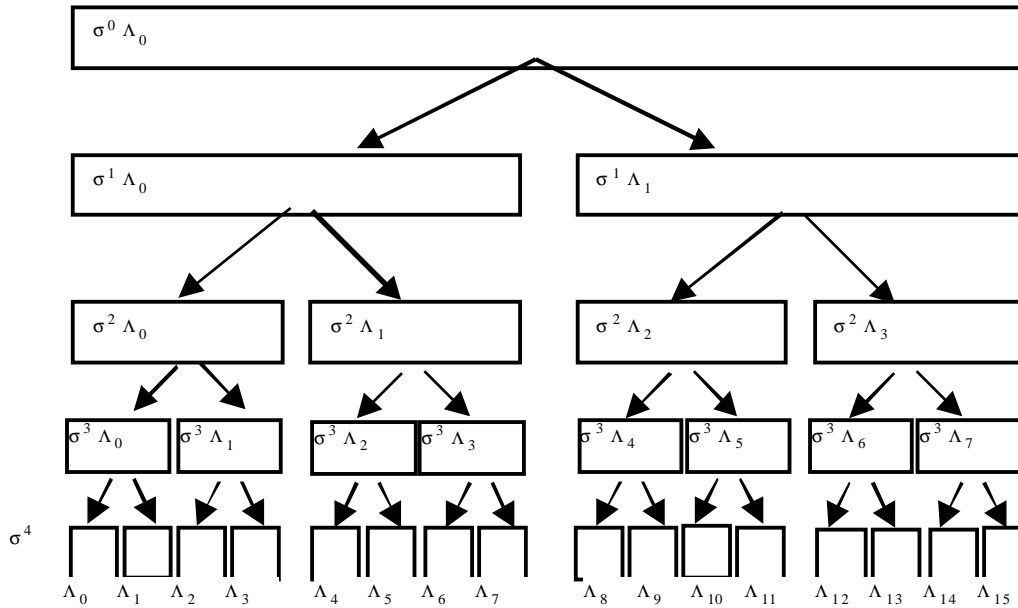


Figura 3-13 Em cada nível de dilatação σ^s tem-se as bases correspondentes aos subespaços Λ_f .

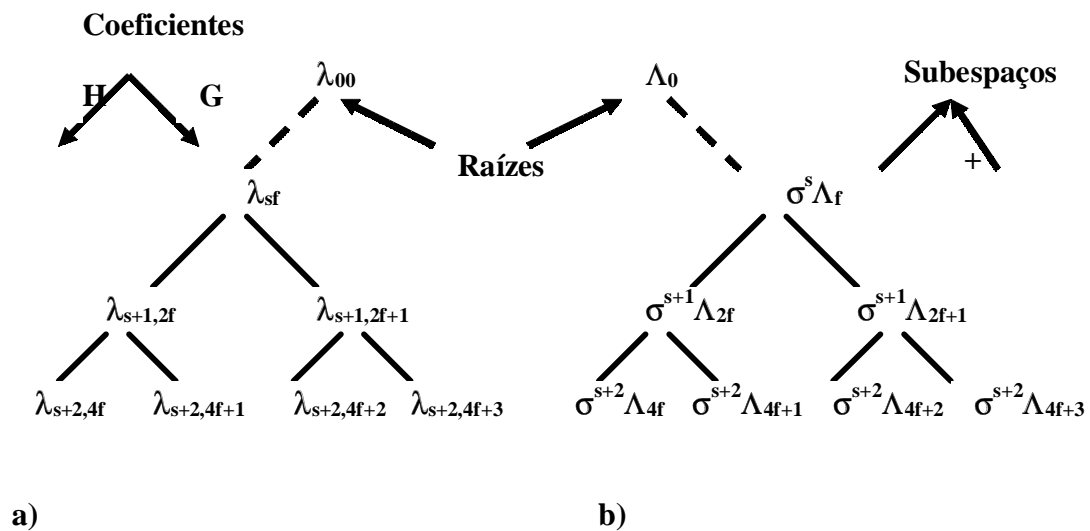


Figura 3-14 Árvore Wavelet Packet a) análise b) Síntese .

A coleção de funções Wavelet Packet w_{sfp} usada no produto interno compreende uma biblioteca de funções, com a organização natural derivada das equações (3.37) e (3.38). Esta coleção forma uma árvore binária cujos nós (ou folhas da árvore) são os espaços $\sigma^s \Lambda_f$. A raiz é $V_0 = \Lambda_0$, os nós são subespaços $\sigma^L \Lambda_0, \dots, \sigma^L \Lambda_{2^L-1}$ como segue na Fig. 3-14 b). Cada nó (espaço) é a soma de seus dois nós descendentes imediatos

(subespaços), ou nós filho. Se os filtros QF são ortogonais, o nó pai é a soma direta de seus nós filhos.

Na Fig. 3-15 encontram-se as Wavelets Packets Haar-Walsh. Cada forma de onda corresponde aos coeficientes em escuro do diagrama de blocos da árvore da Transformada Wavelet Packet diretamente ao lado. Existem 24 formas de ondas para este espaço de dimensão oito. A dimensão oito se refere à escala três ($s=3$) e oito subespaços neste nível de dilatação. Na coluna da esquerda da Fig. 3-15 (forma de onda e a sua localização na árvore WP), tem-se as formas de onda para o primeiro nível de dilatação ($s=1$), gerando as funções w_f com $f=0$ e $f=1$, e 4 deslocamentos da mesma forma de onda (para $p=0..3$). Estas formas de onda são derivadas das equações (3.37) e (3.38) para um mesmo nível de dilatação. Na coluna do meio da Fig. 3-15, tem-se as formas de onda para o segundo nível de dilatação ($s=2$), gerando as funções w_f com $f=0..3$ e 2 deslocamentos da mesma forma de onda (para $p=0$ e $p=1$). Na terceira e última coluna da Fig. 3-15, tem-se as formas de onda para o terceiro nível de dilatação ($s=3$), com $f=0..7$.

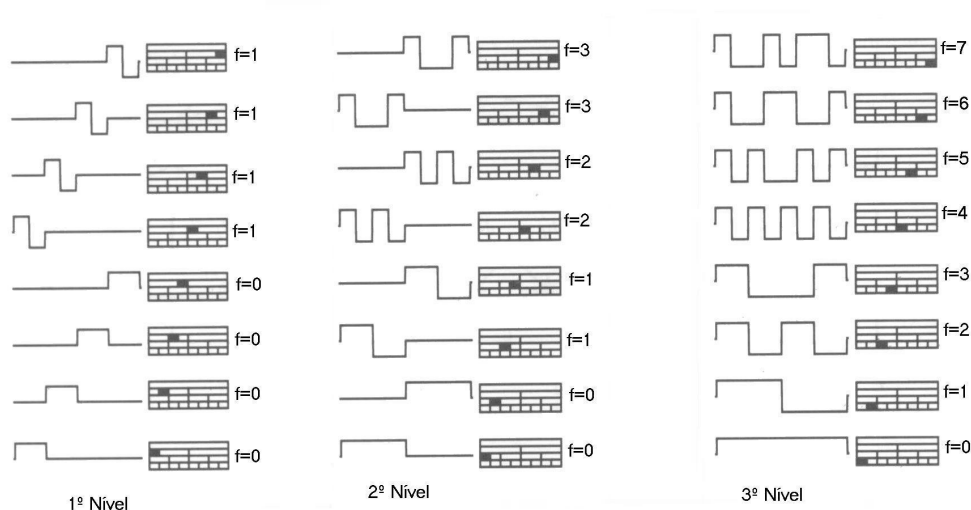


Figura 3-15 Bases Wavelets Packets Haar-Walsh, 1º, 2º e 3º níveis (σ^1) de decomposição (extraído de [WIC94]).

Percebe-se no 3º nível da Fig. 3-15 que o número de oscilações das formas de onda não necessariamente aumentam com o aumento de f . O 3º nível da Fig. 3-15 representa as formas de onda da função Wavelet Packet Haar para o mesmo nível de decomposição. A ordem natural w_f para $f=0,1,\dots,7$, não está de acordo com a idéia de que w_f oscila mais do que w_f se $f > f'$, ou seja, esta ordem não representa uma leitura crescente da frequência nominal ou taxa de oscilação associada as Wavelets Packets.

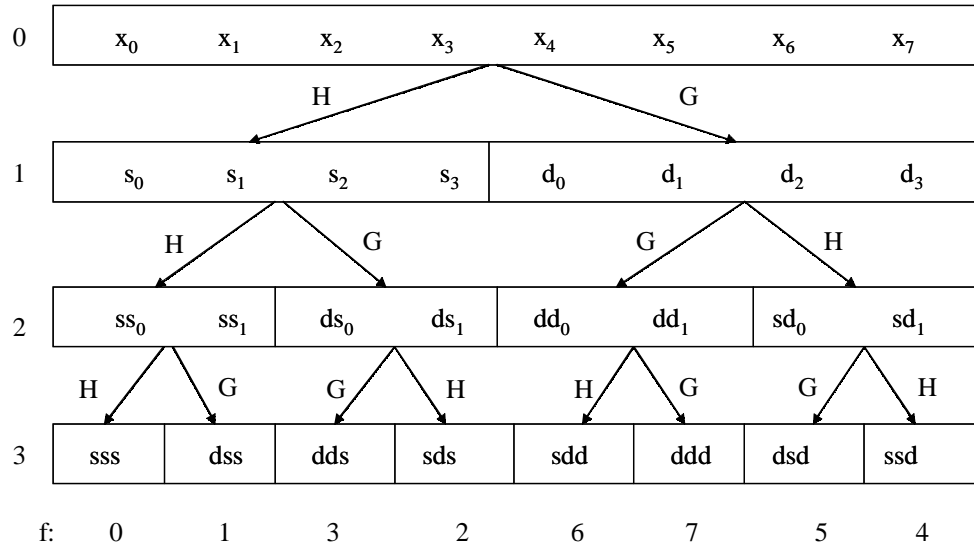


Figura 3-17 Coeficientes Wavelet Packet em ordem crescente de frequência nominal.

A Transformada Wavelet Discreta é uma base ortogonal. A Fig. 3-18 possui um diagrama que corresponde à decomposição $V_0 = \sigma^3 \Lambda_0 \oplus \sigma^3 \Lambda_1 \oplus \sigma^2 \Lambda_1 \oplus \sigma^1 \Lambda_1$ em três níveis, e a área em cinza é um subconjunto da árvore da Wavelet Packet, que corresponde a Wavelet Discreta.

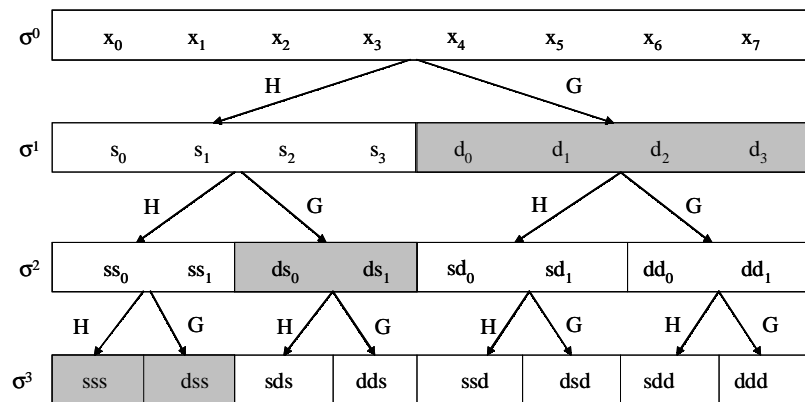


Figura 3-18 Transformada Wavelet Discreta em cinza.

Diferentes escolhas de nós fornecem outros subconjuntos de bases ortogonais. Uma linha de retângulos da Fig. 3-19 corresponde a Wavelet Packet de igual escala para $s=2$. O nível completo corresponde à decomposição nas bases Wavelets $V_0 = \sigma^2 \Lambda_0 \oplus \sigma^2 \Lambda_1 \oplus \sigma^2 \Lambda_2 \oplus \sigma^2 \Lambda_3$.

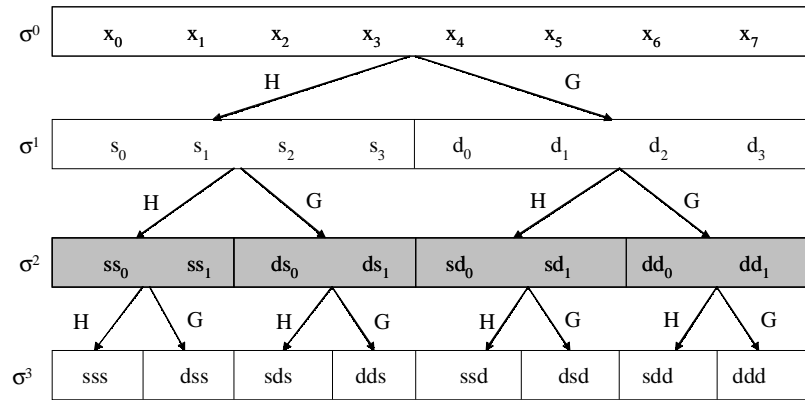


Figura 3-19 Base Wavelet Packet para escolha de um nível fixo (σ^2) em cinza.

Outro possível subconjunto corresponde a base Wavelet $V_0 = \sigma^1 \Lambda_0 \oplus \sigma^2 \Lambda_3 \oplus \sigma^3 \Lambda_4 \oplus \sigma^3 \Lambda_5$, conforme o esquema da Fig. 3-20.

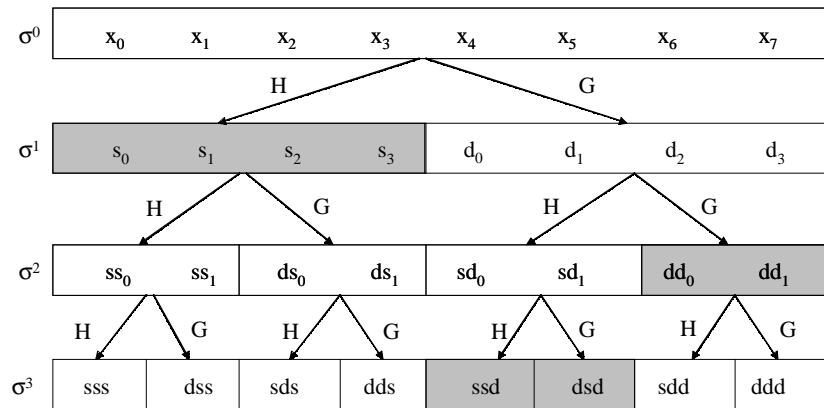


Figura 3-20 Outra possível base Wavelet Packet em cinza.

Uma vez que existem diferentes possíveis bases para a representação de um sinal, é possível escolher qual a melhor entre estas bases conforme algum critério de escolha. A melhor base deve possuir informação substancial sobre o sinal [WIC94].

Antes de se definir qual seria a melhor base (representação ótima), vai-se introduzir o conceito de custo de informação, ou o custo de armazenar a representação escolhida. Seja então uma função custo de informação em seqüências de números reais ou complexos para ser qualquer valor real funcional M satisfazendo a seguinte condição de aditividade:

$$M(u) = \sum_{k \in \mathbb{Z}} \mu(|u(k)|) \quad \mu(0) = 0 \quad (3.49)$$

Sendo β uma coleção de bases e B uma base qualquer, tal que $B \in \beta$, $u(k)$ uma seqüência de coeficientes de $u(k) = \langle b_k, x(k) \rangle = B \cdot x$, onde $b_k \in \beta$ é o k -ésimo vetor da base B , μ uma função definida no intervalo $[0, \infty)$, então o custo de informação de representar um vetor x na base B é $M_x(B \cdot x)$.

A *Best Basis* definida para x (relativa à coleção de bases β e à função custo de informação M) é a base B na qual $M_x(B \cdot x)$ é mínimo [WIC94].

Podem ser definidos diversos tipos de função custo de informação M , mas os mais usados são aqueles que medem concentração. Isto significa que M tem que ser grande quando os elementos da seqüência são basicamente do mesmo tamanho e pequeno quando todos, menos alguns, são negligenciáveis [WIC94].

Alguns exemplos de função custo de informação são:

1. *Threshold*
2. Entropia:
3. Concentração em \mathbb{I}^p :
4. Logaritmo da energia

1. *Threshold*

Este critério envolve escolher um limiar arbitrário ε , fornecendo o número de coeficientes necessários para transmitir ou receber o sinal com precisão do limiar ε .

2. Entropia

Critério de entropia Shannon-Weaver expresso por:

$$\varepsilon_x = - \sum_i \frac{|x_i|^2}{|x|^2} \log_2 \frac{|x_i|^2}{|x|^2}, \text{ onde } |x|^2 = \sum_i |x_i|^2 \quad (3.50)$$

3. Concentração em \mathbb{I}^p

Escolhe-se um valor p arbitrário no intervalo $1 < p < 2$ [WIC94] e determina-se $M(u) = (\|u\|_p)^p$, tal que u pertença ao espaço $l^p(Z)$. Se duas seqüências possuem igual energia $\|u\| = \|v\|$ mas $M(u) < M(v)$, então u possui maior concentração de energia em menor número de elementos.

4. Logaritmo da energia

Seja $M(u) = \sum_{k=1}^N \log|u(k)|^2$. Isto pode ser interpretado como a entropia de um processo de Gauss Markov e que $u(k)$ produz N -vetores cujas coordenadas tem variâncias $\sigma_1^2 = |u(1)|^2, \dots, \sigma_N^2 = |u(N)|^2$. É assumido que não há componentes no processo que não mudam, ou seja, que $\sigma_1^2 \neq 0$ para todos $k=1, \dots, N$. Minimizando $M(u)$ sobre B encontra uma base Karhunen-Loève para o processo.

Algoritmo *Best Basis*

O algoritmo *Best Basis* está baseado na escolha da expansão do sinal na qual o número de coeficientes necessários para se obter 99% da energia [COI98] é menor, ou pode-se escolher a expansão cujo “custo” seja menor. O custo pode ser: custo de informação, custo de codificação, custo de erro, etc...[COI98]. Quando o critério é custo de informação, geralmente utilizam-se as funções custo de informação apresentadas anteriormente.

Sejam H e G os filtros QF e $A_{n,k+1}$ o espaço de vetores no nível $k+1$, onde os sinais são projetados. A operação de filtragem H e G separa o espaço de vetor em 2 subespaços: $A_{2n,k}$ e $A_{2n+1,k}$. A Fig. 3-21 mostra um estágio da decomposição Wavelet.

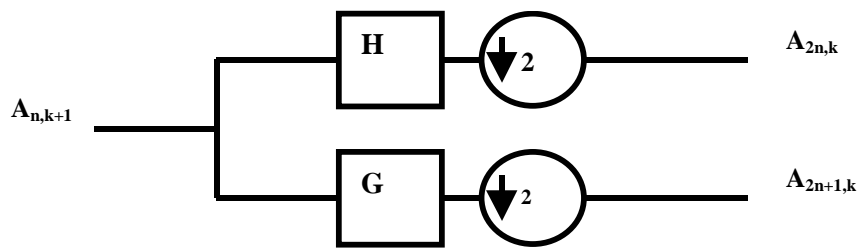


Figura 3-21 Um estágio de decomposição Wavelet ou DWT do sinal.

É mostrado em Wickerhauser [WIC98] que os dois subespaços são ortogonais e $A_{n,k+1}$ é a soma direta deles:

$$A_{n,k+1} = A_{2n,k} \oplus A_{2n+1,k} \quad (3.51)$$

Nas árvores de decomposição Wavelet Packet, conforme a Fig. 3-22, contém diversas bases admissíveis para representar o sinal. É possível extrair a melhor base relativa a uma função custo, conhecido como algoritmo *Best Basis*.

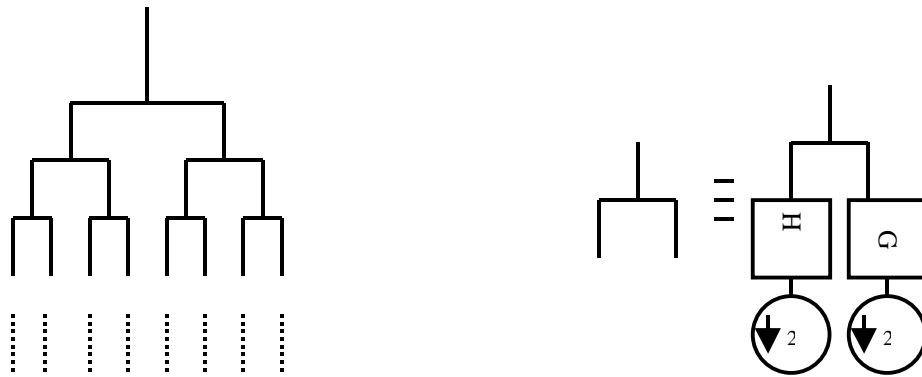


Figura 3-22 Diagrama da árvore de completa obtida pela decomposição.

A base pode ser qualquer sub-árvore da árvore inicial completa e o método *Best Basis* comandará a trajetória de como a árvore será podada. A propriedade da equação (3.51) garante que podando a árvore e obtendo uma sub-árvore, ainda tem-se uma base ortogonal na qual o sinal pode ser decomposto [KAR98].

Para encontrar a *Best Basis*, toda a decomposição da árvore é estendida até o nível de dilatação que se deseja, conforme a Fig. 3-22. No último nível de dilatação, é calculado o custo de informação dos nós filhos, $M(x_a)$ e $M(x_b)$, e calculado o custo de informação do nó pai $M(x)$ conforme a Fig. 3-23.

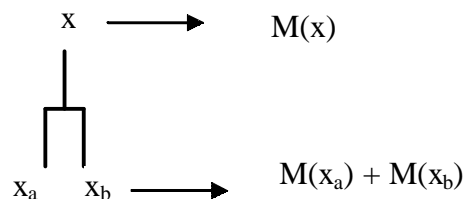


Figura 3-23 Diagrama de um nó pai e seus nós filhos para o cálculo da função custo.

A árvore é podada de baixo para cima indo dos nós filhos à raiz, comparando a menor função custo entre a soma dos nós filhos $M(x_a) + M(x_b)$ com seu nó pai $M(x)$, conforme a ilustração da Fig. 3-24. Os nós que possuem menor custo permanecem, e aqueles que possuem maior custo são podados.

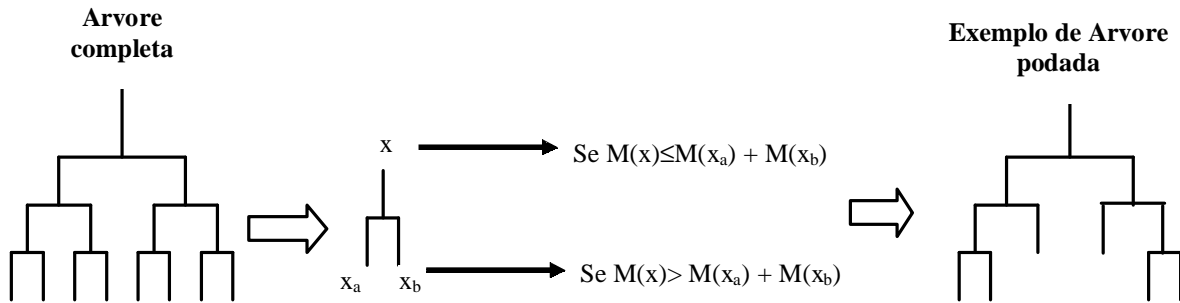


Figura 3-24 Ilustração do princípio de seleção *Best Basis*.

Sejam $A_{2n,k}$ e $A_{2n+1,k}$ as bases no nível k e $B_{n,k+1}$ o nó pai das 2. E seja $M(B.x)$ o valor funcional calculado na projeção do sinal x em alguma base B . A melhor base será:

$$A_{n,k+1} = \begin{cases} B_{n,k+1} & \text{se } M(B_{n,k+1}) \leq M(A_{2n,k}) + M(A_{2n+1,k}) \\ A_{2n,k} \oplus A_{2n+1,k} & \text{se } M(B_{n,k+1}) \geq M(A_{2n,k}) + M(A_{2n+1,k}) \end{cases} \quad (3.52)$$

É atribuído o valor mínimo funcional entre $M(B_{n,k+1})$ e $[M(A_{2n,k}) + M(A_{2n+1,k})]$ para $A_{n,k+1}$, para as próximas comparações.

Isto é feito até atingir o topo da árvore, formando então uma base completa de acordo com a função custo escolhida. A esta árvore resultante dá-se o nome de árvore *Best Basis*.

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 [SCH98].

Na seção 4.3 estão apresentados os testes que foram feitos com objetivo de obter um melhor entendimento da Transformada Wavelet Packet aplicada a sinais de voz. Neste item, são mostrados diversos experimentos utilizando vozes sintéticas, onde se tem o controle dos parâmetros que estão sendo alterados, de forma a avaliar os resultados obtidos aplicando a Transformada Wavelet Packet nestas vozes. Os parâmetros alterados foram o próprio modelo de geração da vogal sintética, as frequências fundamentais dos sinais e por fim, foram inseridos *Jitter* e *Shimmer* nestes sinais. Em todos os sinais sintéticos gerados, foi analisada a melhor função Wavelet Packet para sinais de voz e a melhor função custo, dentro de um conjunto de funções predeterminadas.

Na seção 4.4, aplicaram-se as funções Wavelet Packets e as funções custo escolhidas no item 4.3 a todos os sinais de vozes patológicas do banco de vozes. Nesta seção encontram-se tabelas com o agrupamento de patologias do banco de vozes baseado na acústica e nas similaridades patológicas da voz.

4.2 BANCO DE VOZES PATOLÓGICAS

A descrição do banco de vozes patológicas utilizadas neste trabalho encontra-se na Tabela A.1 do Apêndice A.

As vozes foram gravadas e diagnosticadas no consultório do Dr. Nédio Steffen, no Hospital da PUCRS. O diagnóstico preciso da existência e o tipo de patologia foi obtido pela avaliação dos pacientes através da Videolaringoscopia direta e da Videoestrobolaroscopia. Digitalizou-se toda emissão da vogal sustentada /a/ pelo paciente (vogal cujas estruturas do trato vocal se mantêm mais estáveis [SCH98]) e foi

escolhido o trecho de três segundos que era o mais estável. No total, foram gravadas vozes de 64 pessoas, incluindo normais e patológicos.

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 para um sinal de linha por um circuito baseado em 2 amplificadores operacionais do tipo TL, em cascata, na configuração não-inversor, com alimentação por bateria 9V. O sinal foi então levado a um gravador cassete Technics model RS-B105, 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 do tipo 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) e digital-analógico (D/A), 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:

1. Processador TMS0C25, com *clock* de 40 MHz e ciclo de instruções de 100 ns.
2. Memória de 32 kBytes, permitindo uso conjunto de RAM e EPROM.
3. Barramento para expansão de memória.
4. 2 interfaces de entrada e saída, uma com 8 e outra com 16 bits.
5. 1 interface de comunicação serial, do tipo RS232.
6. 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:

1. Conversor A/D de 14 bits, e capacidade de até 80.000 amostras por segundo (frequência de amostragem, fs), baseado no "chip" AD7871.

2. Duas saídas D/A de 8 bits, baseadas no "chip" PM-7226.
3. Comunicação serial com a placa de processamento de sinais DSP.
4. 3 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 [OPP89]. O processo de aquisição foi feito através da sobre-amostragem do sinal em 3 vezes a frequência de amostragem (f_s), filtragem digital com um filtro do tipo passa-baixas do sinal pela placa DSP e, sub-amostragem (decimação pelo fator 3) do sinal filtrado (de cada 3 amostras, 2 são desprezadas). 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 amostrado em 75 kHz, filtrado digitalmente e então sub-amostrado para $f_s=25\text{kHz}$, obtendo-se uma banda de sinal efetiva de 12,25 kHz. O filtro passa-baixas digital utilizado é um filtro FIR (*Finite Impulse Response*) do tipo janela *Hamming* de 63 pontos, com banda de passagem de $0,16.f_s$ (12kHz) [SCH98]

4.3 TESTE DA FERRAMENTA WAVELET PACKET

Para a aplicação da Transformada Wavelet Packet em sinais de voz e para os algoritmos desenvolvidos neste trabalho, foi utilizado o programa MatLab™ v. 5.3 da MathWorks Inc. e seu *Toolbox* Wavelet.

É desejável que a ferramenta que se está testando como classificador seja insensível a diferentes frequências fundamentais (f_0), devida a variabilidade de f_0 do ser humano, e ao mesmo tempo seja sensível a diferentes formas de ondas que representam as patologias. Estas condições garantem que se possa usar a ferramenta em qualquer sinal de voz independente de sua frequência fundamental, mas que ainda se consiga separar entre normais e patológicos. Com o intuito de se testar estas condições idealizadas, geraram-se vogais sintéticas com uma variabilidade de frequências fundamentais para LPC médio de homens, de mulheres e a média dos dois, chamado de sinal casal. Também foram gerados LPC de sinais patológicos, e foram inseridos *Jitter* e *Shimmer* nos sinais sintéticos de homens e mulheres.

4.3.1 Geração das vogais sintéticas

Para testar a ferramenta Wavelet Packet, foram gerados diversos sinais sintéticos da vogal sustentada /a/. O modelo utilizado para a geração destes sinais é o LPC (*Linear Predicting Code*) [DEL93].

O diagrama de blocos da Fig. 4-1 representa o modelo de geração de sinais vozeados, baseado em LPC:

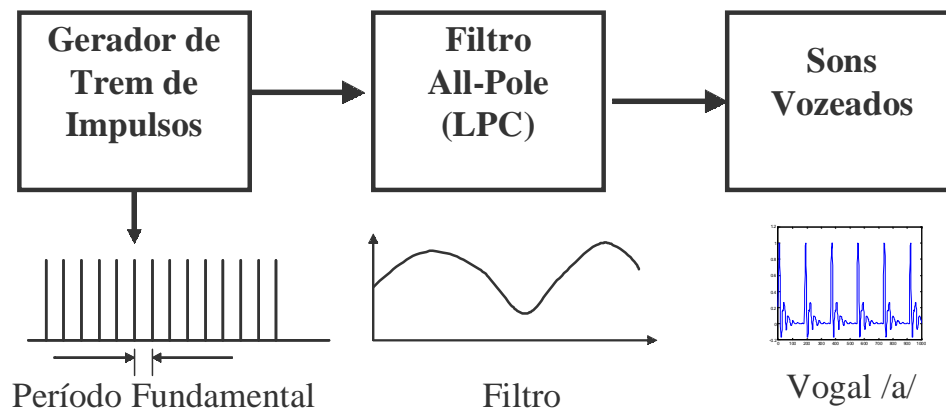


Figura 4-1 Modelo fonte-filtro da geração da vogal sintética.

O sinal de entrada do modelo é um trem de impulsos. A frequência fundamental é gerada variando o período fundamental (T_0) do trem de impulsos, conforme a seguinte equação:

$$T_0 = \frac{1}{f_0} \quad (4.1)$$

onde f_0 é a frequência fundamental desejada.

4.3.2 Vogal Sintética Homem e Mulher

Para a vogal sintética de um homem, foi usada uma frequência fundamental de 120 Hz e para a mulher, de 220 Hz [KEN92].

Para gerar o modelo LPC de homens e mulheres, diversas vozes foram adquiridas no Laboratório de Processamento de Sinais e Imagens (LaPSI). Um grupo de 10

mulheres e um grupo de 10 homens com as cordas vocais supostamente normais (sem nenhuma patologia conhecida) emitiram a vogal sustentada /a/. Os equipamentos utilizados para a aquisição dos sinais foram um microfone modelo SM-58 da Shure, uma mesa de mixagem Mackie 1402 VLZ e os sinais foram gravados em uma fita DAT utilizando um equipamento Tascam DA-30 mk II (LaPSI-UFRGS). Para a conversão A/D os sinais foram reproduzidos no DAT e, via a mesa Mackie, foram injetados na entrada line-in de uma placa de som *on-board* que possui um *chipset* de audio *Crystal CS4236* de um computador Pentium III. A amostragem foi feita com o programa gravador de som do sistema operacional Windows 98, com frequência de amostragem de 22050 Hz e resolução de 16 bits.

Para gerar o LPC das vogais sintéticas foi utilizado o comando 'LPC' do *Toolbox Signal Processing* do MatLab. Para cada voz de mulher do grupo de vozes de mulheres foi calculado o LPC de 22 coeficientes [MAN00b], [PAR01]. Foi então calculado o LPC médio para as mulheres. O mesmo foi feito para os homens, obtendo o LPC médio para os homens. A frequência de amostragem dos sinais sintéticos gerados foi a mesma dos sinais originais que geraram o modelo LPC (22050 Hz).

As formas de onda das vogais sintéticas /a/ para o LPC médio de homem e para o LPC médio de mulher são mostradas na Fig. 4-2.

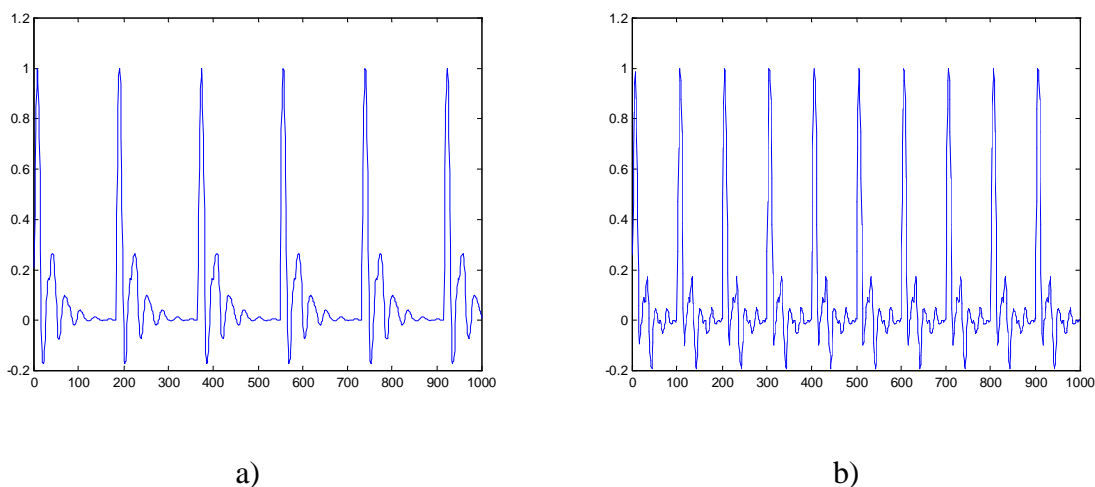


Figura 4-2 Vogal sintética /a/ para a) Homens e $f_0=120\text{Hz}$ e b) Mulheres $f_0=220\text{Hz}$.

Oito sinais sintéticos foram gerados, utilizando quatro diferentes frequências fundamentais para homens e mulheres, conforme a tabela 4-1:

Tabela 4-1 Lista de frequências fundamentais para sinais sintéticos de homem e mulher.

Sexo	Frequências Fundamentais utilizadas (Hz)			
LPC - Homem	80	100	120	140
LPC - Mulher	180	200	220	240

4.3.3 Vogal Sintética Casal

Adicionalmente, fez-se um único LPC médio entre homens e mulheres, gerando um sinal chamado de ‘casal’.

O objetivo deste sinal ‘casal’ é avaliar o comportamento das decomposições em Wavelet Packet de acordo com a variação da frequência fundamental para o mesmo modelo LPC.

Neste sinal variou-se a frequência fundamental de acordo com a tabela 4-2.

Tabela 4-2 Lista de frequências para sinais sintéticos para casal.

Sexo	Frequências Fundamentais utilizadas (Hz)								
LPC Casal	80	100	120	140	160	180	200	220	240

4.3.4 Sinais Sintéticos Patológicos

Com o intuito de avaliar a resposta ou o comportamento da ferramenta Wavelet Packet aplicada a sinais patológicos, geraram-se diversos sinais sintéticos patológicos.

Primeiramente gerou-se o LPC médio para cada uma das quatro patologias:

1. Normais - 9 pacientes
2. Laringite – 6 pacientes
3. Leucoplasia – 3 pacientes
4. Disfonia – 5 pacientes

Como os sinais utilizados para gerar as vogais sintéticas patológicas foram de mulheres, a frequência fundamental do trem de impulsos foi de 220 Hz.

Para as simulações de algumas perturbações de frequência fundamental presente na maior parte das patologias, *Jitter* e *Shimmer* (variação no período da frequência

fundamental e variação na amplitude dos picos do sinal, respectivamente) foram inseridos nos sinais sintéticos.

Para *Jitter*, foram inseridas variações no sinal de entrada (trem de impulsos), introduzindo variações negativas e positivas alternadamente no período dos pulsos como ilustrado na Fig. 4-3.

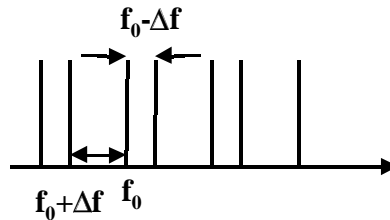


Figura 4-3 Método de inserção de *Jitter* no trem de impulsos.

Para *Shimmer*, variações na amplitude do trem de impulsos foram inseridas adicionando e subtraindo alternadamente valores percentuais na amplitude média do trem de impulsos, conforme a Fig. 4-4.

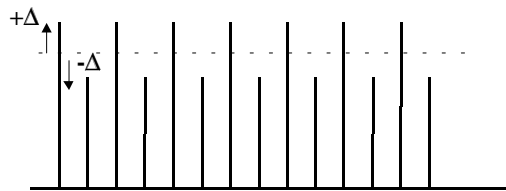


Figura 4-4 Método de inserção de *Shimmer* no trem de impulsos.

Os valores de variação de *Jitter* e *Shimmer* foram escolhidos a partir dos valores médios do banco de vozes, entre normais e patológicas. Os valores da variação para *Jitter* foram 0.5, 1, 2, 5 e 10% e para *Shimmer* foram 5, 10, 15, 20 e 25% .

4.3.5 Funções Wavelet Packet e Função Custo

Para todos os sinais sintéticos foram utilizadas as seguintes funções Wavelet Packets de acordo com a literatura [LIA98], [ENG99], [SOS97]:

1. Haar;
2. Daubechies 6;
3. Daubechies 10;
4. Symlet 5;
5. Symlet 8;
6. Coiflet 2;
7. Coiflet 5.

As funções custo utilizadas neste trabalho foram entropia [LIA98], [ENG99], [SOS97], norma ou concentração em l^p [LIA98], e log da energia. Estas funções custo foram previamente explicadas no capítulo 3 seção 3.3.4. A função custo *threshold* não foi considerada, pois não se tem nenhum limiar prévio que justificasse esta escolha. Já a norma, o valor de p escolhido foi próximo de 2, de forma a se obter a energia do sinal.

4.3.6 Aplicação da Wavelet Packet

Todas as vogais sintéticas geradas nas seções 4.3.2, 4.3.3 e 4.3.4 foram decompostas em árvores Wavelets Packets utilizando as bases e as funções custo do item 4.3.5.

Na Fig. 4-5 está representada a árvore binária completa Wavelet Packet até o quinto nível de decomposição e em cada nó está o índice de posição.

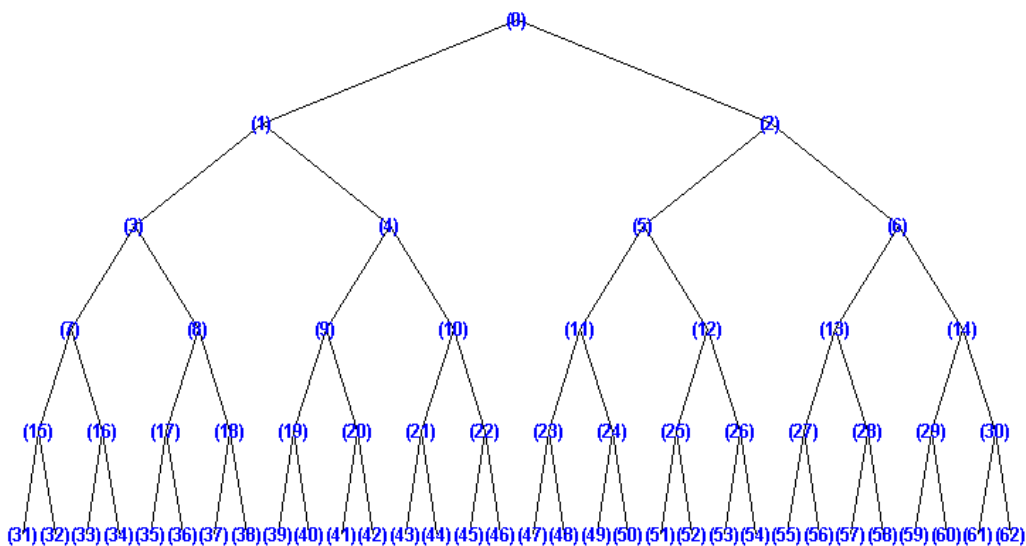


Figura 4-5 Árvore binária Wavelet Packet com o 5º nível de decomposição.

Os resultados utilizando as funções custo e todas as funções Wavelet Packet que estão sendo analisadas nas vogais sintéticas encontram-se no capítulo 5.

4.4 APLICAÇÃO DA TWP NAS VOZES PATOLÓGICAS

Após testar diversas combinações de funções Wavelet Packet e funções custo aplicadas nas vogais sintéticas, aplicou-se as funções WP escolhidas, sym5 e coif2, ao banco de vozes dos pacientes utilizando a função custo escolhida, a entropia. Estas funções foram as escolhidas, pois a sym5 foi a que apresentou a menor variação em termos de tipos diferentes de árvores e a coif2 foi a mais sensível a mudança da frequência fundamental.

Em primeira análise, procurou-se distinguir os sinais de vozes patológicos e normais. Os resultados encontram-se no capítulo 5.

No artigo [ROS98] os autores propõem um agrupamento de patologias baseado na acústica e nas similaridades patológicas da voz ao invés de patologias individuais, devido ao número insuficiente de amostras de algumas patologias para a classificação. Os grupos e as patologias pertencentes a cada grupo encontram-se na tabela 4-3.

Tabela 4-3 Grupos de patologias.

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

A mesma idéia dos autores [ROS98] foi aplicada ao banco de vozes que está descrito na Tabela A.1 do Apêndice A. Os pacientes foram agrupados conforme a Tabela 4-4. 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.

Tabela 4-4 Grupos de patologias para o Banco de vozes descrito na Tabela A.1 do Apêndice A.

Grupo	Pacientes	Total de amostras/ grupo
1	4,9,15,30,32,51,62,63,73	9
2	15,22, 41, 58,59, 62,63,	7
3	10,13,26,27,28,29,30,31,43,46,	10
4	6,23,32,35,36	5
5	1,21,24,26,50, 69,71,	7
6	2,18,19,44,45,47,48,49,55,61,64,68,7	13

A análise dos diversos nós das árvores *Best Basis* para os grupos de patologias encontra-se na tabela 5-19 do capítulo 5 com o intuito de classificar grupos de patologias de acordo com as tabelas 4-3 e 4-4. A tabela 5-19 mostra a ocorrência de nós para cada grupo. Porém, nenhum conjunto de nós foi encontrado que separasse os 6 diferentes grupos de patologias.

5 RESULTADOS

5.1 INTRODUÇÃO

Neste capítulo são apresentados os resultados obtidos neste trabalho na análise de parâmetros discriminantes para classificar sinais patológicos ou normais. Eles são apresentados em forma de Tabelas, descrevendo o número de diferentes *Best Basis* dos sinais sintéticos e sinais patológicos reais para as funções Wavelets Packets e as funções custo utilizadas.

Todas as figuras das *Best Basis* dos sinais sintéticos utilizadas neste trabalho encontram-se em Apêndice B.

5.2 RESULTADOS DA *BEST BASIS* DOS SINAIS SINTÉTICOS

Analisando as *Best Basis* dos sinais sintéticos gerados (homem, mulher e casal) com todas as funções WP e a função custo entropia, gerou-se a tabela 5-1 que mostra quantos tipos diferentes de *Best Basis* surgiram com a variação de frequências fundamentais dentro de cada grupo.

A tabela 5-2 mostra o número de tipos diferentes de árvores com a função custo Log da energia, e a tabela 5-3 mostra o número de tipos diferentes de árvores com a função custo Norma 2.

Analisando as tabelas 5-1, 5-2 e 5-3, as funções custo que resultaram na menor variação de tipos de árvore foram a entropia e a Norma 2. Para estas funções custo, analisou-se a função Wavelet Packet que teve a menor variação de árvore para os sinais sintéticos. Percebeu-se que na análise da *Best Basis* para o grupo de homens, a maioria teve apenas um tipo de árvore. Já para o grupo de mulheres, as árvores tiveram pelo menos 2 variações, com exceção da função sym5, como mostra a tabela 5-1. Nos sinais ‘casais’, as funções que apresentaram apenas uma árvore foi a sym5 e db6. A única função Wavelet Packet que manteve a árvore para uma variação de frequência fundamental na geração dos sinais sintéticos foi a sym5, e a função que mais variou foi a coif2.

Tabela 5-1 Número de tipos diferentes de *Best Basis* ou de árvores com a função custo Entropia.

Sinal / Função	ENTROPIA		
	Homem	Mulher	Casal
Coif2	3	3	3
Coif5	3	2	2
Db10	1	2	2
Db6	1	2	1
Haar	3	2	2
Sym5	1	1	1
Sym8	1	3	2

Nas tabelas 5-2 e 5-3, são mostrados os diferentes tipos de árvores dos mesmos sinais sintéticos da tabela 5-1, com a mesma função, porém com as funções custo Log da energia e Norma 2, respectivamente.

A tabela 5-4 mostra o número de tipos diferentes de árvores para os sinais sintéticos gerados a partir do LPC das patologias para as funções Wavelet Packet.

Analisando as *Best Basis* dos sinais sintéticos de Homem e Mulher com *Jitter* e *Shimmer* e com a função sym5 e entropia, gerou-se a tabela 5-5. Esta tabela apresenta quantos tipos diferentes de *Best Basis* surgiram de um mesmo sinal para diferentes variações de *Jitter* e *Shimmer*.

Tabela 5-2 Número de tipos diferentes de *Best Basis* ou de árvores com a função custo Log Energia.

	LOG ENERGIA		
Sinal / Função	Homem	Mulher	Casal
Coif2	2	3	4
Coif5	4	2	5
Db10	2	4	6
Db6	2	2	4
Haar	4	4	4
Sym5	3	2	4
Sym8	2	3	4

Tabela 5-3 Número de tipos diferentes de *Best Basis* ou de árvores com a função custo Norma 2.

	NORMA 2		
Sinal / Função	Homem	Mulher	Casal
Coif2	4	3	3
Coif5	2	2	3
Db10	1	2	2
Db6	1	2	1
Haar	3	2	2
Sym5	1	1	1
Sym8	1	2	2

Tabela 5-4 Diferentes tipos de árvores de cada função Wavelet Packet para os sinais sintéticos gerados a partir do LPC de quatro patologias.

	Coif2	Coif5	Db10	Db6	Haar	Sym5	Sym8
LPC de 4 Patologias	4	4	3	3	2	4	4

Tabela 5-5 Diferentes tipos de árvores da função sym5 para os sinais sintéticos gerados a partir do LPC de homens e mulheres com variações de *Jitter* e *Shimmer*.

	Variações de <i>Jitter</i>	Variações de <i>Shimmer</i>
LPC Homem	2	1
LPC Mulher	2	1

5.3 RESULTADOS DA *BEST BASIS* NO BANCO DE VOZES

Todos os sinais do banco de vozes patológicas foram decompostos com Wavelet Packet utilizando as funções sym5 e coif2 e a função de custo entropia.

Analisando as árvores visualmente dos sinais do banco de vozes, percebeu-se a presença de alguns nós da árvore Wavelet Packet que ocorriam frequentemente nos sinais patológicos e ao mesmo tempo não ocorriam nas vozes normais. Estes nós que se diferenciavam entre normais e patológicos ocorriam apenas no 5º nível de decomposição, ou seja, entre os nós 31 a 62, conforme a árvore Wavelet Packet da Fig. 4-5.

Para a função sym5 os resultados estão mostrados nas tabelas 5-6 a 5-10. Baseadas nestas tabelas utilizou-se os nós em cinza (41,42,49,50,51,52,55,56) como referência para separar sinal patológico ou não. Ou seja, a presença de um dos nós na *Best Basis* indica que o sujeito é patológico, e a ausência indica normal. Se um sinal é patológico e não possui nenhum dos nós de referência, este é chamado de falso

negativo. Se um sinal é normal e possui um dos nós de referência, este é chamado de falso positivo.

As marcas 'X' nas tabelas indicam a presença dos nós da *Best Basis* para determinado sinal.

Tabela 5-6 Presença dos nós das *Best Basis* do Banco de vozes para sinais Normais com a função Sym5.

Banco de Vozes - Normais Sym5 e Entropia													
No	Pac2	Pac18	Pac19	Pac44	Pac45	Pac47	Pac48	Pac49	Pac55	Pac61	Pac64	Pac68	Pac72
31		X	X						X			X	X
32		X	X						X			X	X
33			X	X	X					X	X	X	X
34			X	X	X					X	X	X	X
35		X	X	X			X	X					
36		X	X	X			X	X					
37			X	X	X		X					X	X
38			X	X	X		X					X	X
39		X				X	X					X	X
40		X				X	X					X	X
41				X									
42				X									
43		X	X	X	X	X		X			X		
44		X	X	X	X	X		X			X		
45			X	X		X							X
46			X	X		X							X
47	X						X						
48	X						X						
49	X												
50	X												
51				X									
52				X									
53							X		X				
54							X		X				
55						X							
56						X							
57				X			X						
58				X			X						
59				X				X		X			
60				X				X		X			
61			X	X									
62			X	X									

Nas tabelas 5-7 a 5-10 apresentam os nós das árvores *Best Basis* para os sinais patológicos do banco de vozes utilizando a função sym5 e entropia.

Tabela 5-7 Presença dos nós das *Best Basis* do Banco de vozes para sinais Patológicos com a função Sym5.

Banco de Vozes - Pacientes Sym5 e Entropia												
No	Pac1	Pac4	Pac5	Pac6	Pac8	Pac9	Pac10	Pac11	Pac13	Pac14	Pac15	Pac16
31			X	X			X		X			X
32			X	X			X		X			X
33	X		X			X	X	X				
34	X		X			X	X	X				
35							X	X	X	X	X	X
36							X	X	X	X	X	X
37		X	X					X	X		X	X
38		X	X					X	X		X	X
39	X		X		X							
40	X		X		X							
41			X	X	X			X	X		X	
42			X	X	X			X	X		X	
43					X			X	X	X	X	
44					X			X	X	X	X	
45		X	X	X		X		X	X		X	
46		X	X	X		X		X	X		X	
47			X		X						X	X
48			X		X						X	X
49		X			X				X		X	X
50		X			X				X		X	X
51	X				X							
52	X				X							
53	X		X		X			X				
54	X		X		X			X				
55	X					X						X
56	X					X						X
57	X			X	X	X	X		X			
58	X			X	X	X	X		X			
59			X		X	X			X			
60			X		X	X			X			
61			X		X	X			X		X	
62			X		X	X			X		X	

Tabela 5-9 continuação da tabela 5-8.

Banco de Vozes - Pacientes Sym5 e Entropia													
No	Pac34	Pac35	Pac36	Pac37	Pac41	Pac42	Pac43	Pac46	Pac50	Pac51	Pac54	Pac56	Pac57
31	X		X			X		X	X			X	
32	X		X			X		X	X			X	
33		X	X			X		X	X			X	X
34		X	X			X		X	X			X	X
35	X		X		X			X	X		X		X
36	X		X		X			X	X		X		X
37	X	X	X	X		X	X	X		X		X	X
38	X	X	X	X		X	X	X		X		X	X
39			X	X	X	X	X	X	X	X	X		X
40			X	X	X	X	X	X	X	X	X		X
41	X		X		X	X						X	
42	X		X		X	X						X	
43	X		X			X		X	X				
44	X		X			X		X	X				
45	X		X	X	X	X	X		X		X		X
46	X		X	X	X	X	X		X		X		X
47	X			X				X	X				
48	X			X				X	X				
49	X			X				X	X	X			
50	X			X				X	X	X			
51	X			X	X	X		X	X				X
52	X			X	X	X		X	X				X
53	X			X		X	X	X					X
54	X			X		X	X	X					X
55	X		X	X	X	X	X	X		X	X		X
56	X		X	X	X	X	X	X		X	X		X
57		X	X	X	X	X				X			
58		X	X	X	X	X				X			
59	X					X		X	X				X
60	X					X		X	X				X
61	X	X		X	X				X	X			
62	X	X		X	X				X	X			

Tabela 5-10 continuação da tabela 5-9.

Banco de Vozes - Pacientes Sym5 e Entropia										
No	Pac58	Pac59	Pac60	Pac62	Pac63	Pac66	Pac69	Pac71	Pac73	Pac74
31		X			X					
32		X			X					
33		X		X		X		X		
34		X		X		X		X		
35		X	X	X	X					
36		X	X	X	X					
37					X				X	X
38					X				X	X
39				X	X			X	X	
40				X	X			X	X	
41		X		X			X			
42		X		X			X			
43				X	X			X	X	
44				X	X			X	X	
45	X			X	X		X	X	X	X
46	X			X	X		X	X	X	X
47							X			
48							X			
49					X		X			
50					X		X			
51		X		X		X				X
52		X		X		X				X
53	X					X	X		X	X
54	X					X	X		X	X
55		X			X	X		X	X	
56		X			X	X		X	X	
57			X		X	X	X	X		X
58			X		X	X	X	X		X
59		X				X				X
60		X				X				X
61						X		X	X	X
62					X		X	X	X	X

Os nós em cinza (41, 42, 49, 50, 51, 52, 55, 56) foram os nós de referência para separar sinal patológico ou não. Para os normais ocorreram 3 falsos negativos e para os patológicos 7 falsos positivos. A Tabela 5-11 mostra o resultado da classificação normal *versus* patológicos.

Tabela 5-11 Classificação de Normais *versus* Patológicos para a função sym5 e entropia.

	Amostras	Percentual de acerto	Percentual de erro
Normais	13	76,92%	3/13 – 23,07%
Patológicos	48	85,42%	7/48 – 14,58%

Tabela 5-13 Presença dos nós das *Best Basis* do Banco de vozes para sinais patológicos com a função Coif2.

Banco de Vozes Patológicas - Coif2 e Entropia													
No	Pac1	Pac4	Pac5	Pac6	Pac8	Pac9	Pac10	Pac11	Pac13	Pac14	Pac15	Pac16	Pac17
31			X	X			X		X			X	
32			X	X			X		X			X	
33	X					X	X	X					
34	X					X	X	X					
35					X		X	X	X	X	X	X	X
36					X		X	X	X	X	X	X	X
37		X	X		X			X	X		X	X	
38		X	X		X			X	X		X	X	
39	X		X		X	X			X			X	
40	X		X		X	X			X			X	
41				X	X			X	X		X		
42				X	X			X	X		X		
43					X			X	X	X	X		X
44					X			X	X	X	X		X
45			X		X			X	X		X		
46			X		X			X	X		X		
47		X	X		X						X		
48		X	X		X						X		
49		X			X			X			X	X	
50		X			X			X			X	X	
51	X								X				X
52	X								X				X
53			X		X			X					X
54			X		X			X					X
55				X	X	X			X				
56				X	X	X			X				
57	X			X	X				X				
58	X			X	X				X				
59			X		X				X				X
60			X		X				X				X
61			X		X				X		X		
62			X		X				X		X		

Tabela 5-14 continuação da Tabela 5-13.

Banco de Vozes Patológicas - Coif2 e Entropia													
No	Pac21	Pac22	Pac23	Pac24	Pac25	Pac26	Pac27	Pac28	Pac29	Pac30	Pac31	Pac32	Pac34
31	X	X			X		X	X		X	X		X
32	X	X			X		X	X		X	X		X
33		X			X	X	X				X		
34		X			X	X	X				X		
35	X	X			X	X	X					X	X
36	X	X			X	X	X					X	X
37		X		X	X			X		X			X
38		X		X	X			X		X			X
39	X	X			X				X		X	X	
40	X	X			X				X		X	X	
41		X	X		X		X		X				X
42		X	X		X		X		X				X
43	X	X	X		X	X				X	X		X
44	X	X	X		X	X				X	X		X
45					X	X			X		X	X	X
46					X	X			X		X	X	X
47	X		X		X	X			X	X	X	X	X
48	X		X		X	X			X	X	X	X	X
49	X		X		X					X			X
50	X		X		X					X			X
51			X		X		X			X		X	X
52			X		X		X			X		X	X
53	X			X	X		X	X		X	X	X	X
54	X			X	X		X	X		X	X	X	X
55	X	X		X	X	X			X	X			X
56	X	X		X	X	X			X	X			X
57		X			X	X			X	X			X
58		X			X	X			X	X			X
59			X	X			X	X		X	X	X	X
60			X	X			X	X		X	X	X	X
61			X	X							X	X	X
62			X	X							X	X	X

Tabela 5-15 continuação da Tabela 5-14.

Banco de Vozes Patológicas - Coif2 e Entropia														
No	Pac35	Pac36	Pac37	Pac41	Pac42	Pac43	Pac46	Pac50	Pac51	Pac54	Pac56	Pac57	Pac58	Pac59
31		X	X		X		X	X						X
32		X	X		X		X	X						X
33	X	X			X		X	X			X	X		X
34	X	X			X		X	X			X	X		X
35		X		X			X	X		X		X		X
36		X		X			X	X		X		X		X
37		X	X	X			X		X				X	
38		X	X	X			X		X				X	
39			X		X	X	X	X	X	X		X		X
40			X		X	X	X	X	X	X		X		X
41				X	X			X						X
42				X	X			X						X
43		X			X		X	X		X				
44		X			X		X	X		X				
45		X		X	X	X				X			X	X
46		X		X	X	X				X			X	X
47	X	X	X				X	X						
48	X	X	X				X	X						
49	X						X	X	X					
50	X						X	X	X					
51		X			X		X	X				X		X
52		X			X		X	X				X		X
53				X		X	X						X	
54				X		X	X						X	
55					X	X	X	X	X					
56					X	X	X	X	X					
57	X			X	X		X		X				X	
58	X			X	X		X		X				X	
59		X			X		X	X				X		X
60		X			X		X	X				X		X
61	X	X	X	X				X	X			X		
62	X	X	X	X				X	X			X		

Tabela 5-16 continuação da Tabela 5-15.

Banco de Vozes Patológicas - Coif2 e Entropia									
	Pac60	Pac62	Pac63	Pac66	Pac67	Pac69	Pac71	Pac73	Pac74
31	X		X		X				
32	X		X		X				
33	X	X		X			X		
34	X	X		X			X		
35	X	X	X						
36	X	X	X						
37		X	X		X			X	X
38		X	X		X			X	X
39		X	X	X			X	X	X
40		X	X	X			X	X	X
41	X	X	X			X			
42	X	X	X			X			
43		X	X				X		
44		X	X				X		
45		X				X	X	X	X
46		X				X	X	X	X
47						X			
48						X			
49			X						
50			X						
51		X		X					X
52		X		X					X
53		X	X	X				X	X
54		X	X	X				X	X
55							X		
56							X		
57							X	X	X
58							X	X	X
59				X		X	X		X
60				X		X	X		X
61						X	X	X	X
62						X	X	X	X

O resultado da classificação normal *versus* patológicos é mostrado na Tabela 5-17

Tabela 5-17 Classificação normais *versus* patológicos para a função coif2 e entropia.

	Amostras	Percentual de acerto	Percentual de erro
Normais	13	69,23%	4/13 – 30,76%
Patológicos	48	81,25%	9/48 – 18,75%

Aplicou-se também a função *sym5* com a função custo norma 2 no banco de vozes. Apenas os resultados da classificação estão apresentados na tabela 5-18. Os nós de referência utilizados foram os (41, 42, 49, 50, 51, 52, 55, 56).

Tabela 5-18 Classificação normais *versus* patológicos para a função *sym5* e Norma 2.

	Amostras	Percentual de acerto	Percentual de erro
Normais	13	69,23%	4/13 – 30,76%
Patológicos	48	83,33%	8/48 – 16,66%

Analisaram-se diversos nós das árvores *Best Basis* na tentativa de classificar grupos de patologias de acordo com as tabelas 4-3 e 4-4 do capítulo 4. Porém, nenhum conjunto de nós foi encontrado que separasse os 6 diferentes grupos de patologias. A tabela 5-19 mostra a ocorrência de nós para cada grupo.

Tabela 5-19 Número de ocorrências dos pares de nós do último nível da árvore Wavelet Packet para os grupos de patologias.

Par de nós	Grupo 1 9 amostras	Grupo 2 7 amostras	Grupo 3 10 amostras	Grupo 4 5 amostras	Grupo 5 8 amostras	Grupo 6 13 amostras
31,32	2	3	8	2	3	5
33,34	2	3	4	2	4	7
35,36	4	6	6	2	4	5
37,38	6	3	6	2	2	6
39,40	4	4	3	1	4	5
41,42	3	4	3	2	3	1
43,44	5	4	5	2	4	7
45,46	7	6	4	3	5	4
47,48	2	1	4	1	4	2
49,50	6	3	4	2	3	1
51,52	1	3	2	1	2	1
53,54	1	1	5	1	4	2
55,56	4	4	4	1	5	1
57,58	3	4	5	3	4	2
59,60	3	1	5	2	2	3
61,62	5	2	2	3	3	2

6 DISCUSSÕES E CONCLUSÕES

6.1 INTRODUÇÃO

Na seção 6.2 deste capítulo são discutidos os resultados obtidos neste trabalho. Também é apresentada a conclusão do uso da Transformada Wavelet Packet como um extrator de características para classificar vozes patológicas ou não com base nas tabelas apresentadas no capítulo 5.

Na seção 6.3 se fez sugestões para trabalhos futuros com o intuito de dar continuidade para a pesquisa de análise e classificação de vozes patológicas utilizando a Transformada Wavelet Packet.

6.2 ANÁLISE DOS RESULTADOS E CONCLUSÕES

Diversos sinais sintéticos foram gerados para teste da ferramenta Transformada Wavelet Packet. Escolheu-se um conjunto de funções Wavelets entre um universo de funções possíveis. Da mesma forma, três funções custo foram escolhidas e justificadas no capítulo 4.

Para escolher a melhor função Wavelet Packet para esta aplicação, aplicou-se a Transformada Wavelet Packet nos sinais sintéticos gerados e avaliaram-se as *Best Basis* que não variavam com a frequência fundamental. As funções custo que tiveram os melhores resultados em relação às *Best Basis* foram a entropia e a Norma 2, conforme as tabelas 5-1 e 5-3. Dentro destas duas opções, avaliou-se a melhor função. Analisaram-se as *Best Basis* que tiveram a menor variação de tipos de árvores para diferentes valores de frequência fundamental. De todas as funções Wavelet Packet escolhidas, as funções que não apresentaram nenhuma variação foram a db6 e a sym5 para sinais sintéticos homens e casais, conforme as Fig. A.1 a A.7 e as Fig. A.15 a A.28. Porém, a única função Wavelet Packet que se manteve constante no grupo de homens, mulheres e casais foi a sym5, de acordo com as Fig. A.8 a A.14. Assim, conclui-se que esta função não apresentou sensibilidade à variação de frequência, pois manteve a mesma árvore *Best Basis* para todos os sinais sintéticos. Já a função Wavelet Packet que mais teve variação foi a coif2.

A seguir, verificou-se se a função Wavelet Packet escolhida era sensível a diferentes patologias. De acordo com as Fig. A.29 a A.35, existe variabilidade dos tipos de árvores *Best Basis* de uma patologia para outra, incluindo a função Wavelet Packet sym5, conforme apresentado na tabela 5-4.

Analisando as *Best Basis* nos sinais sintéticos nos quais foram inseridos *Jitter* e *Shimmer* para a função Wavelet Packet escolhida, conclui-se que os tipos de árvores permaneceram constantes com relação a estes sinais, conforme as Fig. A.36 a A.39 e a tabela 5-5. Porém, olhando o nó 7 das árvores *Best Basis* dos sinais com *Shimmer*, Fig. A.37 e A.39, percebe-se que o valor de entropia diminui à medida que se aumenta a variação de *Shimmer*. Ao se inserir *Shimmer* em um sinal, se está alterando apenas a amplitude do sinal e não a frequência. Assim, como os coeficientes Wavelet representam a correlação entre a forma de onda do sinal e a função que se está analisando, e a forma de onda continua a mesma porém com variação na amplitude, é compreensível que o valor de entropia diminua à medida que o sinal varie a sua amplitude. Já para *Jitter*, era esperado que as árvores não apresentassem variação, pois a função sym5 foi escolhida exatamente por não apresentar sensibilidade à frequência fundamental. Pelas razões mencionadas acima, a função Wavelet Packet sym5 foi a escolhida para analisar os sinais do banco de vozes patológicas, independente da frequência fundamental do paciente.

Após aplicar a função Wavelet Packet sym5 em todo o banco de sinais de vozes patológicas, buscou-se um padrão que discriminasse normais dos patológicos. Analisando as árvores de todos os pacientes, percebeu-se que alguns nós ocorriam mais que outros e ao mesmo tempo não ocorriam com frequência nos normais, que poderiam servir de discriminante. Os resumos destes resultados estão nas tabelas 5-6 a 5-10. Com os nós escolhidos, obteve-se um índice de classificação de 85,42% para os patológicos (7 pacientes falsos negativos para 48 amostras) e 76,92% normais (3 pacientes falsos positivos para 13 amostras).

Assim como se testou a função Wavelet Packet que menos variou em relação as *Best Basis*, utilizou-se a título de experimentação a função que mais variou (coif2) no banco de vozes para medir seu desempenho. Os resultados obtidos estão nas tabelas 5-12 a 5-16. Com os nós escolhidos, obteve-se um índice de classificação para os patológicos de 81,25% (9 pacientes falsos negativos para 48 amostras) e 69,23% para os normais (4 pacientes falsos positivos para 13 amostras).

Os resultados aqui apresentados são comparáveis aos resultados obtidos por outros autores, conforme visto na seção 2.2. Exemplos como em [CHI92], onde os autores separaram normais de patológicos com um índice de acerto de 75,9%. Outro exemplo encontra-se em [ROS98] onde a média de acerto de classificação dos grupos de patologias foi de 62,33% utilizando redes neurais.

Alguns resultados foram superiores como 96,1% em [ACC98]. Porém este resultado foi alcançado com condições restritas pois investiga a existência de uma patologia específica. Já em [ROS99] foram utilizados índices acústicos e uma rede neural artificial para discriminar entre vozes patológicas e normais com um acerto em torno de 100%.

Observando-se a tabela 5-3 nota-se que a função custo Norma 2 teve resultados próximos aos da função custo entropia na análise dos sinais sintéticos. Aplicou-se ao banco de vozes esta função custo com a função sym5 e obtiveram-se os resultados da Tabela 5-18. Os índices de classificação foram de 83,33% para os patológicos e 69,23% para os normais.

Não se encontrou nenhum conjunto de nós que discriminasse os grupos de patologias conforme a Tabela 5-19.

Verificou-se que a Transformada Wavelet Packet é uma ferramenta capaz de extrair parâmetros com o intuito de separar patológicos de normais, conforme os índices de classificação alcançados. Comparando os resultados obtidos com os relatados na literatura, apresentados na seção 2.2, os índices apresentados neste trabalho utilizando um classificador linear apresentaram bom desempenho. A melhor função Wavelet Packet para isso foi a sym5 e a entropia, concordando com os resultados do artigo [ENG99]. Conclui-se portanto que a Transformada Wavelet Packet é uma ferramenta promissora para auxiliar o diagnóstico de patologias nas pregas vocais, assim como acompanhar o desenvolvimento do tratamento de doenças.

6.3 SUGESTÕES PARA TRABALHOS FUTUROS

Apesar de não ter se encontrado nenhum conjunto de nós que discriminasse os grupos de patologias, sugere-se para um trabalho futuro o uso dos resultados obtidos, utilizando os valores de entropia dos nós em um classificador não-linear como uma rede neural, por exemplo, anteriormente utilizada na tentativa de classificação de patologias como visto na literatura na seção 2.3.

Neste trabalho não foi possível testar a possibilidade de usar um classificador de patologias devido ao baixo número de amostras de cada grupo de patologias para treinar, testar e validar a classificação. Para isso, seria preciso um banco de pacientes com um número maior de amostras, ficando também como sugestão de trabalho futuro aumentar o banco de vozes.

Com um banco de vozes maior, poderia se testar diferentes funções custo como a *Local Discriminant Basis* (LDB) proposta por [SAI94] para melhorar desempenho dos resultados obtidos.

REFERÊNCIAS

- [AKA98] Akay M. “Time Frequency and Wavelets in Biomedical Signal Processing”. IEEE Press, 1998.
- [ACC98] Accardo A. P. and 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, Vol. 28, Issue 1, Elsevier Pages 75-89, January 1998.
- [ARC00] - Arce G. R., *Fellow* , IEEE and Hasan S. R. “Elimination of Interference Terms of the Discrete Wigner Distribution Using Nonlinear Filtering”. IEEE Transactions On Signal Processing, vol. 48, NO 8, August 2000.
- [BAK87] - Baken R. J. “Clinical Measurement of Speech and Voice”. Allyn and Bacon, a division of Simon and Schuster, Inc., Needham Heights, MA. 1987.
- [BAL99] - Balmelli L. and Mojsilovic A. “Wavelet domain features for textures description, classification and replicability analysis”. Proceedings of IEEE International Conference on Image Processing (ICIP), vol. 4, pp 440-444, october 1999.
- [BEN86] – Bendat J. S. and Piersol A. G.: Random Data: Analysis and measurement procedures, 2e. John Wiley & Sons, Inc., 1986.
- [BEN94] - Bentley P. M. and McDonnell. “Wavelet transforms: an introduction”, Eletronics & Communication engineering journal, August 1994.
- [BEN97] - Benediktsson J. A., Sveinsson J. R., Ersoy O. K., Swain P. H. “Parallel consensual neural networks”, IEEE Transactions on Neural Networks, Vol. 8, Issue 1, Pages 54-64, January 1997.
- [BOO94] - Daniel R., Boone S., McFarlane C. “A voz e a terapia vocal”. Artes médicas, Porto Alegre, 1994.
- [CHA93] - Chang T. and Kuo C. J. “Texture analysis and classification with tree-structured wavelet transform”, IEEE Transactions on Image Processing, Vol. 2, No. 4, pp.429-441, Oct. 1993.
- [CHA98] - Chang, Sungwook; Kwon, Y.; Yang, Sung-il. “Speech feature extracted from adaptive wavelet for speech recognition”, Electronics Letters, Vol. 34, Issue 23, Pages 2211-2213, 12 November 1998.
- [CHI92] Childers D. G., Bae K. S. “Detection of Laryngeal Function Using Speech and Electroglottographic Data”, IEEE Transactions on Biomedical Engineering, Vol. 39, Issue 1, Pages 19-25, January 1992.

- [COI92] Coifman R. R. and Wickerhauser M. V. "Entropy-based algorithms for best basis selection." IEEE Transaction on Information Theory, 38(2):1713--1716, March 1992.
- [COI98] – Coifman R. R. and Wickerhauser M. V. "Experiments with Adapted Wavelet De-noising for medical signals and Images", pp 323-346, In Metin Akay [AKA98].
- [COL96] – Colton R. H., Casper J. K., "Compreendendo os problemas de voz". Artes médicas, Porto Alegre, 1996.
- [DAU90]- Daubechies I. "The Wavelet Transform, Time-Frequency Localization and signal Analysis". IEEE Transaction on. Information Theory, vol. 36., 1990.
- [DEL80] - Deller Jr., J. R. and Anderson, D. J. "Automatic Classification of Laryngeal Dysfunction using the Roots of the Digital Inverse Filter, ". IEEE Trans. on Biom. Eng. , Vol.BME 27, pp.714-721, 1980.
- [DEL93] Deller Jr., P. R.; Proakis, J. G. and Hansen, J. H. L. Discrete-Time Processing of Speech Signals. Macmillan Publishing Company, New York, NY, USA. 1993.
- [ENG99] - K. Englehart, B. Hudgins, P. A. Parker and M. Stevenson. "Classification of the myoelectric signal using time-frequency based representations", Elsevier: Medical Engineering & Physics, Vol. 21, Issues 6-7, Pages 431-438, July 1999.
- [FRO98] Frohlich M., Michaelis D., Werner S. H. "Acoustic 'breathiness measures' in the description of pathologic voices", Proceedings of the 1998 IEEE International Conference on Acoustics Speech and Signal Processing, Vol.: 2 , Page(s): 937 -940 vol.2. 1998.
- [HAD00]- Hadjitodorov S., Boyanov B., Teston B. "Laryngeal pathology detection by means of class-specific neural maps", IEEE Transactions on Information Technology in Biomedicine, Vol. 4 Issue: 1 , Page(s): 68 -73, March 2000.
- [GAR00] - Garcia C., Zikos G. and Tziritas G. "Wavelet packet analysis for face recognition", Image and Vision Computing, Vol. 18, Issue 4, Pages 289-297, March 2000.
- [GUO89] Guoxin K., Jiarui L., Qiang Y., Guoying C. "A method of spectra-analysis for abnormal voice signals ", Engineering in Medicine and Biology Society, 1989. Images of the Twenty-First Century., Proceedings of the Annual International Conference of the IEEE Engineering in , 1989.
- [HOM98] Homingo J. and Cristóbal G. "High Resolution Spectral Analysis of images Using the Pseudo-Wigner Distribution." IEEE Transactions On Signal Processing, vol. 46, NO 6, pg 1757-1762, june 1998.

- [HSI00] – Hsin, Hsi-Chin. “Texture segmentation using modulated wavelet transform” IEEE Transactions on Image Processing, Vol. 9, Issue 7, Pages 1299-1302, , July 2000.
- [INT97] –Intrator N., Wong Y. S., Huynh Q. Q., Lee B. H. L. “Wavelet Feature Extraction for Discrimination tasks”, Proceedings ofThe 1997 Canadian Workshop on information Theory. Toronto, June 1997.
- [JIA99] – Jiang Q., Goh S. S. and Lin Z. “Local discriminant time-frequency atoms for signal classification”, Signal Processing, Vol. 72, Issue 1, Elsevier Pages 47-52, 4 January1999.
- [JO01] - Jo C., Kim K., Kim D., Wang S. “Screening of Pathological Voice from ARS using Neural Networks”, proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications, Firenze-Italy, september 2001.
- [KAD92] Kadambe S. and Boudreaux-Bartels G. F. "Application of the wavelet transform for *Pitch* detection of speech signals," IEEE Trans. on Inf. Theory, Vol. 38, pp. 917-924, March 1992.
- [KAR98] –Karrakchou M., Kunt M. “From Continous Wavelet Transform to Wavelet Packet: Application to the Estimation of Pulmonary Microvascular Pressure”, pp367-387, In Metin Akay [AKA98].
- [KEN92] Kent R. D. and Read C. “The Acoustics Analysis of Speech”, Singular Publishing Group, Inc., San Diego, CA.1992.
- [KES99] - Keshava N., Moura J. M.F. “Matching wavelet packets to Gaussian random processes”, IEEE Transactions on Signal Processing, Vol. 47, Issue 6, Pages 1604-1614, 1999.
- [KUH82] – Kuhl I. “Manual Prático de Laringologia”. Editora da Universidade, Porto Alegre 1982.
- [LIA98] Liang H., Nartimo I. “A feature extraction algorithm based on wavelet packet decomposition for heart sound signals”. Time-Frequency and Time-Scale Analysis. Proceedings of the IEEE-SP International Symposium on , Page(s): 93 –96, 1998.
- [MAL89] Mallat S. G. “A theory for multiresolution signal decomposition: The wavelet representation”. IEEE Trans Pat Anal Mach Intell, 11:674--693, July 1989.
- [MAL93] Mallat S. G. and Zhang Z. “Matching pursuits with time-frequency dictionaries.” IEEE Transactions on Signal Processing, 41(12), December 1993.
- [MAL97] Mallet Y., Coomans D., Kautsky J. and De Vel O. “Classification Using Adaptive Wavelets for future extraction”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 10, pp. 1058-1066, October 1997.

- [MAN00a] Manfredi C., D'Aniello M., Brusciaglioni P. and Ismaelli A. "A comparative analysis of fundamental frequency estimation methods with application to pathological voices", *Medical Engineering & Physics*, Vol. 22, Issue 2, Pages 135-147, March 2000.
- [MAN00b] Manfredi C. "Adaptive noise energy estimation in pathological speech signals", *IEEE Transactions on Biomedical Engineering*, Vol. 47, Issue 11, Pages 1538-1543, November 2000.
- [MAR00] -Martinez C. E., Rufiner H. L. "Acoustic Analysis of speech for detection of Laryngeal Pathologies", *Information Technology Applications in Biomedicine. Proceedings. 2000 IEEE EMBS International Conference on*, 2000.
- [MEY00] Meyer F. G., Member IEEE, Averbuch A. Z., Stromberg J. "Fast Adaptive Wavelet Packet Image Compression", *IEEE Transactions on Image Processing*, Vol. 9, No. 5, pp. 1-9, May 2000.
- [NIE01] Niedzielska G., Glijer E. and Niedzielski A. "Acoustic analysis of voice in children with noduli vocales, *International Journal of Pediatric Otorhinolaryngology*", Vol. 60, Issue 2, Elsevier Pages 119-122, 20 August 2001.
- [OBA98] Obaidat M. S., Brodzik A. and Sadoun B. "A performance evaluation study of four wavelet algorithms for the *Pitch* period estimation of speech signals", *Information Sciences*, Vol. 112, Issues 1-4, Pages 213-221, December 1998.
- [OBA99] Obaidat M. S., Lee C., Sadoun B. and Nelson D. "Estimation of *Pitch* period of speech signal using a new dyadic wavelet algorithm, *Information Sciences*", Vol. 119, Issues 1-2, Pages 21-39, 1 October 1999.
- [OPP89] Oppenheim, V. A.; Schaffer W. R. "Discrete-time signal processing". Prentice-Hall signal processing series 1989.
- [PAY97] Paya B. A., Esat I. I. and Badi M. N. M. "Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor". *Mechanical Systems and Signal Processing*, 11(5):751--765, Sept 1997.
- [PAR01] Parraga A., Zaro M. A., Schuck Jr. A. "Quantitative Assessment of the Use of Continuous Wavelet Transform in the Analysis of the Fundamental Frequency Disturb of the Synthetic Voice", *proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications, Firenze-Italy, september 2001.*
- [RAM93] Ramchandran K. and Vetterli M. "Best wavelet packet bases in a rate-distortion sense", *IEEE Transactions on Image Processing*, Vol. 2, No. 2, pp. 160-175, April 1993.

- [RIT01] – Ritchings RT., McGillion M., Moore CJ. “Pathological voice quality assesment using artificial neural networks”, proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications, Firenze-Italy, september 2001.
- [RON75] Rontal, E., Rontal, M. and Rolnick, M. I. “Objective Evaluation of Vocal Pathology using Voice Spectrography,”. Ann. Otol. Rhinol. Laryngol., Vol.84, pp.662-671. 1975.
- [ROS98] Rosa M. de O., Pereira C. J. and C.P.L.F. Carvalho A. “Evaluation of Neural Classifiers using Statistic Methods for Identification of Laryngeal Pathologies”. Proceedings of the Vth Brazilian Symposium on Neural Networks , 1998.
- [ROS99] Rosa M. de O., Greller M., Carvalho C.P.L.F. A. “Signal processing and statistical procedures to identify laryngeal pathologies”, in Electronics, Circuits and Systems, 1999 6th IEEE International Conference on IEEE Circuits and Systems Society, 1999.
- [ROS00] Rosa M. de O., Pereira M. J. C., Grellet M. “Adaptive estimation of residue signal for voice pathology diagnosis”, IEEE Transactions on Biomedical Engineering, Vol. 47, Issue 1, Pages 96-104, January 2000.
- [SAI94] Saito N. and Coifman R. R. “Local discriminant bases”, A. F. Laine and A. M. Unser, ed., Mathematical Imaging: Wavelet Applications in Signal and Image Processing, Proc. SPIE, Vol. 2303, Jul. 1994.
- [SAR00] Sarikaya R., Hansen J. H.L. “High resolution speech feature parametrization for monophone-based stressed speech recognition”, IEEE Signal Processing Letters, Vol. 7, Issue 7,Pages 182-185, 2000.
- [SCH98] Schuck Jr, A. “Análise e processamento de sinais de voz disfônica através da transformada wavelet discreta”, Tese para obtenção do título de Doutor em engenharia, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais – PPGEM, 1998.
- [SHE96] Sheng Y. “Wavelet Transform”. The Transforms and Applications Handbook, POULARIKAS, A. D.(editor-in-chief), CRC Press in cooperation with IEEE press, Boca Raton, 1996.
- [SHI00] Shirazi M.N., Noda H. and Takao N. “Texture classification based on Markov modeling in wavelet feature space,” Image and Vision Computing, Vol. 18, Issue 12, Pages 967-973, September 2000.

- [SIM93] Simpson, D. M. "An Introduction to the Discrete Orthogonal Wavelet Transform," *Revista Brasileira de Engenharia, Caderno de engenharia Biomédica*, vol. RBE 9 (1), pp.57-81. 1993.
- [SOS97] Šoštarič A., Zazula D., Doncarli C. "Feature extraction using wavelet packets," In: B. ZAJC (ed.), *Proceedings of the Sixth Electrotechnical and Computer Science Conference ERK '97, Portorož, Slovenia*. Ljubljana: IEEE Region 8, Slovenian IEEE Section, 1997, pp. A/159-162. Sept. 1997.
- [TAD01] Tadeusiewicz R., Wszolek W., Izvorski A., Wszolek T. "Understanding of Vocal tract pathology using speech signals analysis", *proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications, Firenze-Italy, september of 2001*.
- [TAS96] Taswell C. "Satisfying search algorithms for selecting near-best bases in adaptive tree-structured wavelet transforms", *IEEE Transactions on Signal Processing*, Vol. 44, Issue 10, Pages 2423-2438, October 1996.
- [RIO91] Rioul O. and Vetterli M. 'Wavelet and Signal processing', *IEEE Signal Processing Magazine*, Vol. 8 (4), October 1991.
- [XIA98] - Xiaoli L. and Zhejun Y. "Tool wear monitoring with wavelet packet transform-fuzzy clustering method", *Wear*, Vol. 219, Issue 2, Pages 145-154, September 1998.
- [WIC94] Wickerhauser M. V. "Adapted Wavelet Packet Analysis from Theory to software" . Wellesley, MA: AK Peters, Ltd., 9 May 1994.
- [WU96] Y. Wu and Du R. "Feature extraction and assessment using wavelet packets for monitoring of machining processes". *Mechanical Systems and Signal Processing*, 10(1):29--53, Jan. 1996.
- [YEN00] Yen G. G., Kuo-Chung L. "Wavelet packet feature extraction for vibration monitoring", *IEEE Transactions on Industrial Electronics*, Vol. 47, Issue 3, 2000, Pages 650-667

APÊNDICES

APÊNDICE A

Tabela A.1 – Descrição do banco de vozes utilizado.

Pacientes	Sexo	Descrição	Patologia
PAC01	F	Professora de canto, voz normal	micronódulo nas pregas vocais
PAC02	M	Cantor nativista, vibrato na emissão, voz normal	Pregas normais
PAC04	M	cantor, voz muito rouca	Laringite crônica
PAC05	F	voz soprosa	Sulco vocálico
PAC06	F	voz soprosa.	Edema pós-cirúrgico
PAC08	M	Rapaz pré-cirúrgico	Voz supra-glótica (ventricular)
PAC09	F	Paciente 25 dias sem falar	Pós-laringite por abuso vocal
PAC10	M	Paciente de 78 anos, voz quase normal	Paralisia unilateral bem compensada
PAC11	M	Paciente pós-operatório de 6 anos, feita neocorda	Laringectomia fronto-lateral
PAC13	F	voz mal-compensada	Paralisia uni-lateral
PAC14	F	Paciente pós-operatória de cisto, em tratamento fonoterápico	Sulco vocálico
PAC15	F	fumante	Leucoplasia+laringite crônica
PAC16	F		Rinite e Sinusite alérgica
PAC17	M	Paciente de 66 anos, pós operatório+radioterapia	CA na laringe, lesão prega vocal direita.
PAC18	F	Paciente fonastênica, cansaço na voz	Pregas normais
PAC19	F	Paciente voz normal	Pregas normais
PAC21	F	Paciente de 16 anos, voz rouca por secreção na traquéia	Cisto pequeno na prega vocal direita

PAC22	F	pós-operatório de 3 semanas	Leucoplasia
PAC23	F	fumante, voz rouca e grave	Edema de Reike
PAC24	F	voz muito soprosa e rouca	Cisto, uma prega vocal inchada e sem ondas
PAC25	M	pós operatório 45 laringectomia parcial	Laringectomia fronto-lateral
PAC26	F	voz limpa, trat. fonoterápico	Disfonia funcional+micro-nódulo
PAC27	F	voz limpa mas levemente soprosa, em tratamento fonoterápico	Disfonia funcional
PAC28	F	voz limpa, trat. fonoterápico	Disfonia funcional
PAC29	F	voz limpa, fina e levemente rouca, trat. fonoterápico	Disfonia funcional
PAC30	F	voz grossa e rouca	Disfonia funcional+laringite
PAC31	M	voz alterada	Paralisia prega vocal direita
PAC32	F	voz grossa e rouca, fumante	Edema de Reike+laringite crônica
PAC34	M	voz com sibilização	Papilomas
PAC35		Senhora c/voz rouca e grossa	Edema de Reike-pregas frouxas
PAC36	F		Edema de Reike+formações polipóides
PAC37	F	voz levemente soprosa, pós-cirúrgica de papiloma recorrente	Pregas vocais limpas
PAC41	M	Paciente meia idade, voz rouca	neoplasia na laringe
PAC42	M	Paciente pós-cirúrgico de muito tempo, revisão. Voz muito rouca	Laringectomia fronto-lateral
PAC43	M	Paciente 59 anos	Paralisia unilateral prega esquerda
PAC44	M	voz normal	pregas vocais normais
PAC45	F	voz normal	pregas vocais normais
PAC46	F	Paciente com voz muito soprosa, em	Paralisia prega vocal direita

		fonoterapia	
PAC47	M	Paciente c/voz normal,	pregas vocais normais
PAC48	F	voz normal	pregas vocais normais
PAC49	M	voz normal	pregas vocais normais
PAC50	F		micronódulo nas pregas vocais
PAC51	F		Laringite + Uso Aerolyn
PAC53	F	voz rouca e soprosa	Em teste
PAC54	M	voz normal (talvez alteração ressonância)	Linfonodo no pescoço
PAC55	M	Cantor nativista, Padrão Fry, voz normal	pregas vocais normais
PAC56	F	voz normal, língua presa, <i>Jitter</i> e <i>Shimmer</i> visualmente baixos	Prega vocal direita vascularizada
PAC57	M	voz muito ruim	Laringectomia fronto-lateral
PAC58	M	Paciente com quebra de fonação, voz rouca	Granuloma abaixo prega vocal direita
PAC59	M	Cantor nativista, voz rouca e cansada	Leucoplasia+Irritação
PAC60	M	voz rouca	Laringectomia fronto-lateral
PAC61	F	voz normal, nervosa	pregas vocais normais
PAC62	M	Paciente fumante, pouco tempo de fonação	Leucoplasia+ Laringite crônica
PAC63	F	voz horrível	Leucoplasia+ Laringite crônica
PAC64	F	voz normal, nervosa	pregas vocais normais
PAC66	M	Paciente idoso	Presbifonia
PAC67	F	ex-fumante	Secreção nas pregas
PAC68	F	voz normal	pregas vocais normais
PAC69	M	Paciente com diplofonia	Pólipo grande na prega vocal

PAC70	M	voz normal, pós-cirurgico de tireóide à Laser	Pregas vocais limpas
PAC71	F	Criança	Nódulo bilateral
PAC72	F	Paciente pós-operatória'3 dias	pregas vocais normais
PAC73	M	voz áspera	Ceratoze+Laringite crônica
PAC74	F	Adolescente, voz levemente soprosa.	Fenda Posterior aumentada

APÊNDICE B

Best Basis das Vogais Sintéticas Homens

Nas figuras a seguir, encontram-se *as Best Basis* do LPC de Homens com as funções Wavelet escolhidas e com a variação de frequência fundamental de 80Hz a 140Hz, conforme o item 4.3.2. Em cada nó da árvore está o valor de entropia.

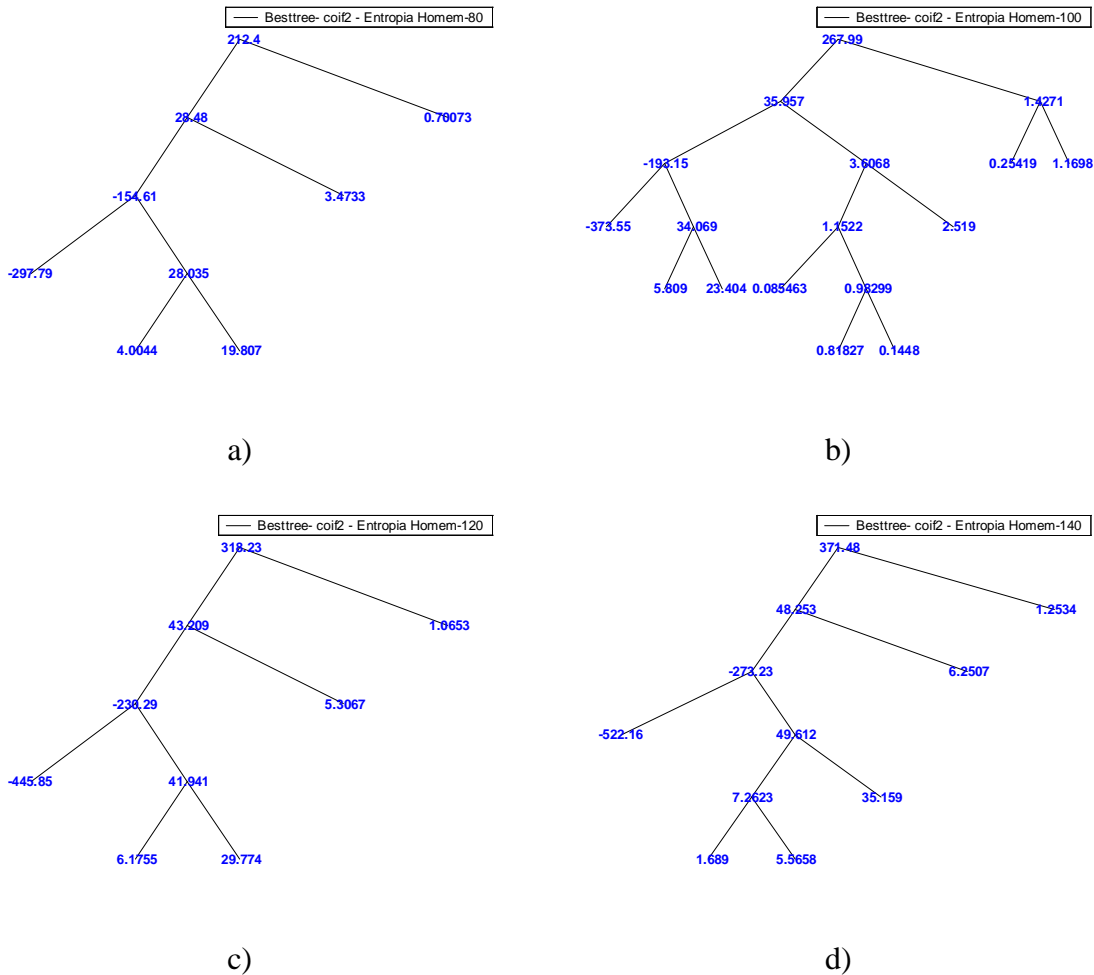


Figura A. 1 *Best Basis* do Homem com a função coif2 para a) $f_0=80$ Hz b) $f_0=100$ Hz c) $f_0=120$ Hz d) $f_0=140$ Hz.

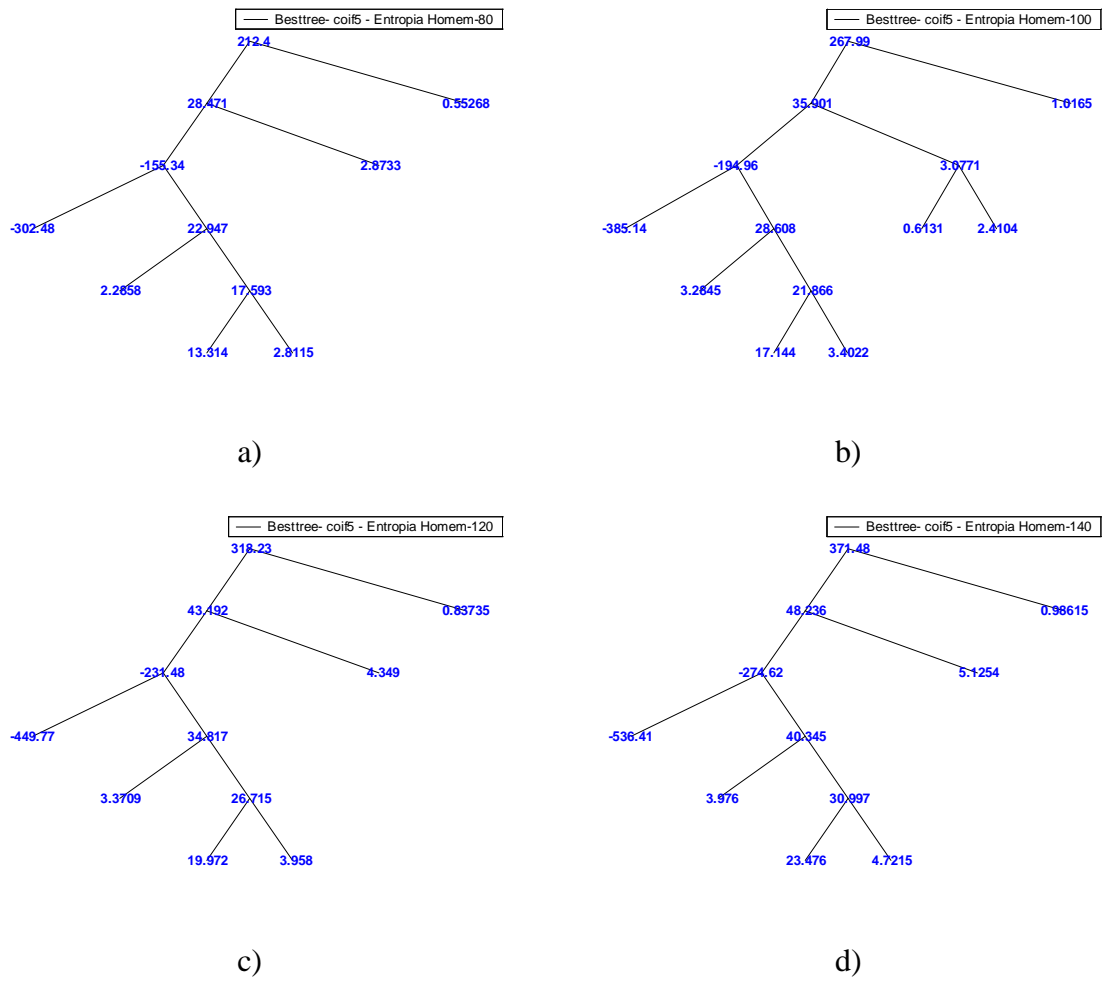


Figura A. 2 Best Basis do Homem com a função *coif5* para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

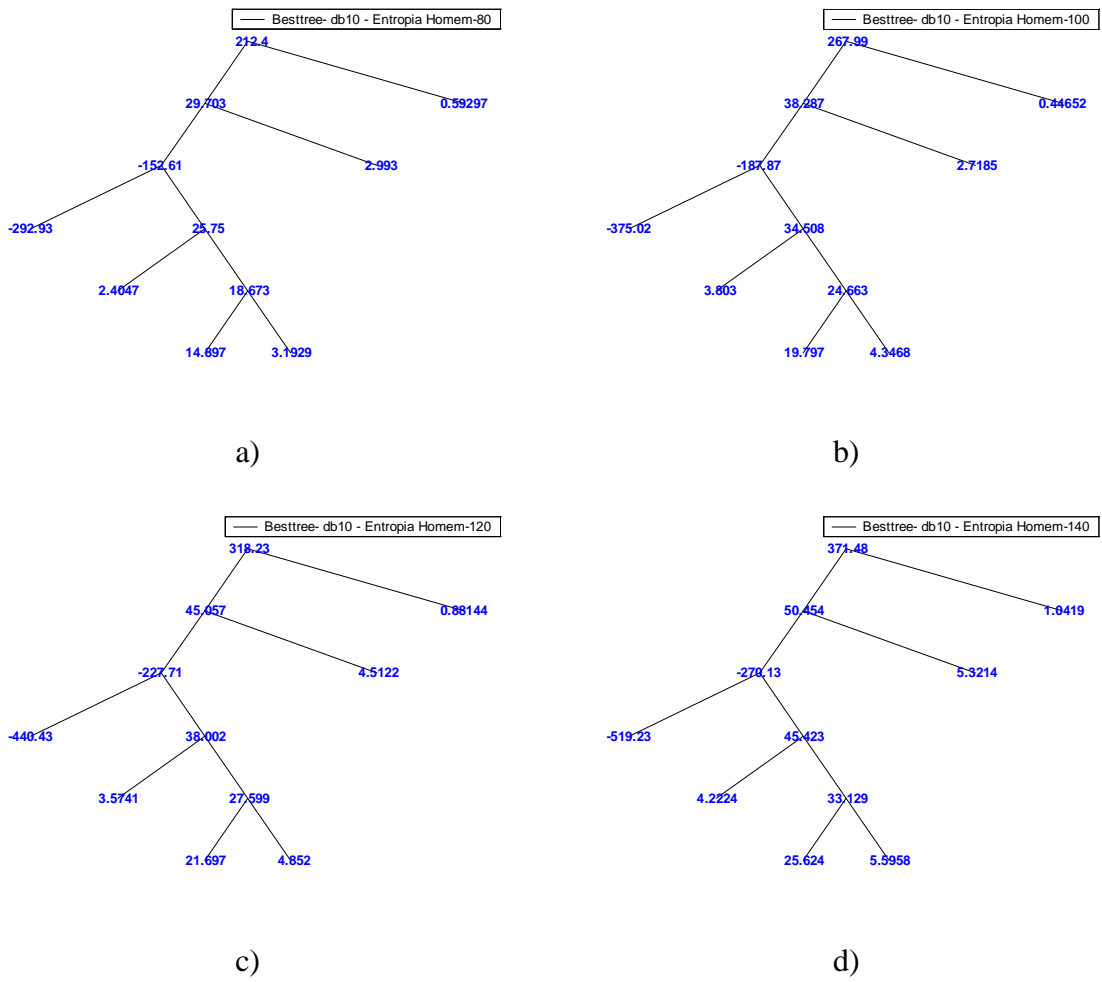


Figura A. 3 Best Basis do Homem com a função db10 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

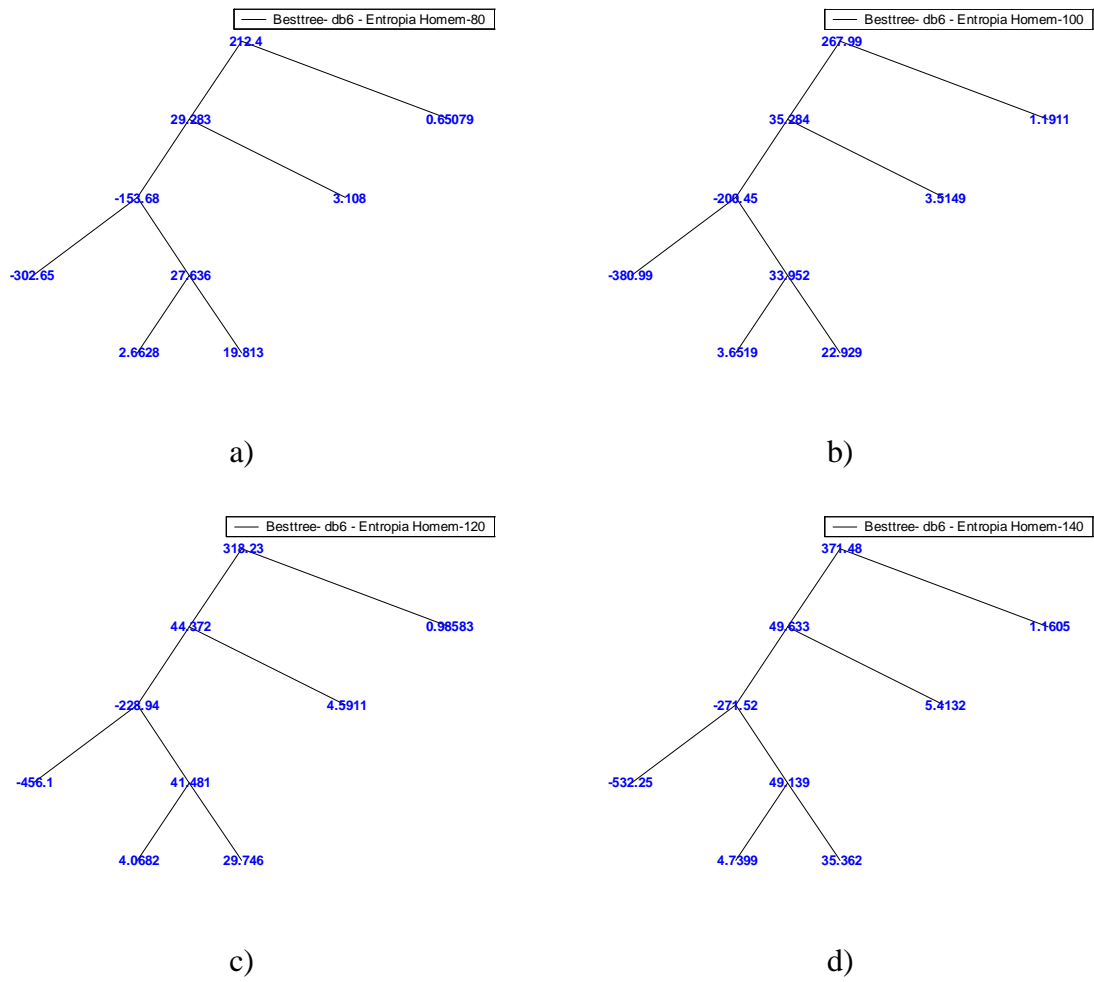


Figura A. 4 Best Basis do Homem com a função db6 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

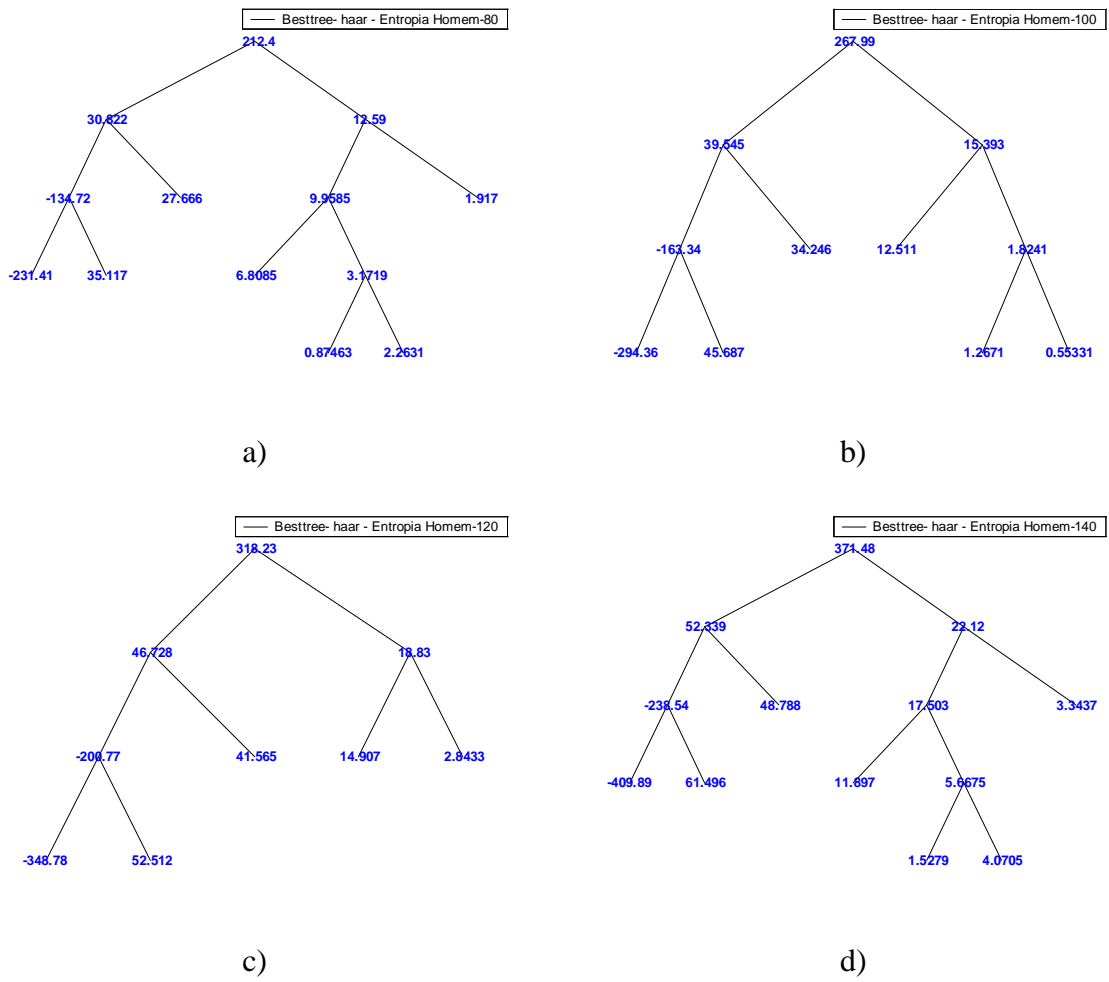
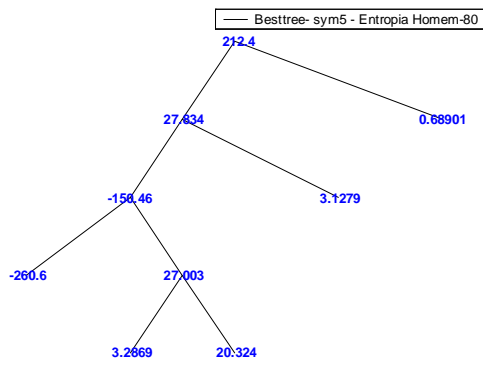
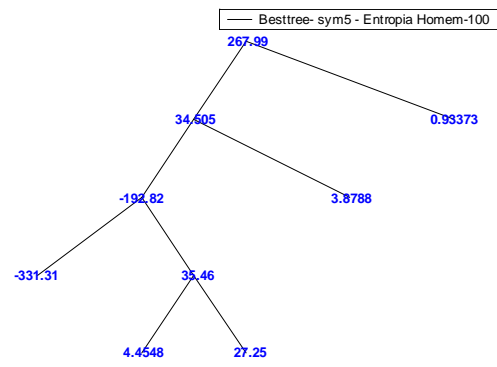


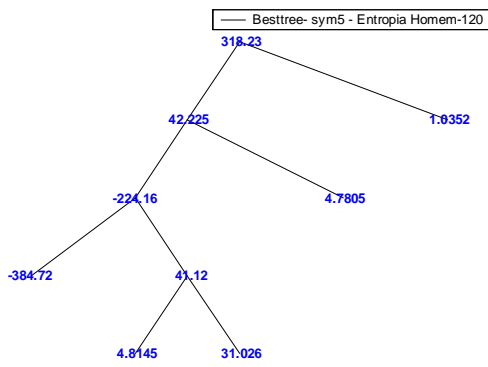
Figura A.5 Best Basis do Homem com a função haar para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.



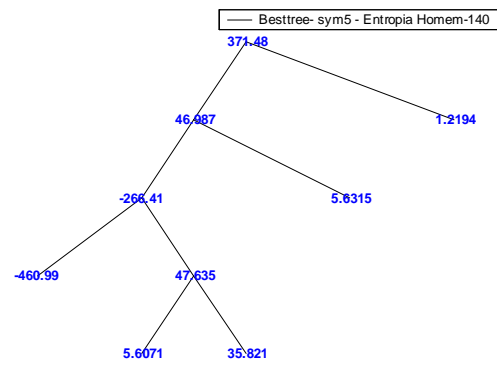
a)



b)

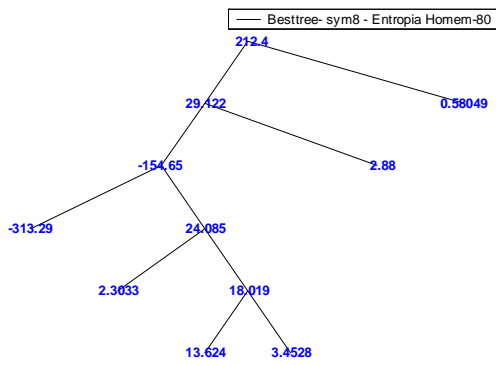


c)

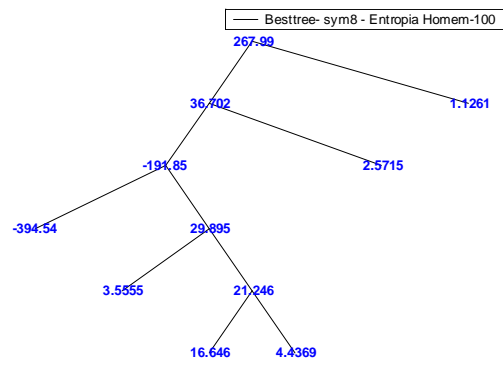


d)

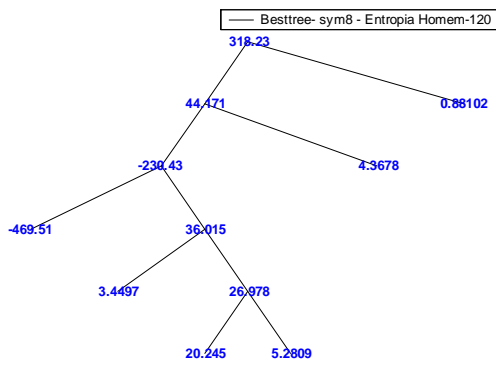
Figura A.6 Best Basis do Homem com a função sym5 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.



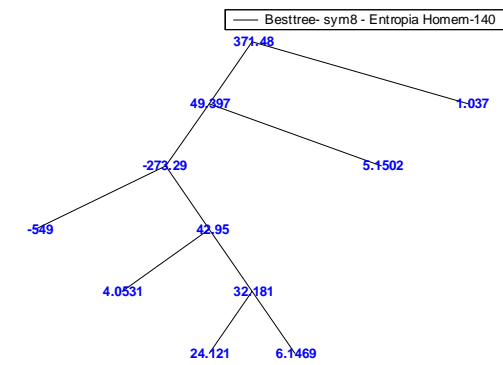
a)



b)



c)



d)

Figura A.7 Best Basis do Homem com a função sym8 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

Best Basis das Vogais Sintéticas Mulheres

Nas figuras a seguir, encontram-se as *Best Basis* do LPC das vogais mulheres com as funções Wavelet escolhidas e com a variação de frequência fundamental de 180Hz a 240Hz, conforme o item 4.3.2.

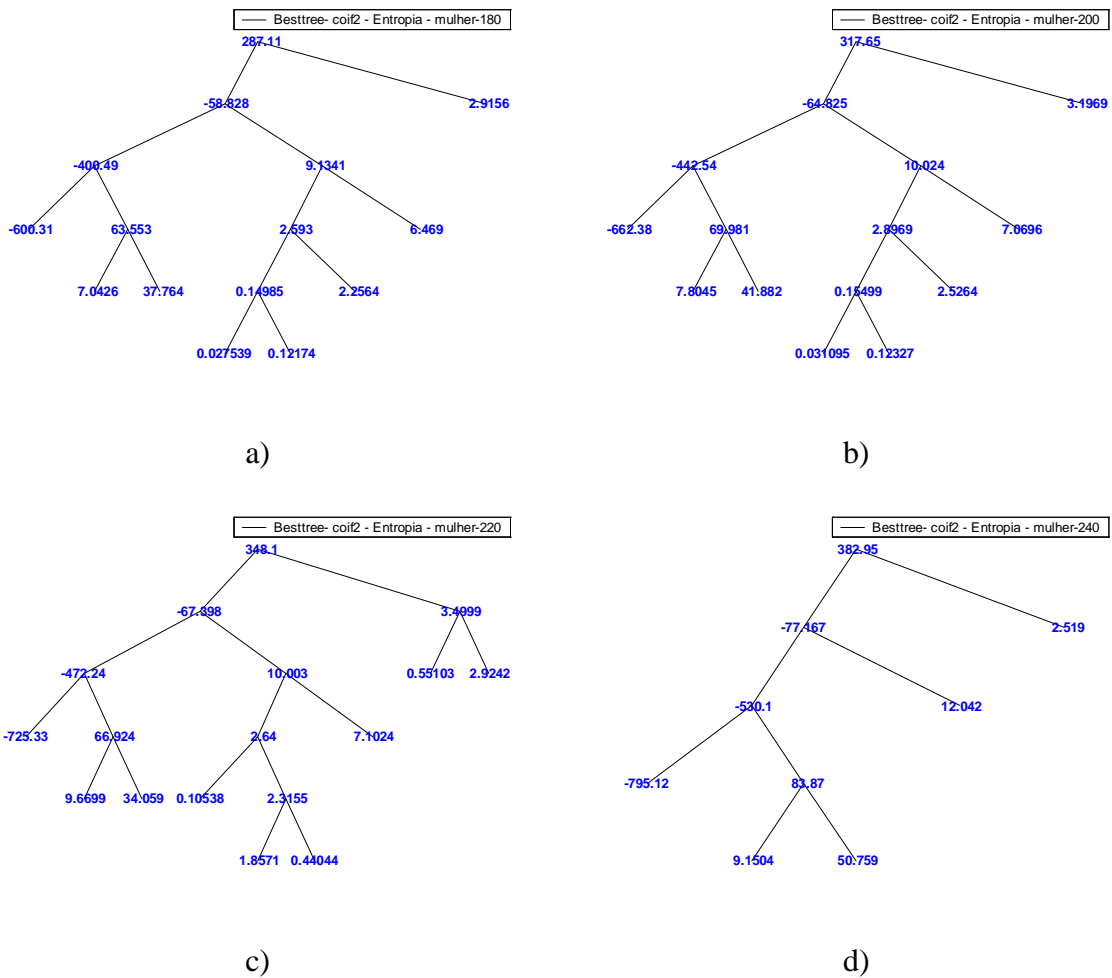


Figura A.8 *Best Basis* da Mulher com a função *coif2* para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$.

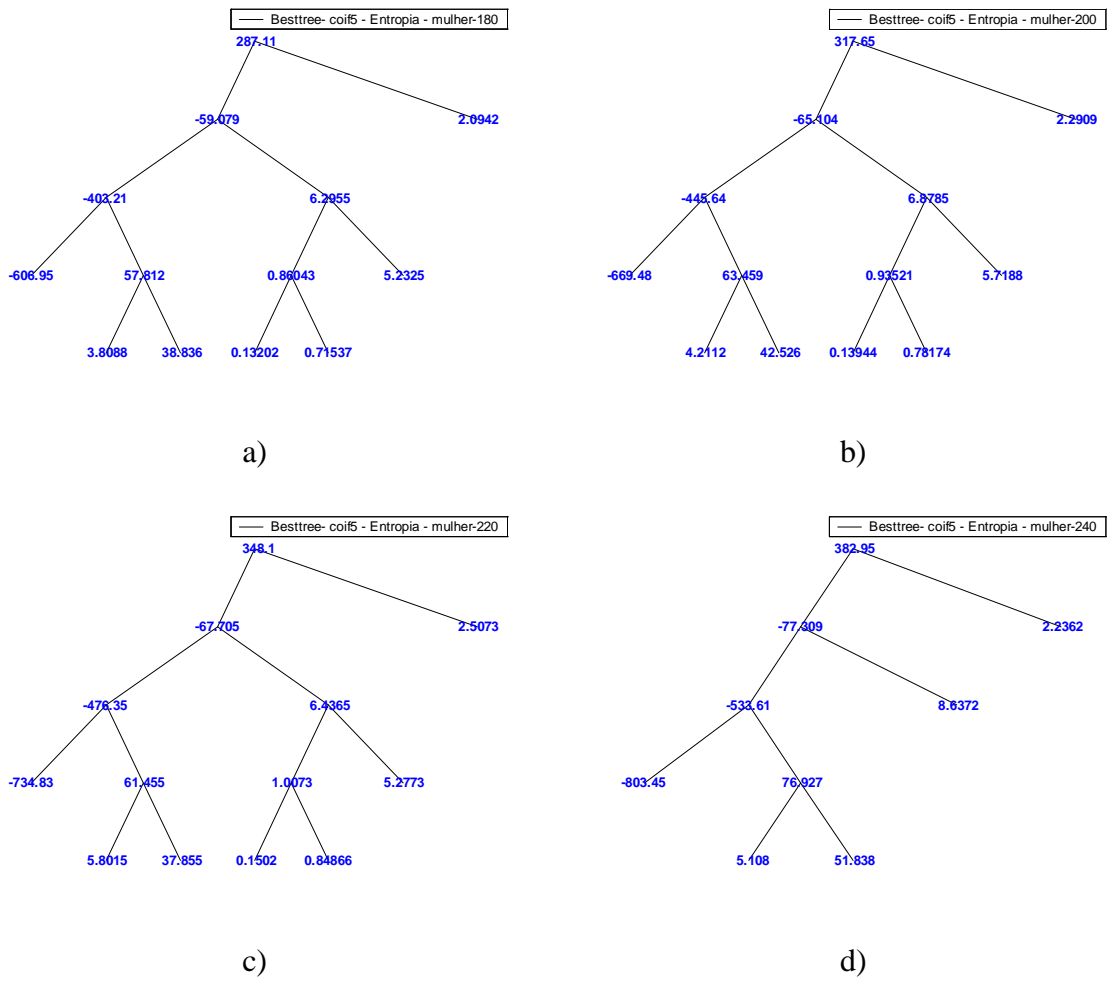


Figura A.9 Best Basis da Mulher com a função *coif5* para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$.

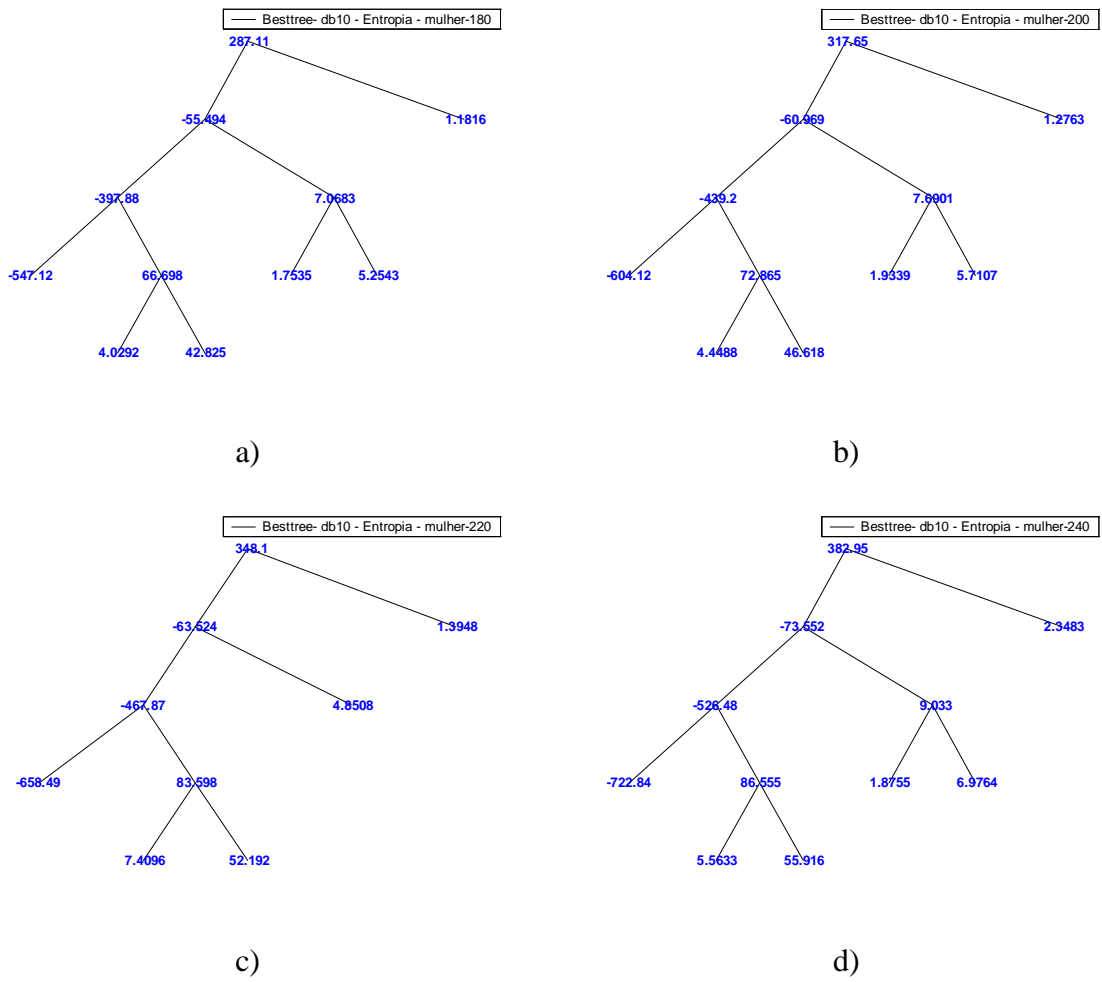


Figura A.10 Best Basis da Mulher com a função db10 para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$.

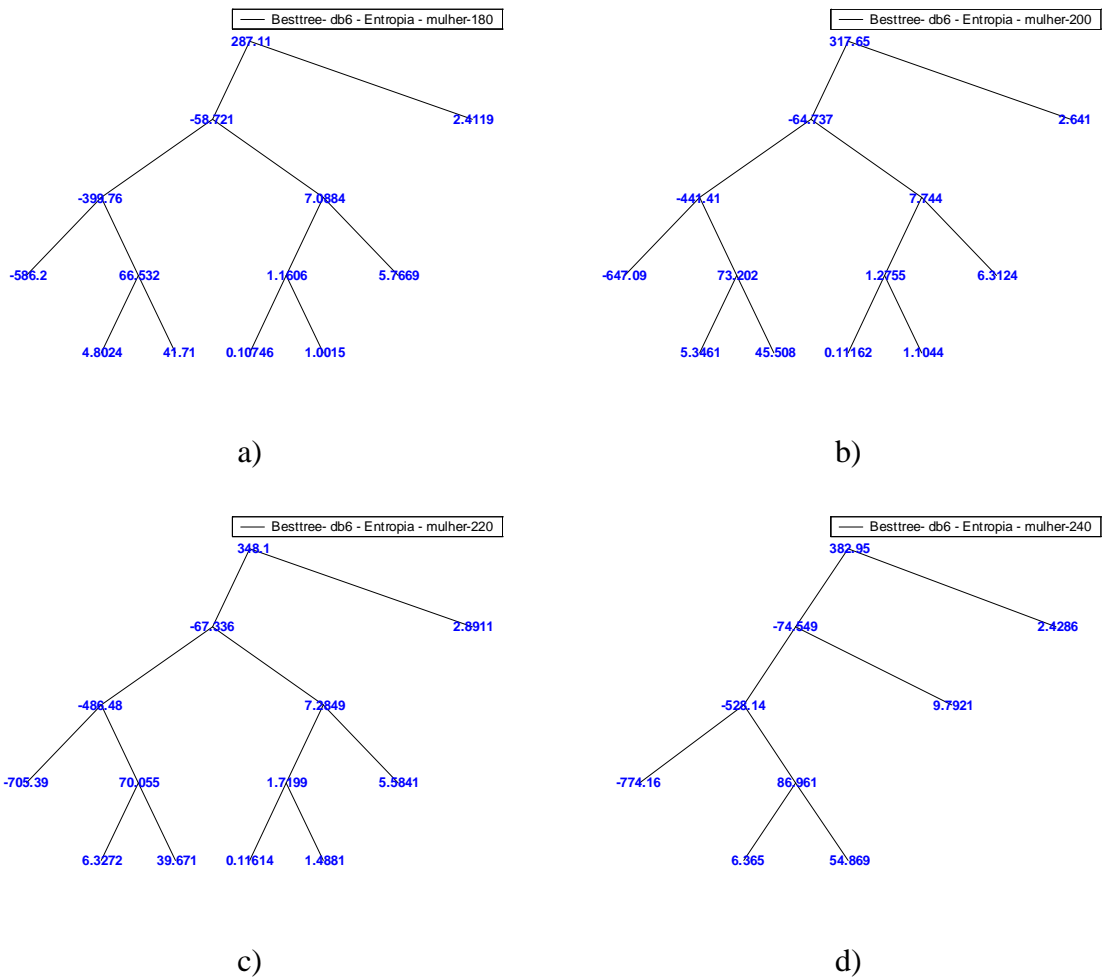


Figura A.11 Best Basis da Mulher com a função db6 para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$.

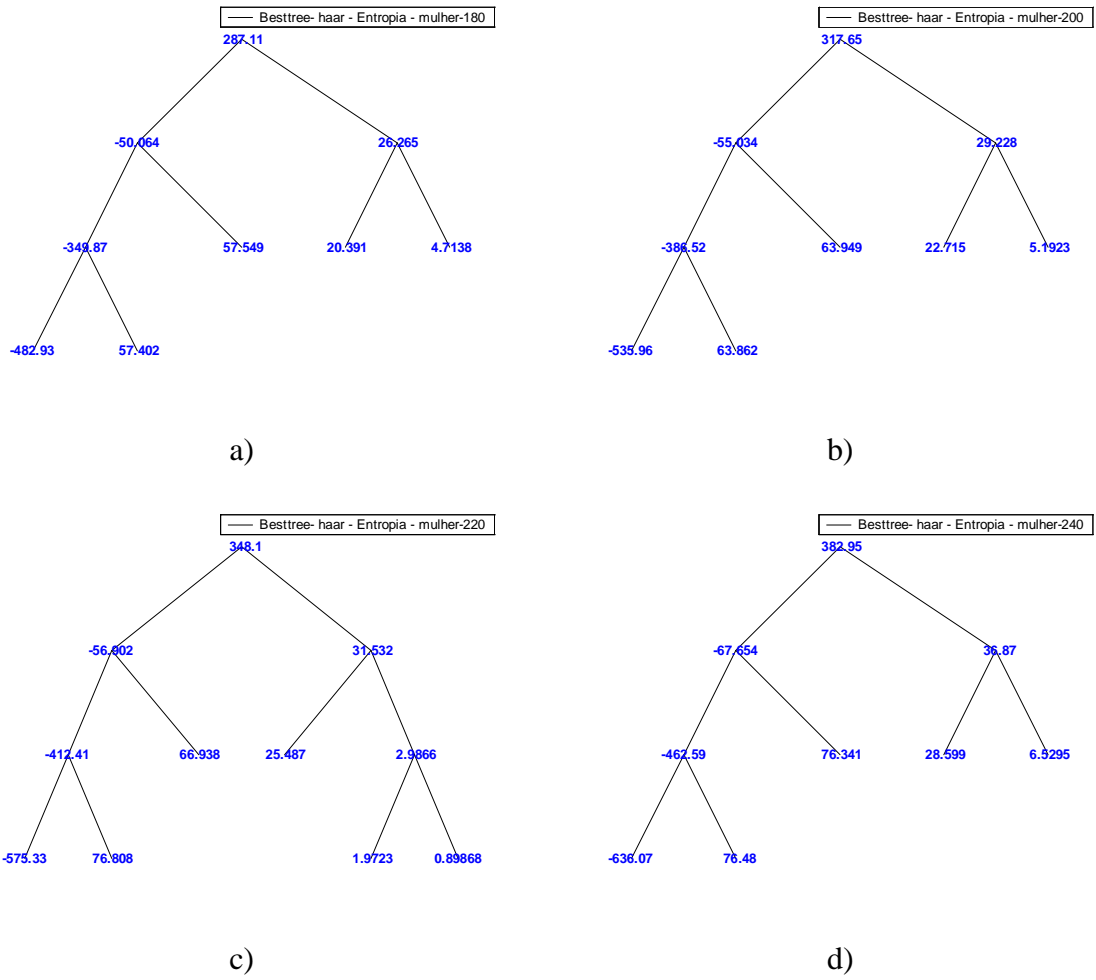


Figura A.12 Best Basis da Mulher com a função haar para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$.

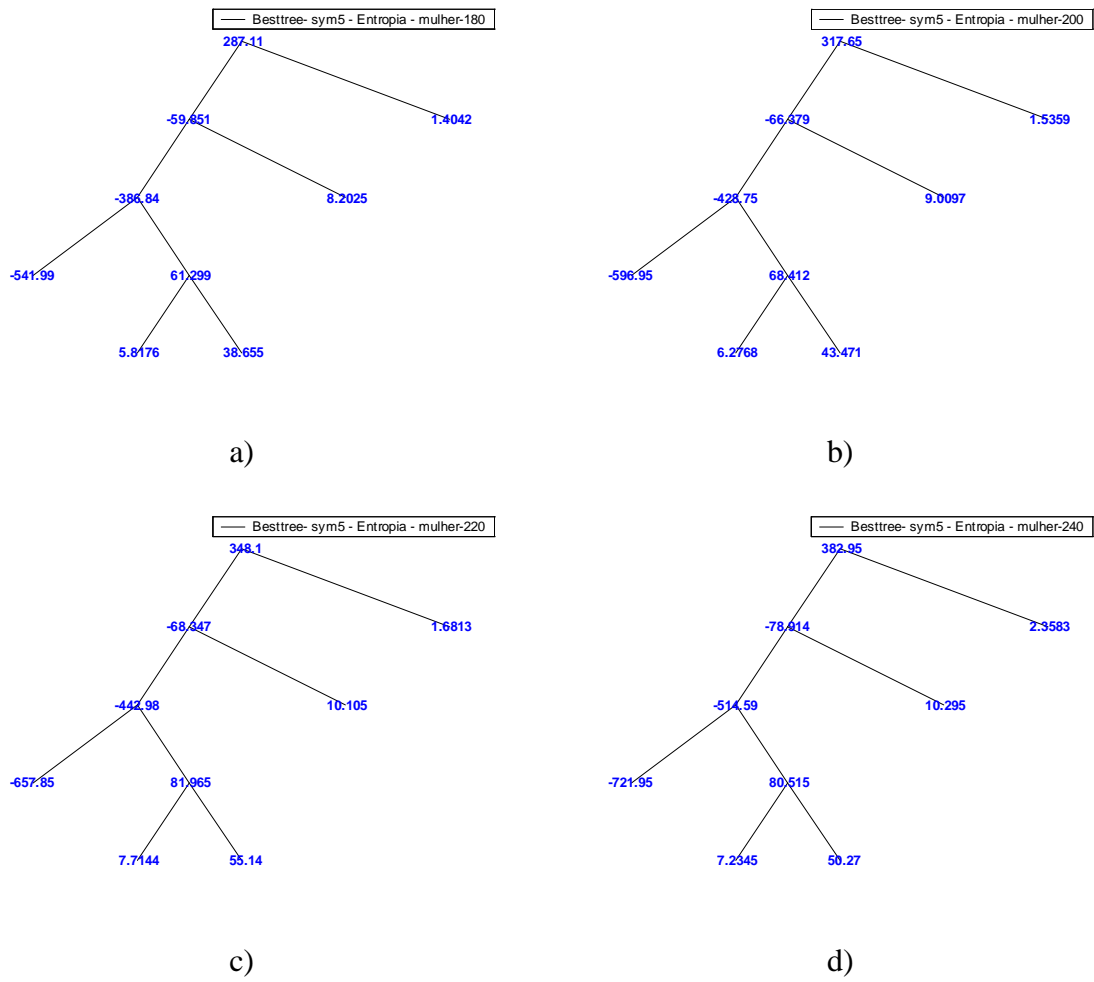


Figura A.13 Best Basis da Mulher com a função sym5 para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$.

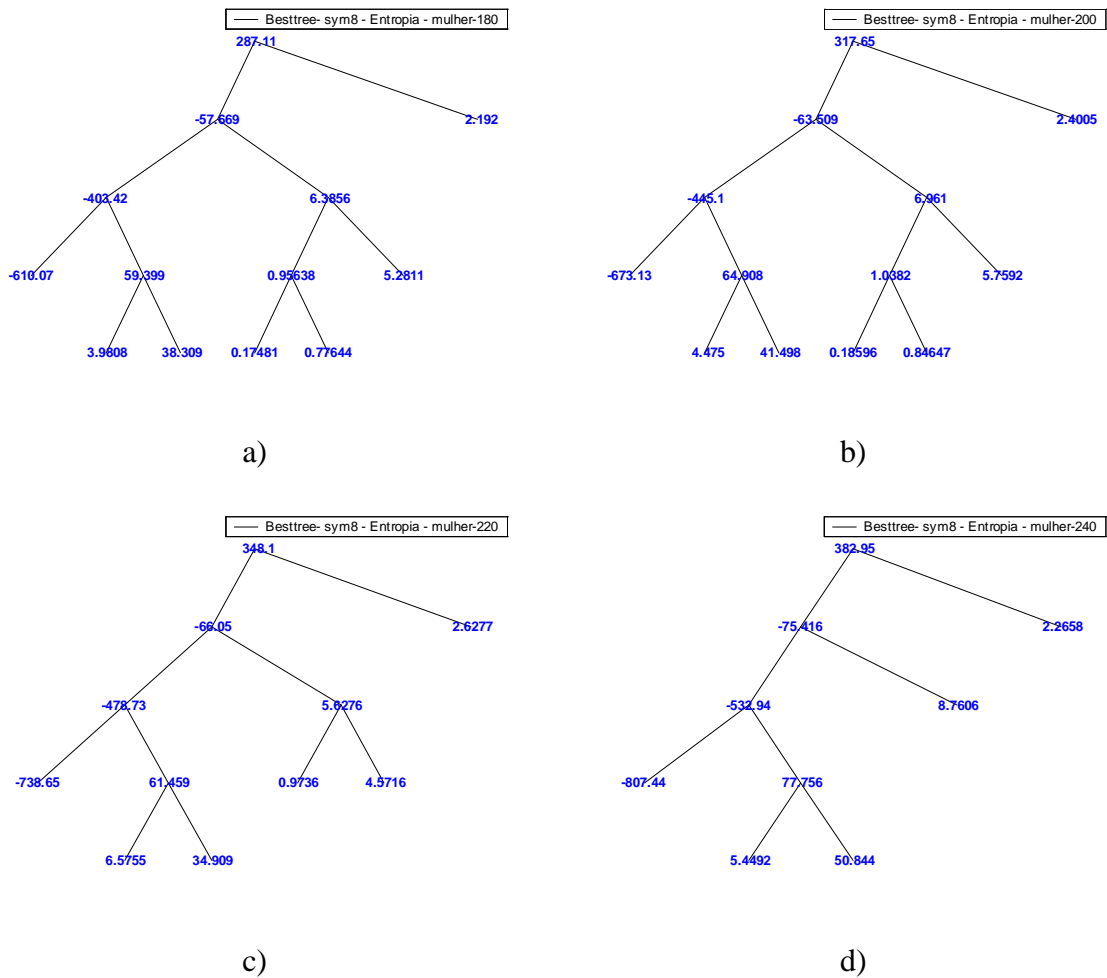


Figura A.14 Best Basis da Mulher com a função sym8 para a) $f_0=180\text{Hz}$ b) $f_0=200\text{Hz}$ c) $f_0=220\text{Hz}$ d) $f_0=240\text{Hz}$.

Best Basis das Vogais Sintéticas Casais

Nas figuras a seguir, encontram-se as *Best Basis* do LPC das vogais casais, com as funções Wavelet escolhidas, e com variação de frequência fundamental de 80Hz a 240Hz, conforme o item 4.3.3.

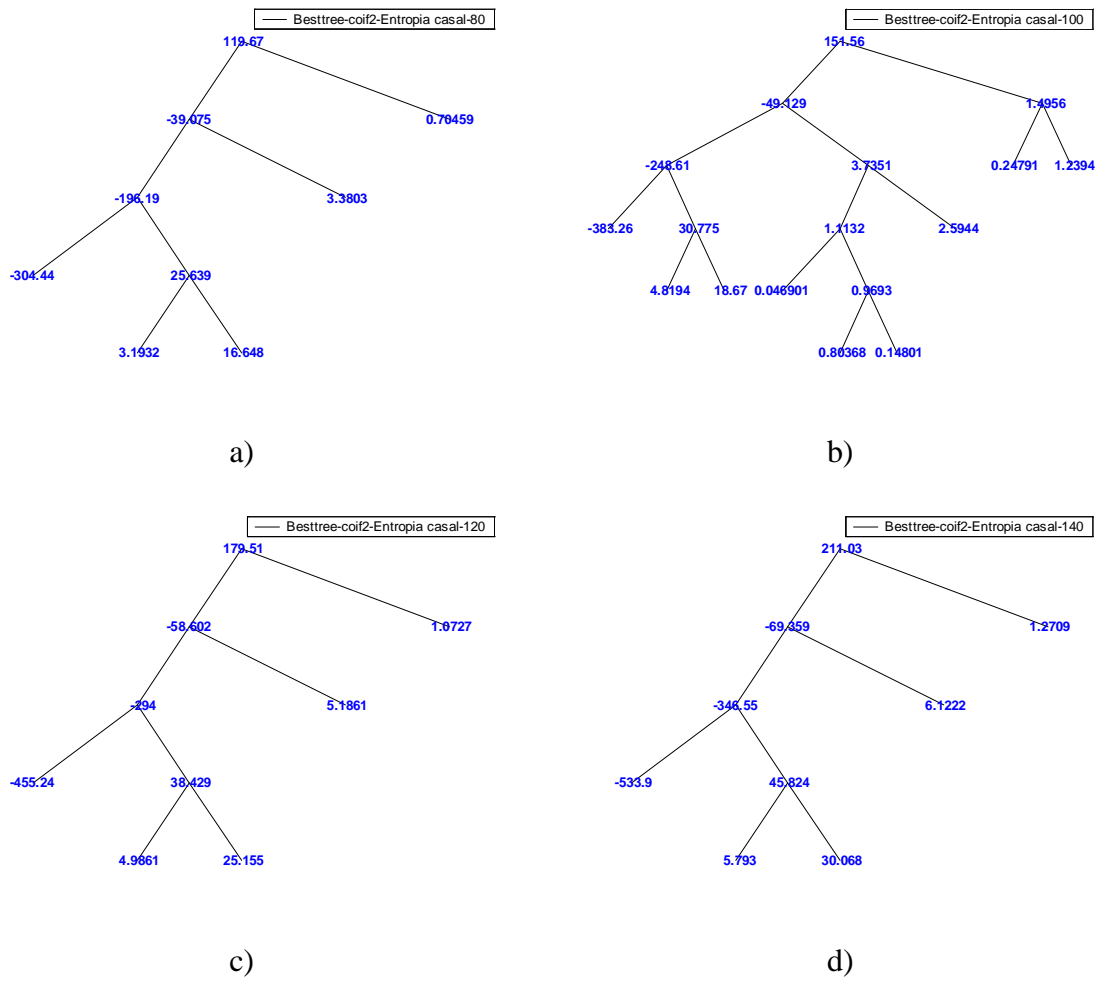
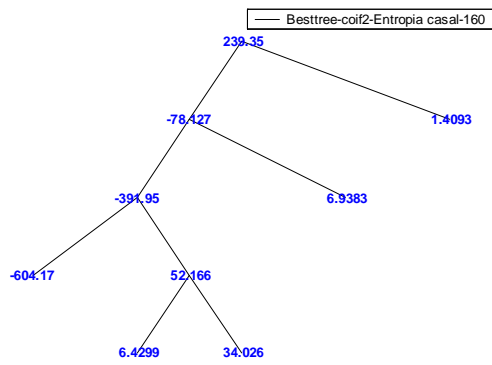
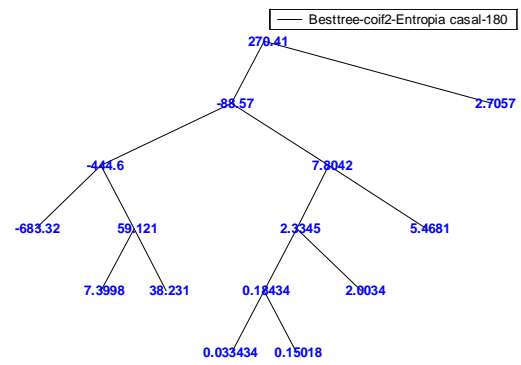


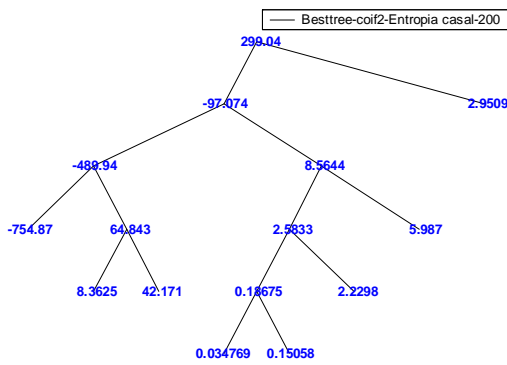
Figura A.15 Best Basis Casal com a função coif2 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.



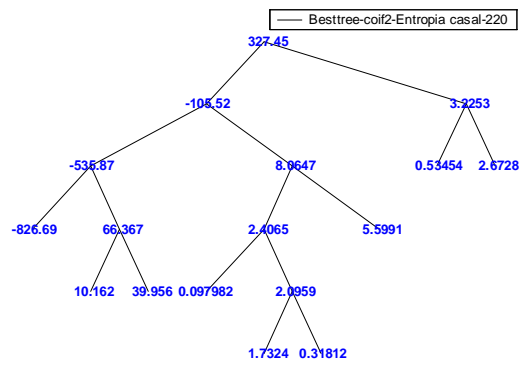
a)



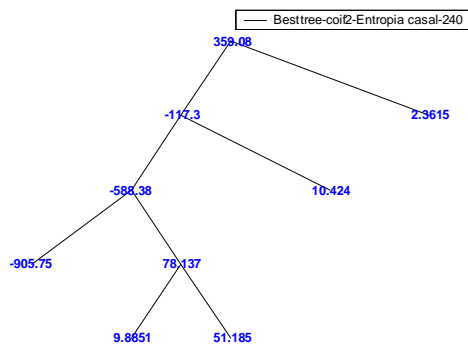
b)



c)



d)



e)

Figura A.16 Best Basis Casal com a função coif2 para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$.

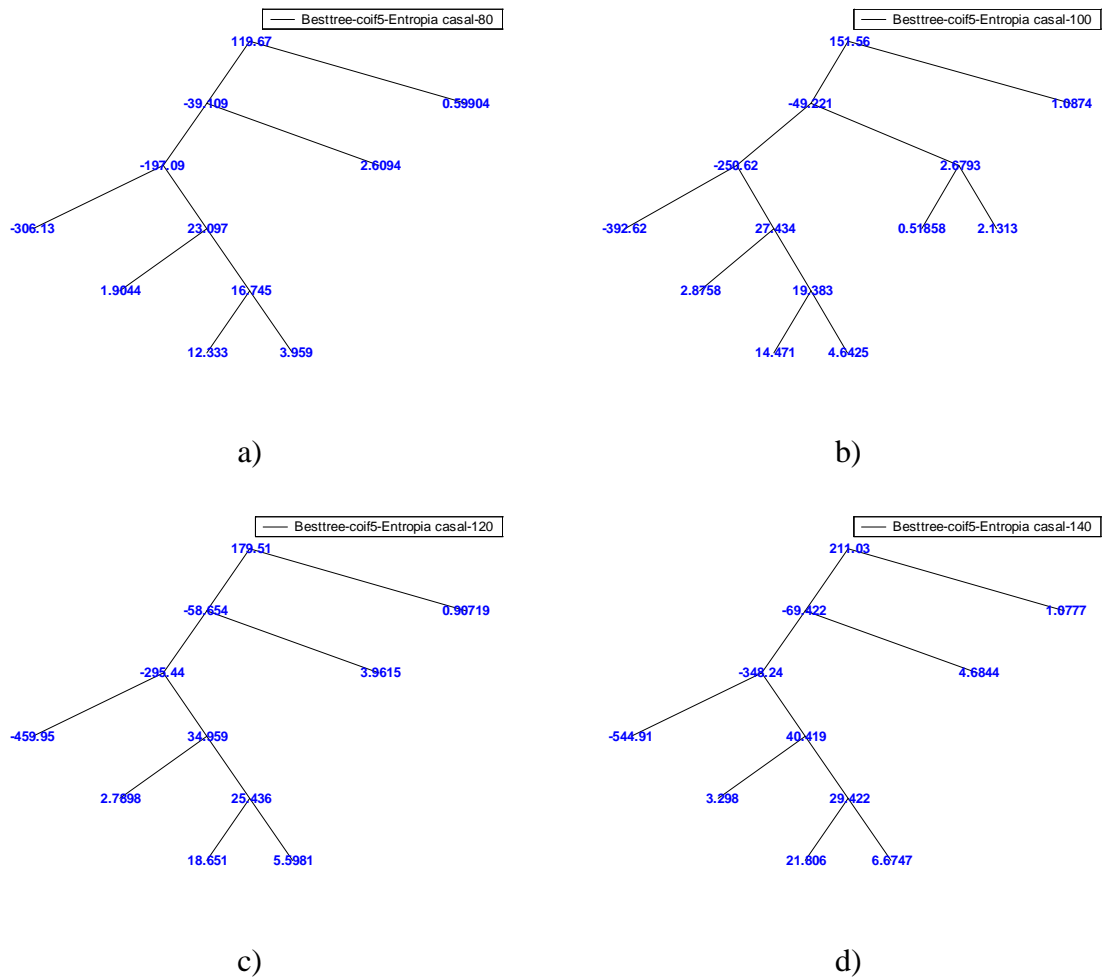


Figura A.17 *Best Basis Casal* com a função *coif5* para a) $f_0=80$ Hz b) $f_0=100$ Hz c) $f_0=120$ Hz d) $f_0=140$ Hz.

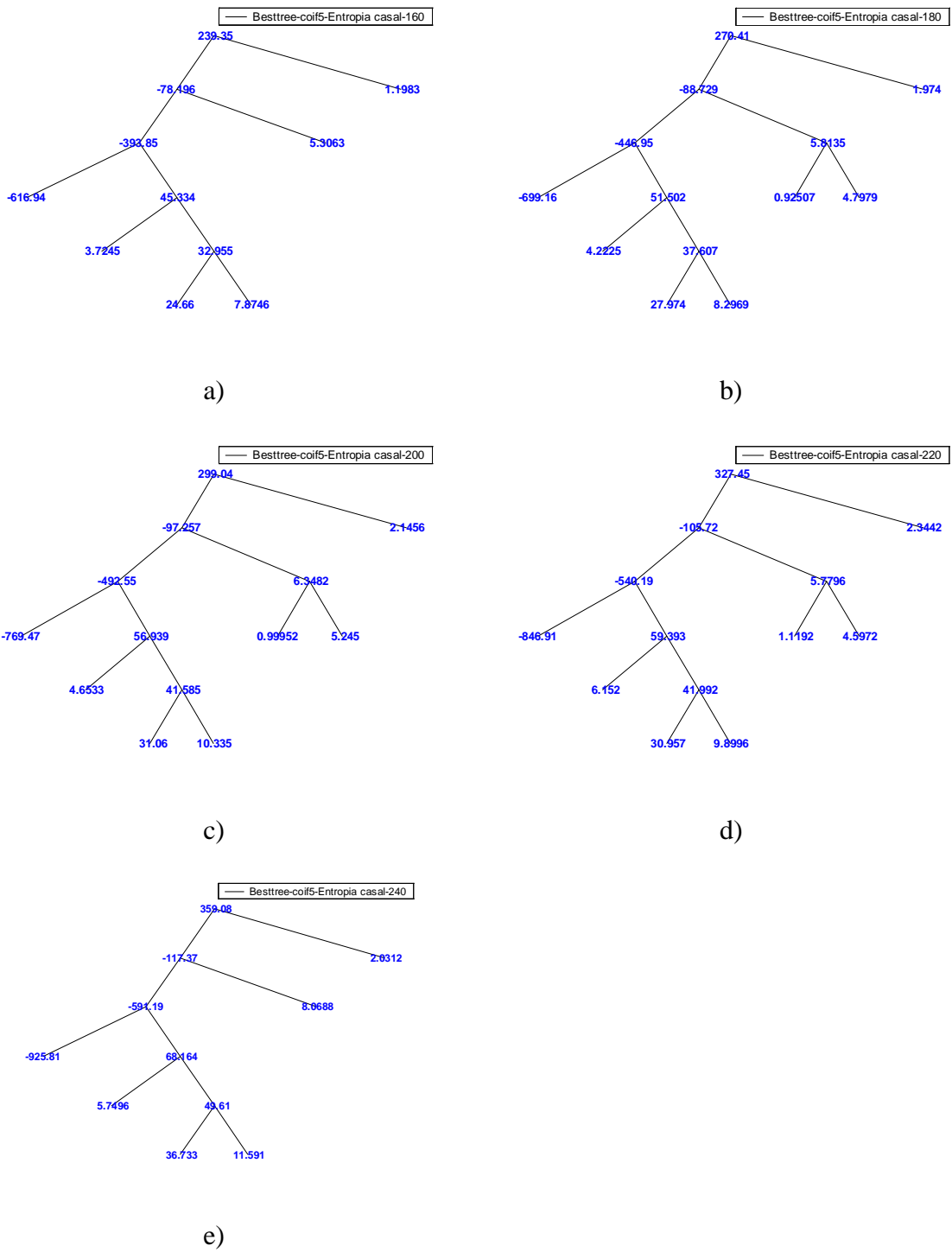


Figura A.18 *Best Basis Casal* com a função *coif5* para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$
 c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$.

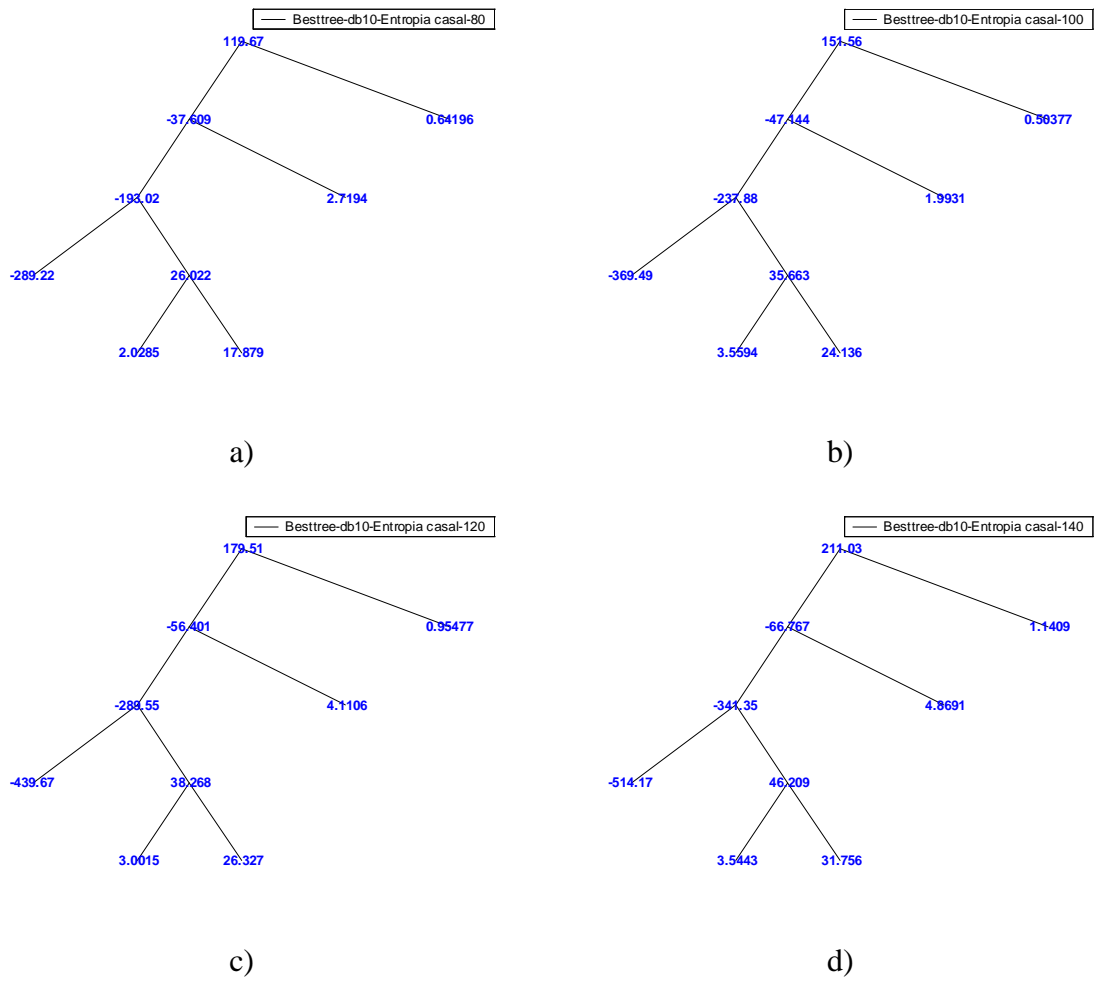


Figura A.19 Best Basis Casal com a função db10 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

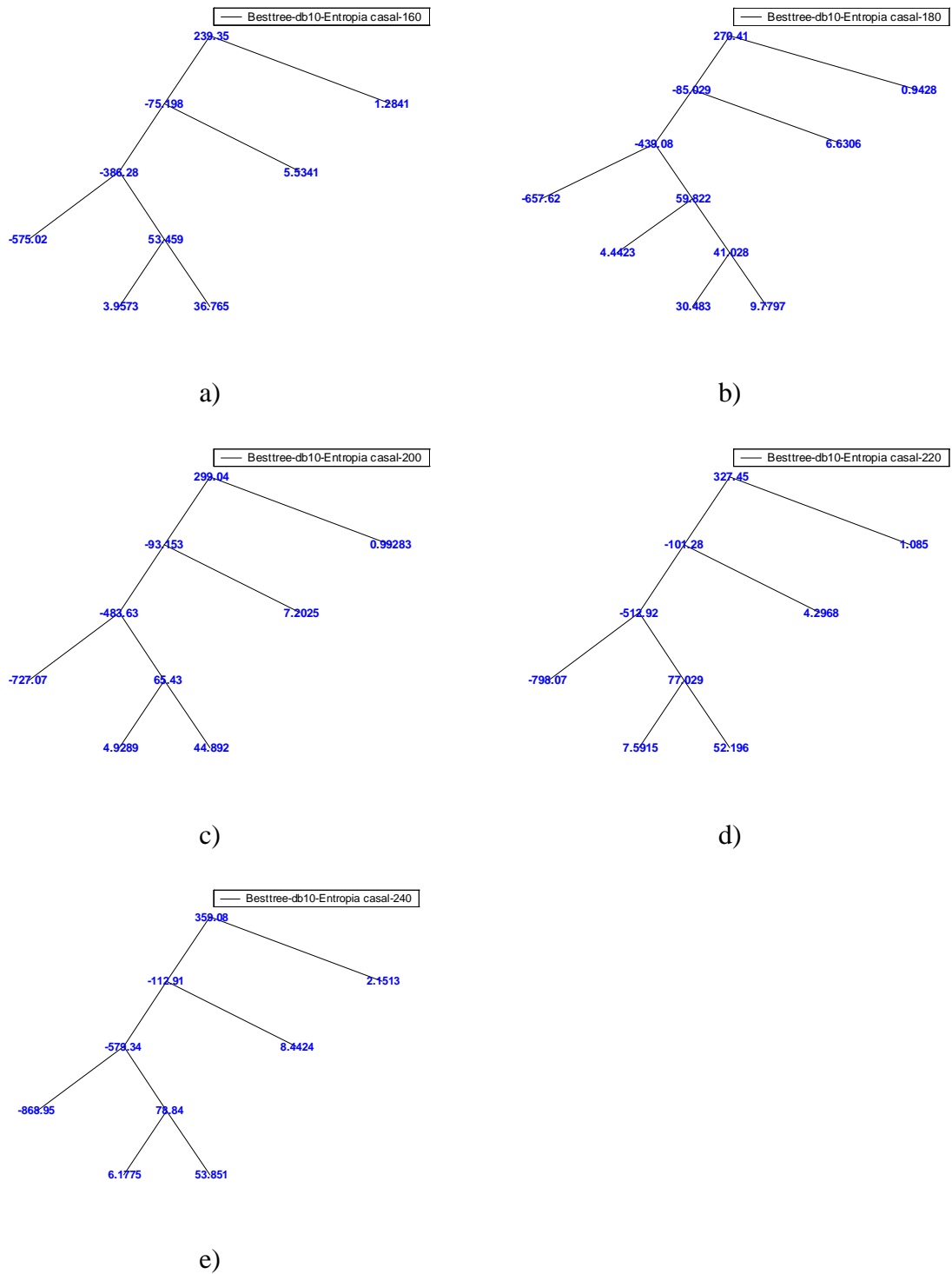


Figura A.20 Best Basis Casal com a função db10 para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$.

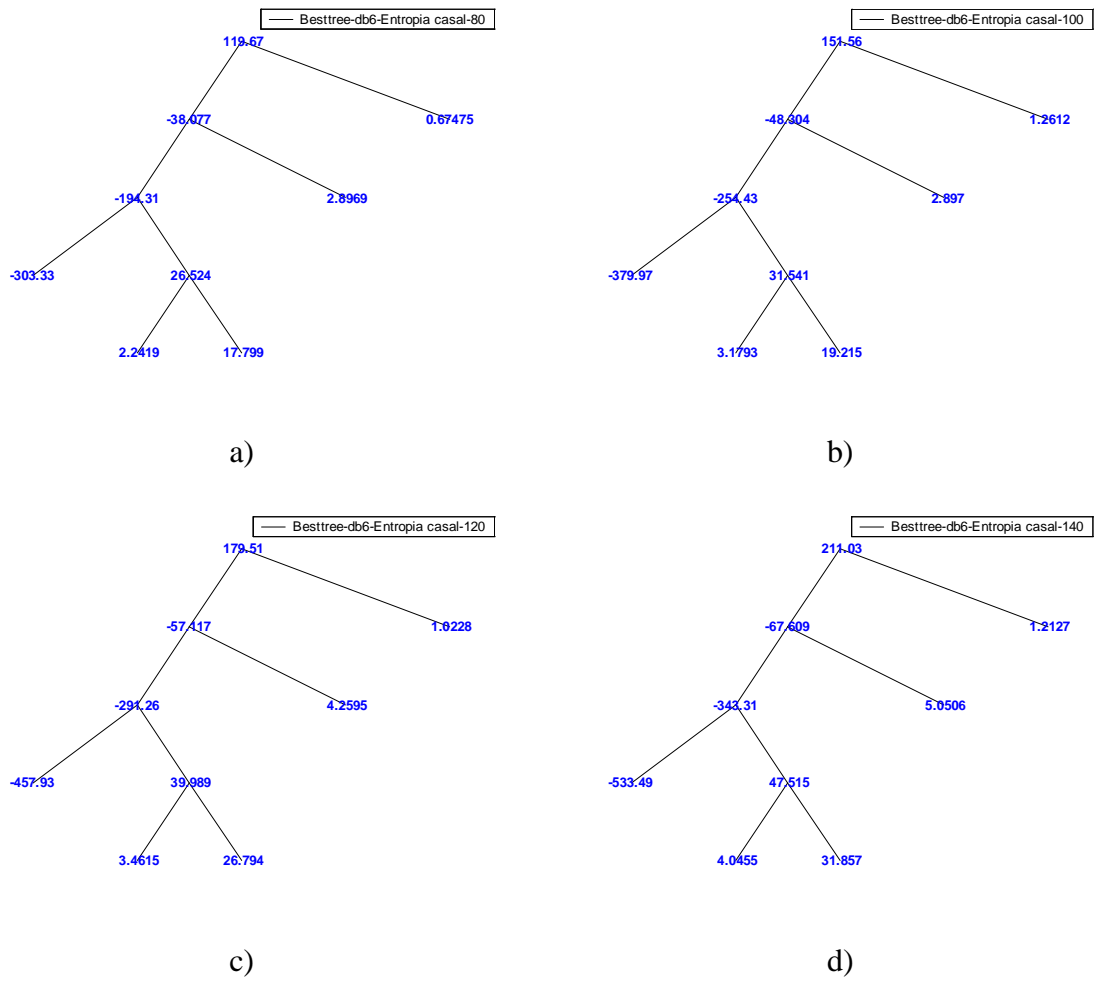


Figura A.21 Best Basis Casal com a função db6 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

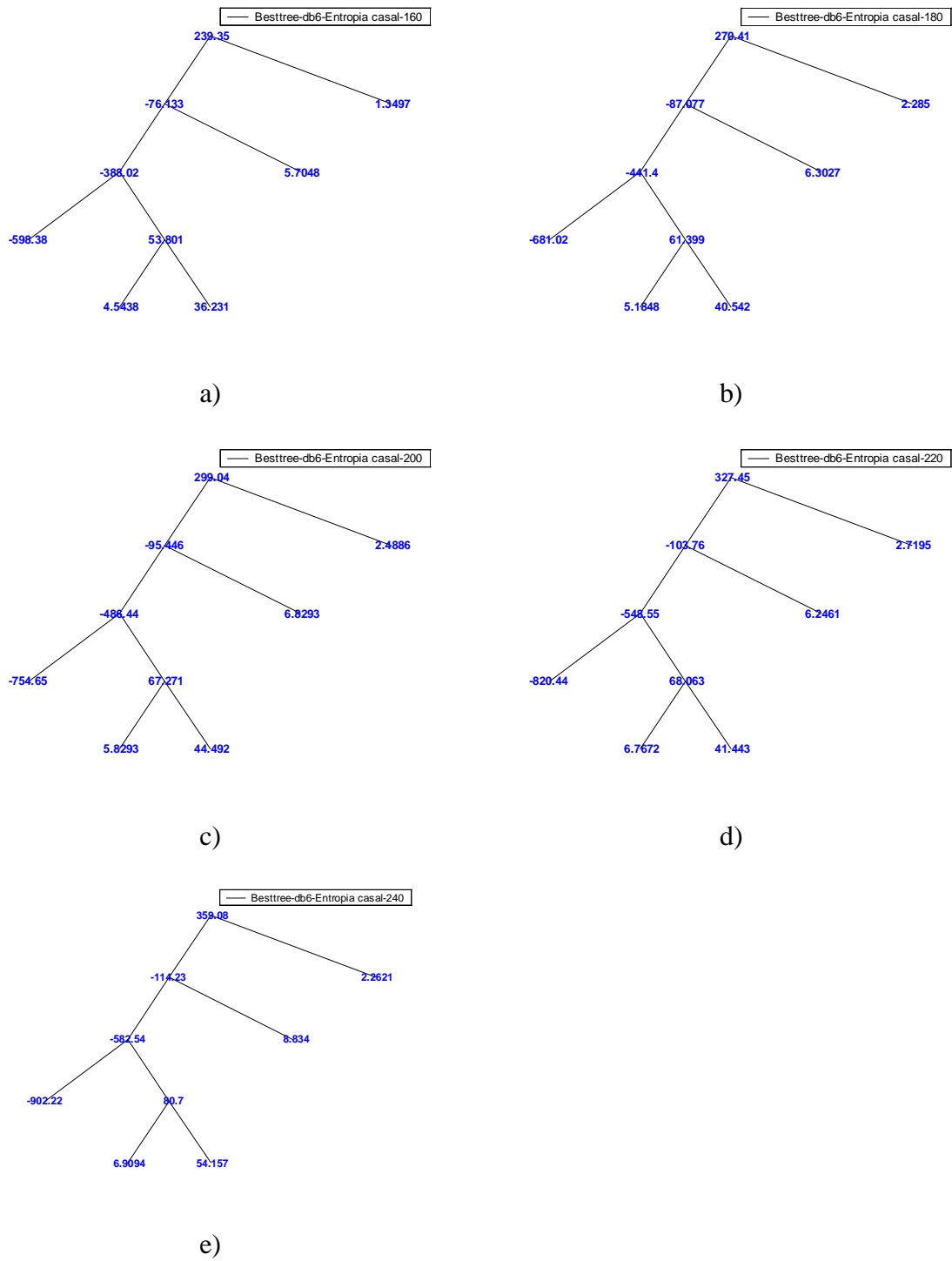


Figura A.22 Best Basis Casal com a função db6 para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$.

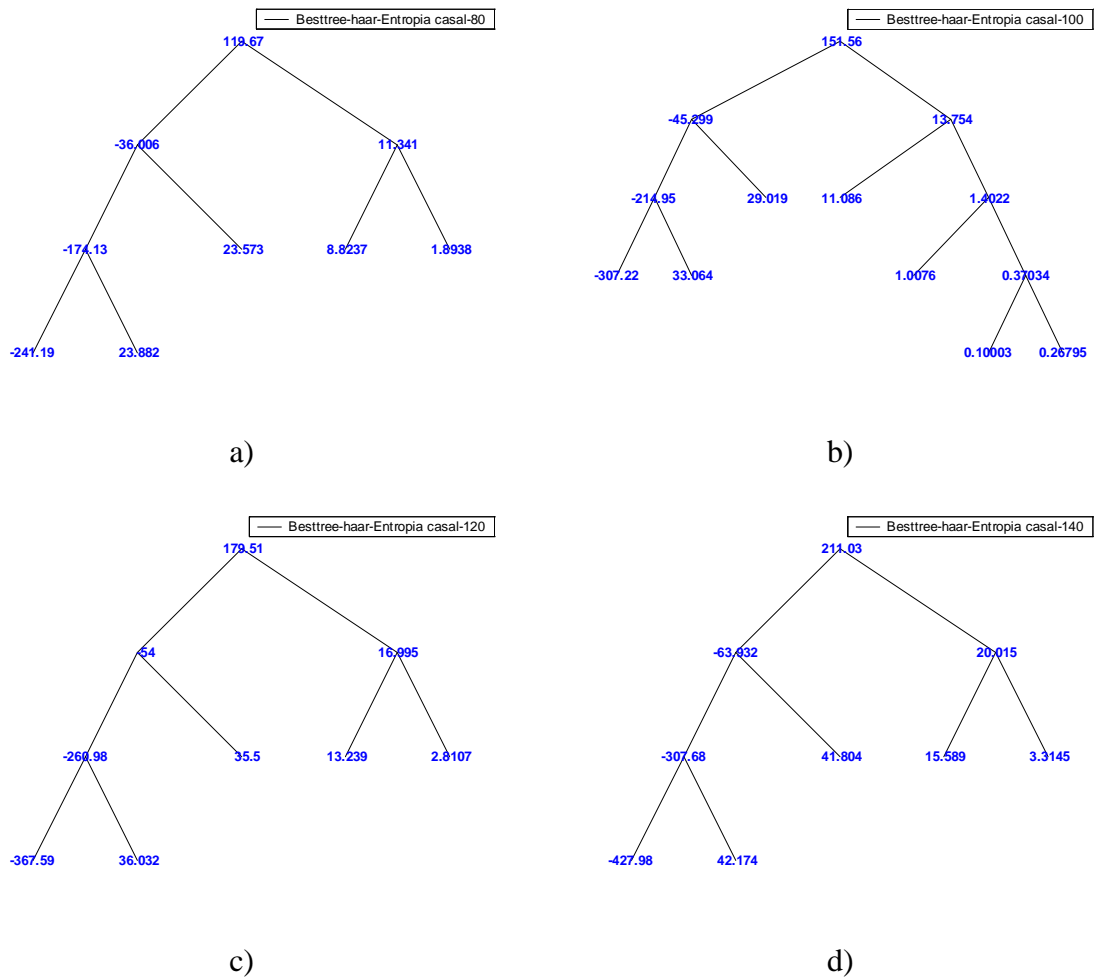


Figura A.23 Best Basis Casal com a função haar para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

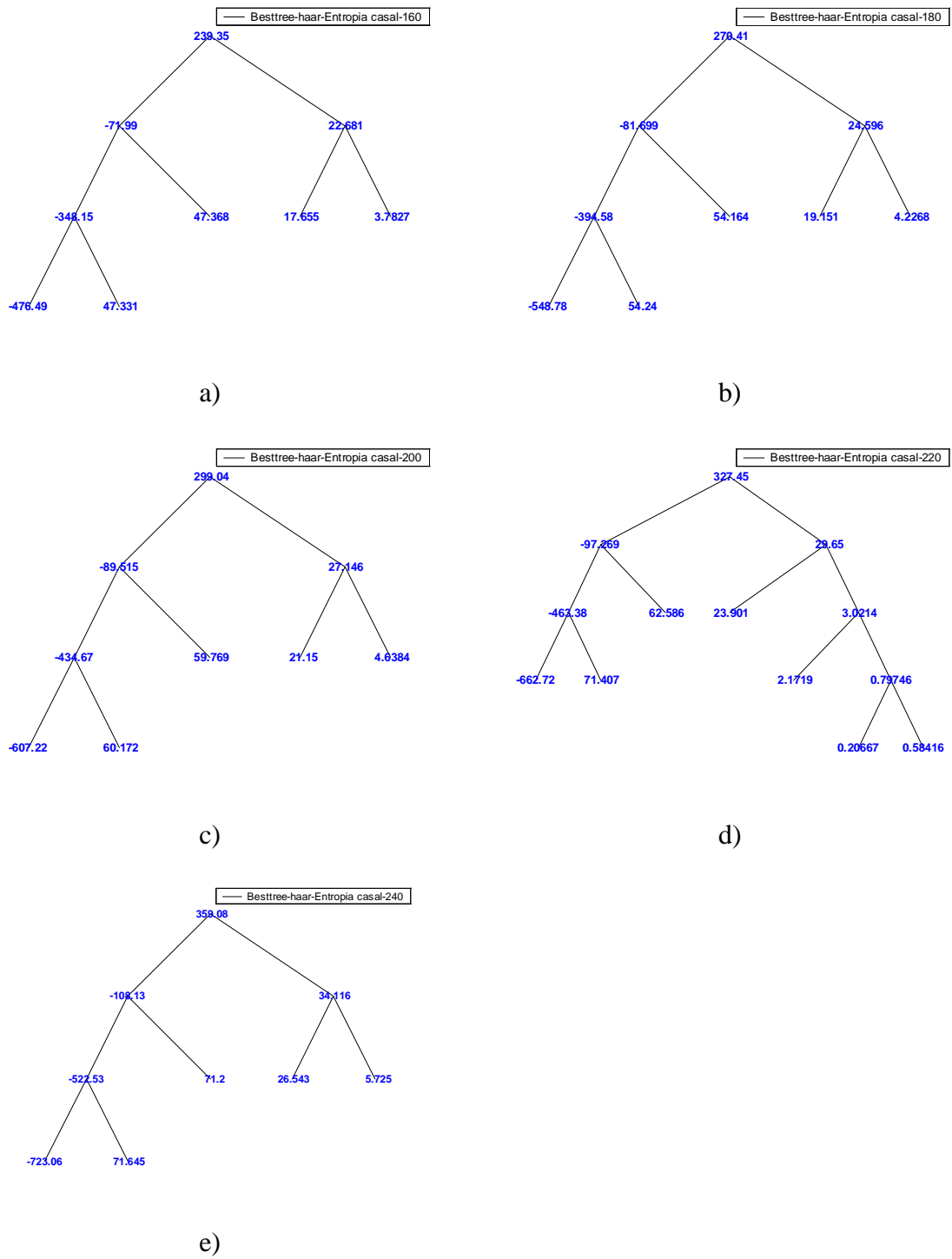


Figura A.24 Best Basis Casal com a função haar para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$.

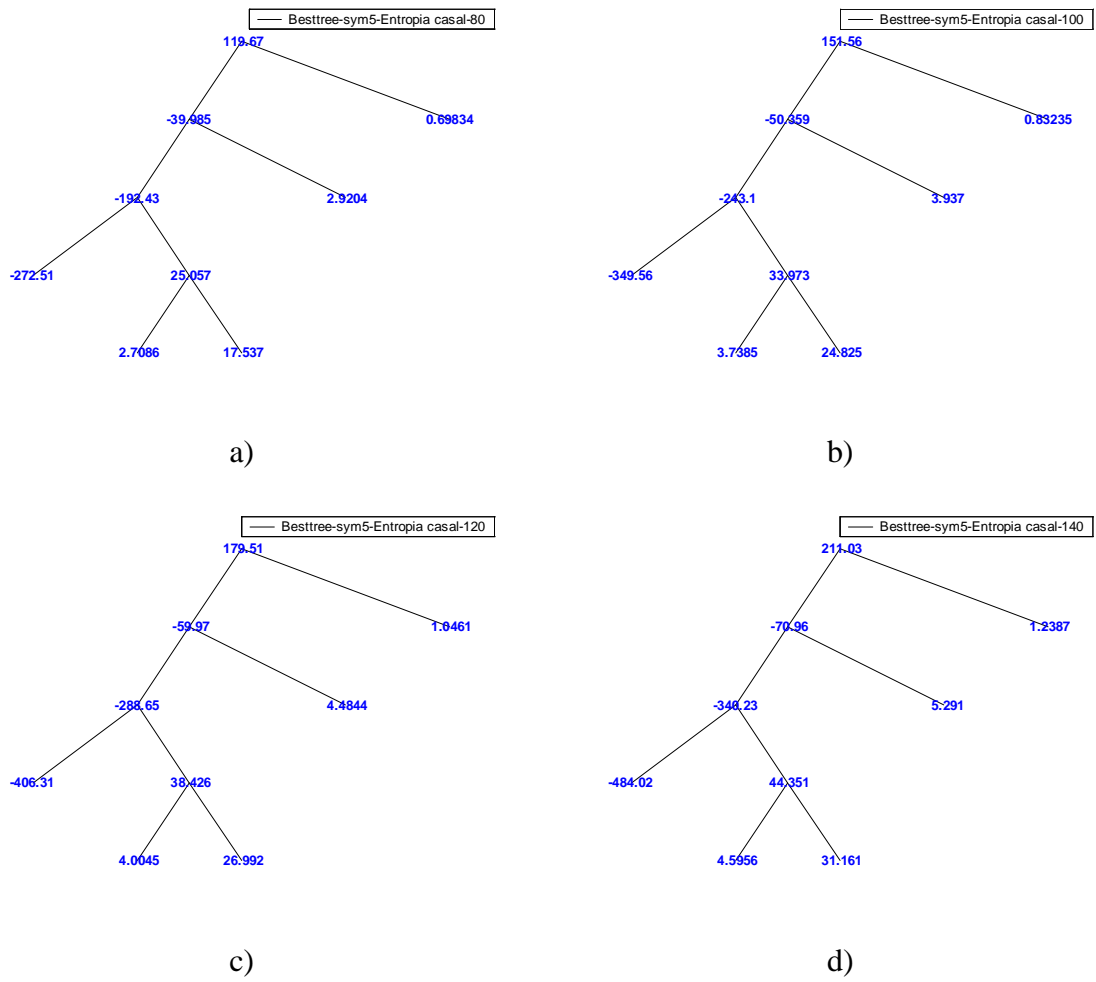


Figura A.25 Best Basis Casal com a função sym5 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.

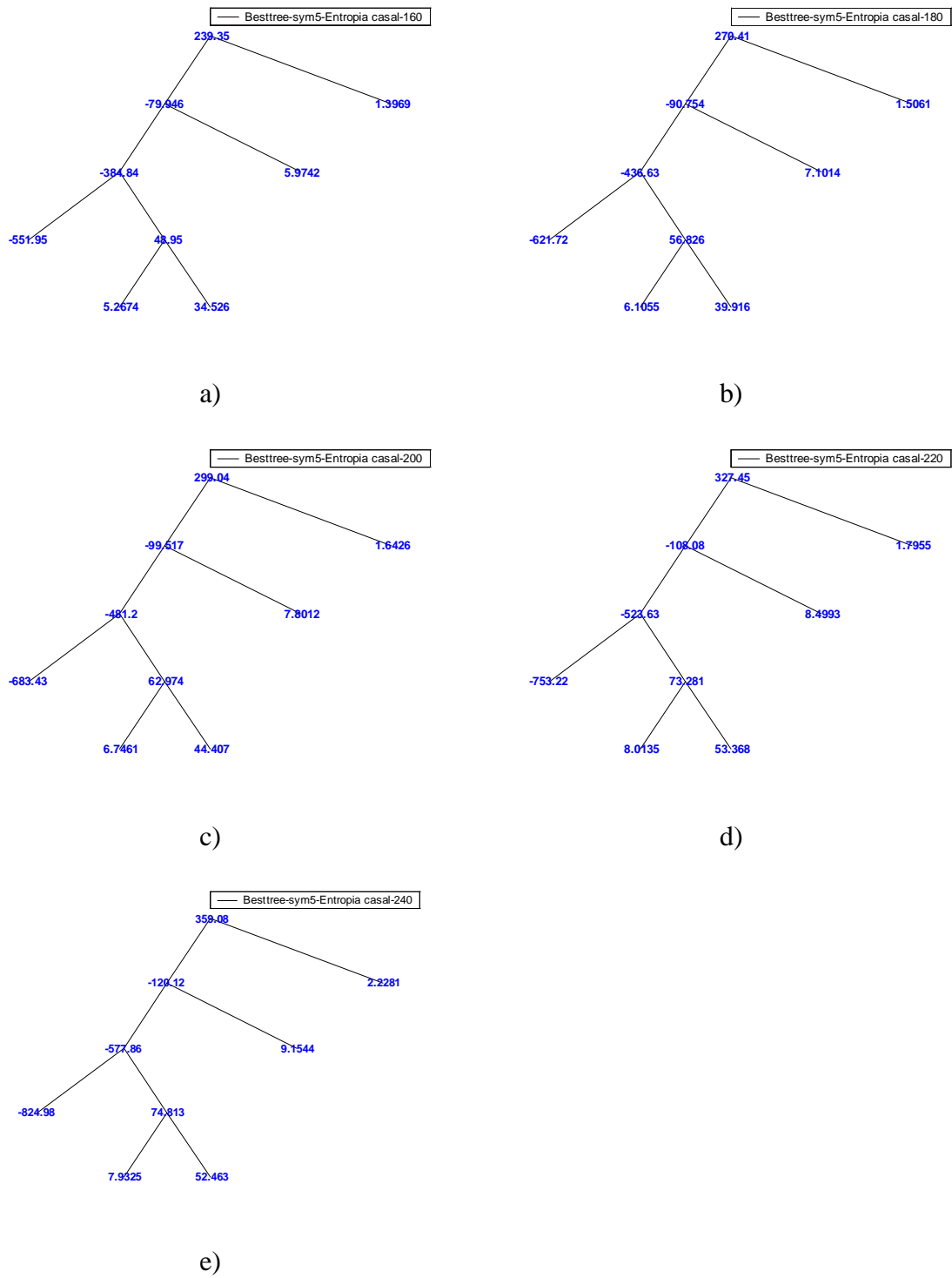


Figura A.26 Best Basis Casal com a função sym5 para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$.

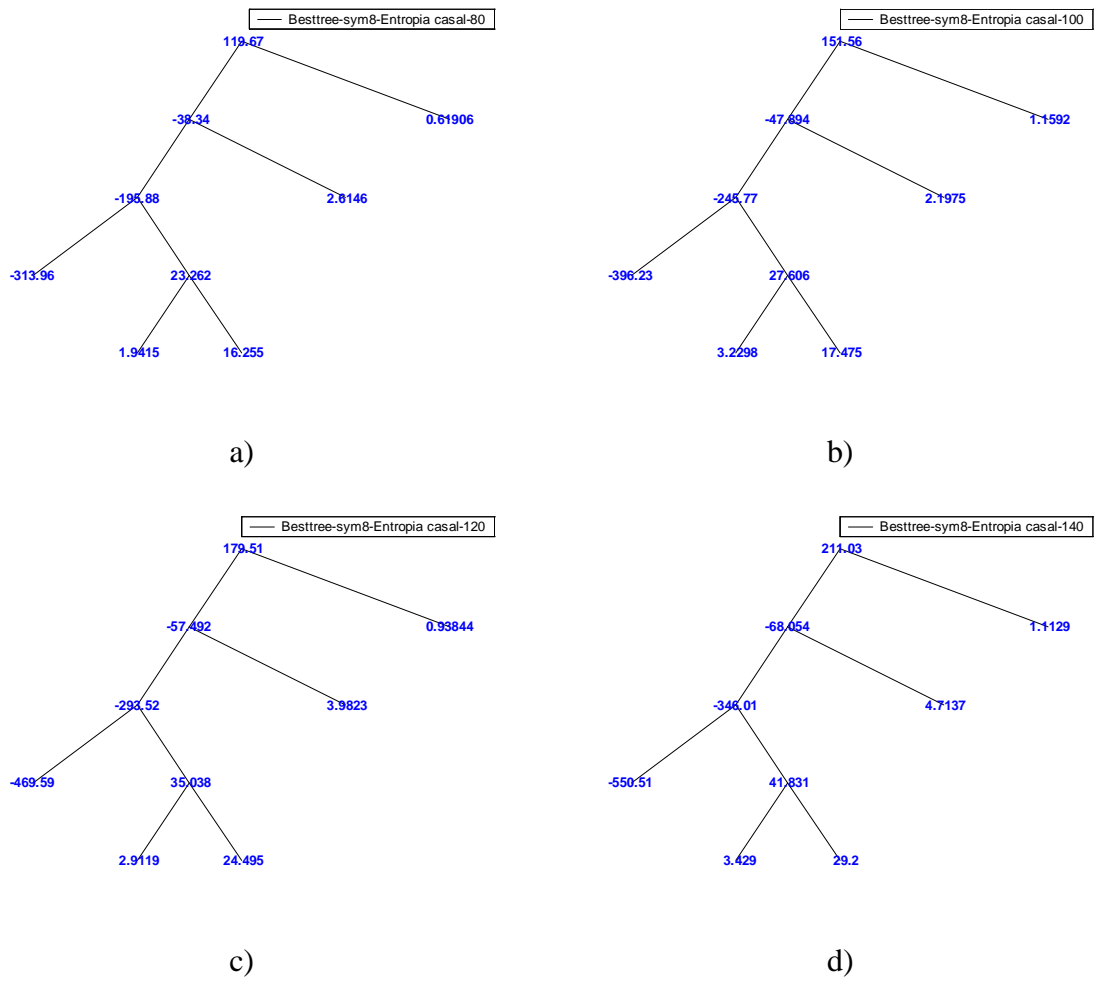
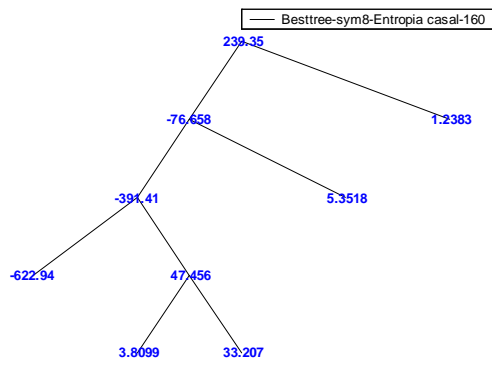
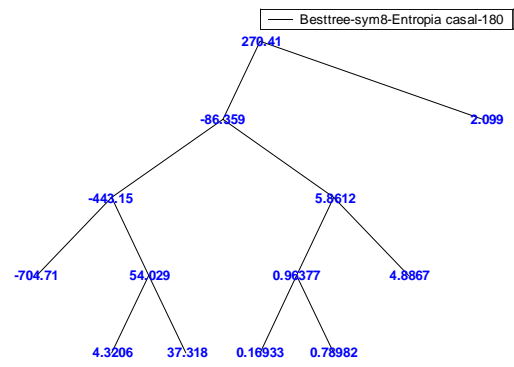


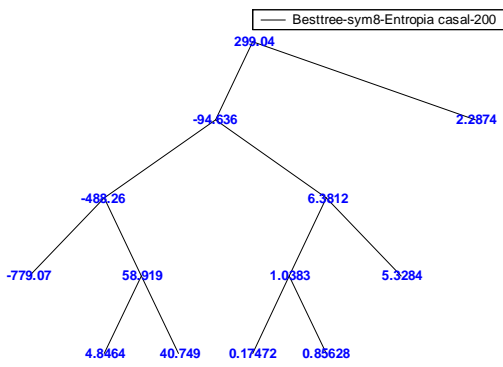
Figura A.27 Best Basis Casal com a função sym8 para a) $f_0=80\text{Hz}$ b) $f_0=100\text{Hz}$ c) $f_0=120\text{Hz}$ d) $f_0=140\text{Hz}$.



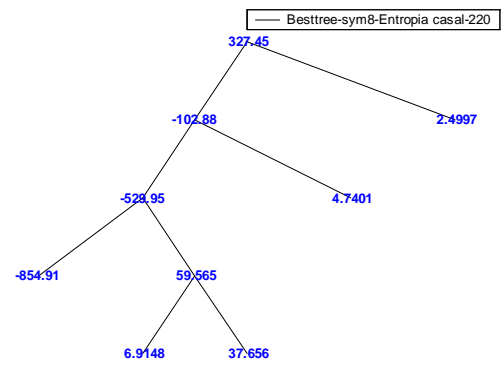
a)



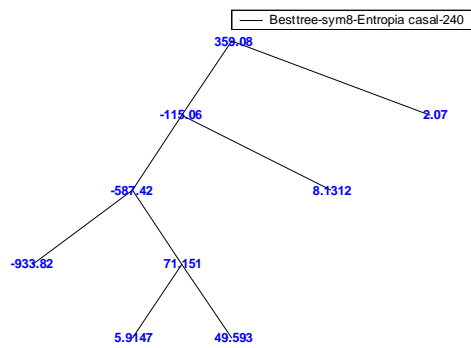
b)



c)



d)



e)

Figura A.28 Best Basis Casal com a função sym8 para a) $f_0=160\text{Hz}$ b) $f_0=180\text{Hz}$ c) $f_0=200\text{Hz}$ d) $f_0=220\text{Hz}$ e) $f_0=240\text{Hz}$.

Best Basis de Sinais Sintéticos Patológicos

Nas figuras A.29 a A.35, encontram-se as *Best Basis* dos sinais sintéticos gerados a partir do LPC de quatro patologias para as funções Wavelet em teste, mas mantendo a frequência fundamental constante, conforme os itens 4.3.4 e 4.3.5 e utilizando a função custo entropia.

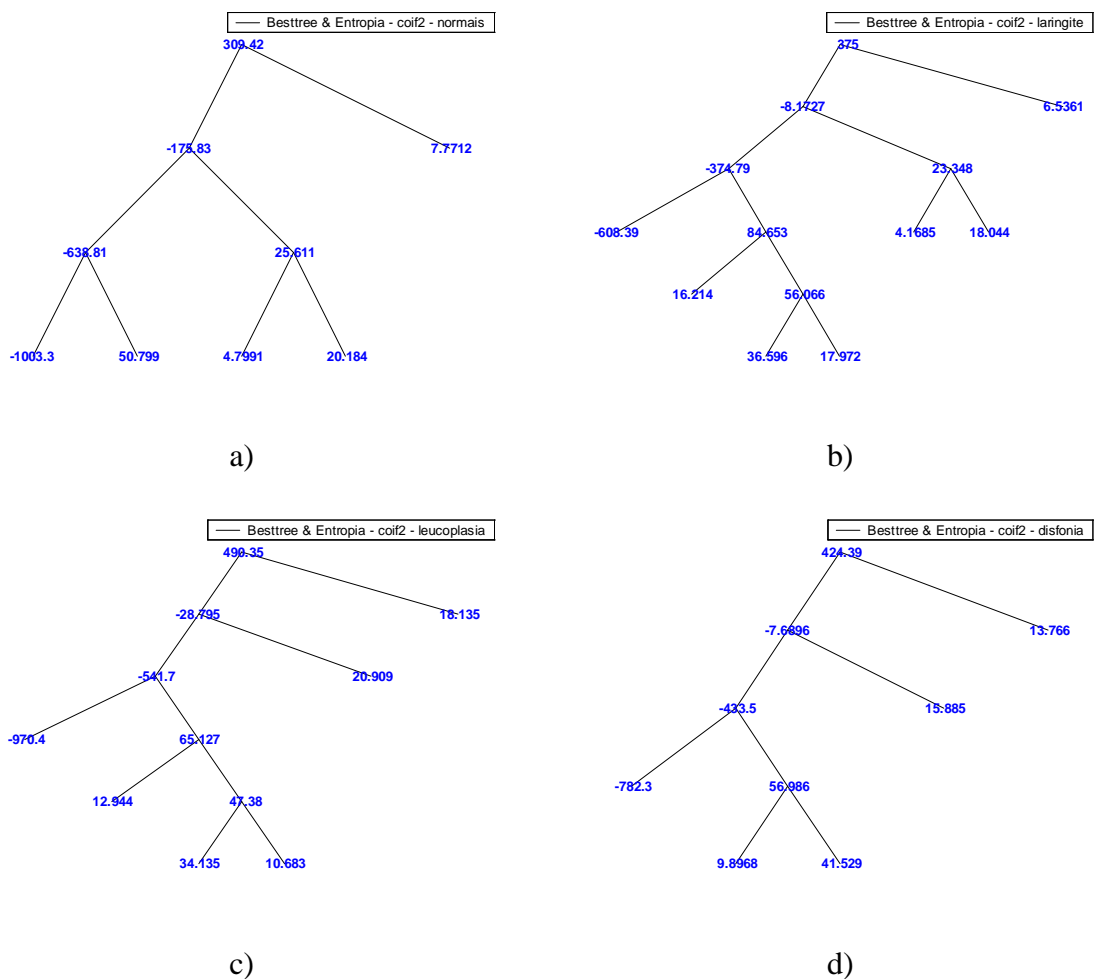


Figura A.29 Best Basis LPC patológicos com a função coif2 para a) normais b) laringite c) leucoplasia d) disфонia.

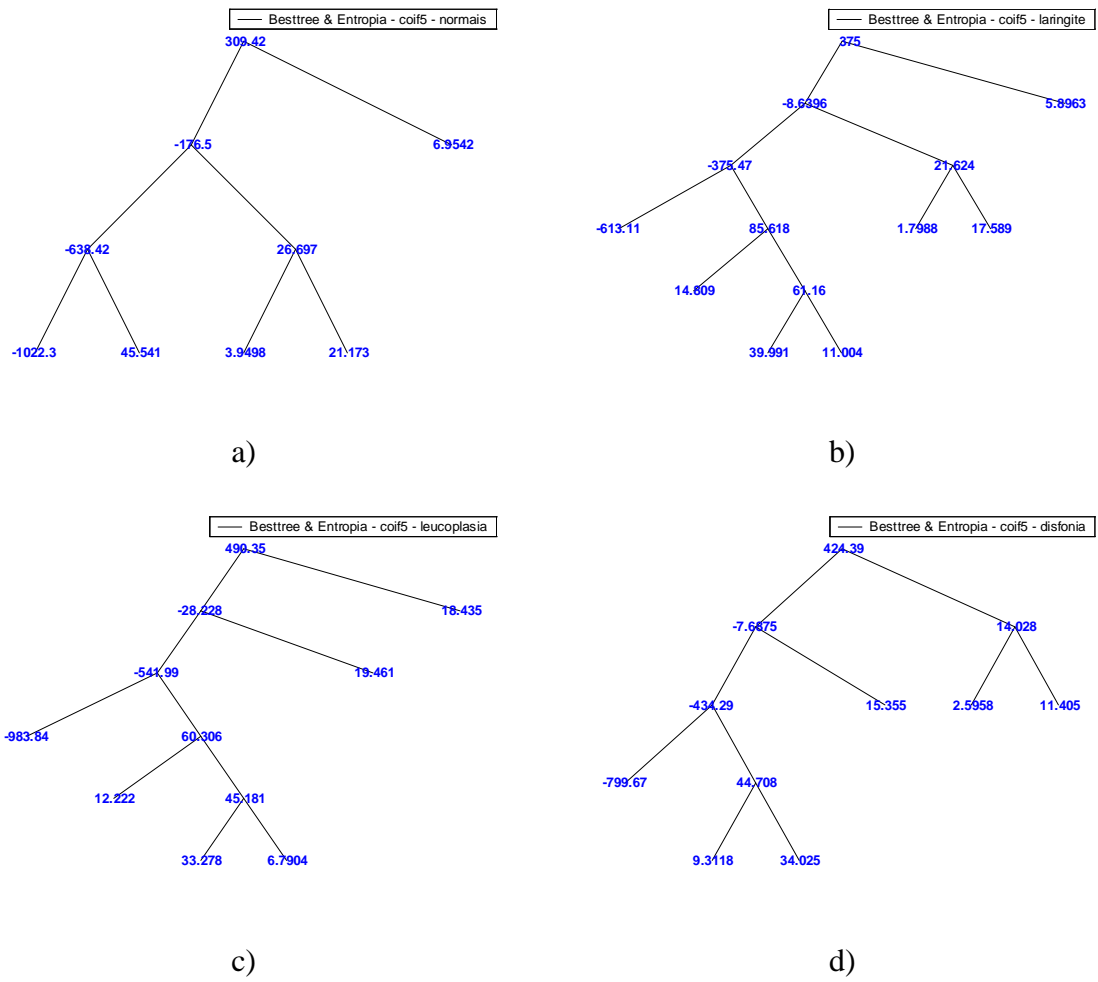


Figura A.30 Best Basis LPC patológicos com a função `coif5` para a) normais b) laringite c) leucoplasia d) disфонia.

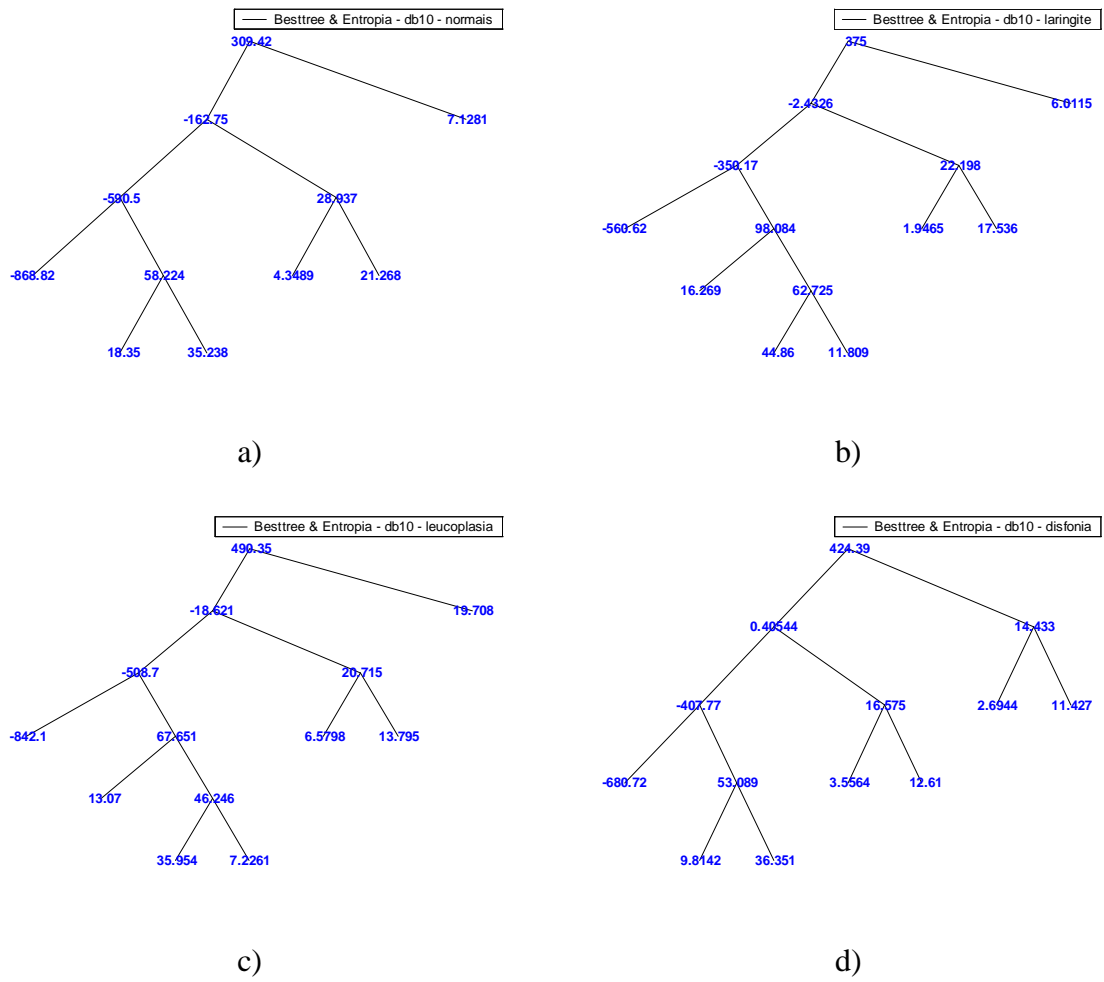


Figura A.31 Best Basis LPC patológicos com a função db10 para a) normais b) laringite c) leucoplasia d) disfonia.

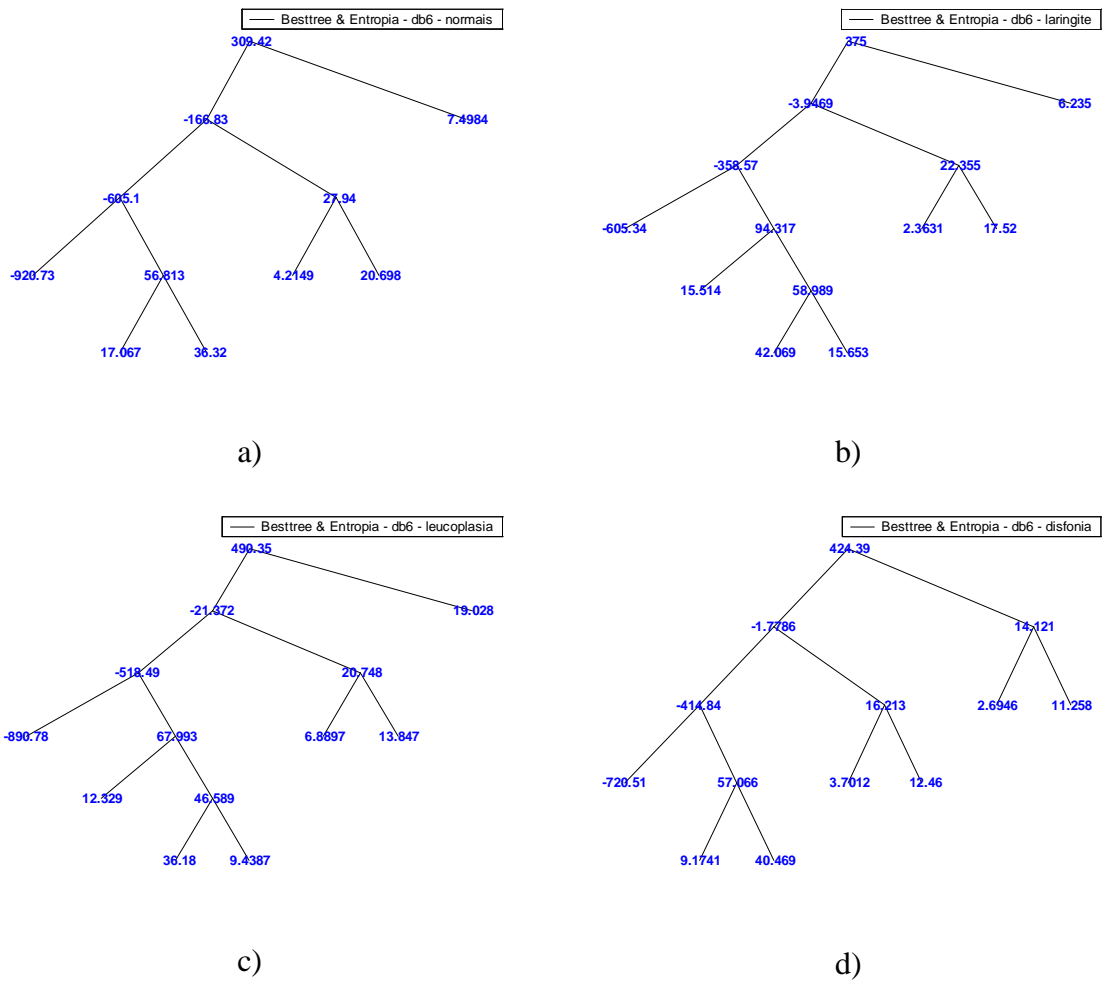


Figura A.32 Best Basis LPC patológicos com a função db6 para a) normais b) laringite c) leucoplasia d) disфонia.

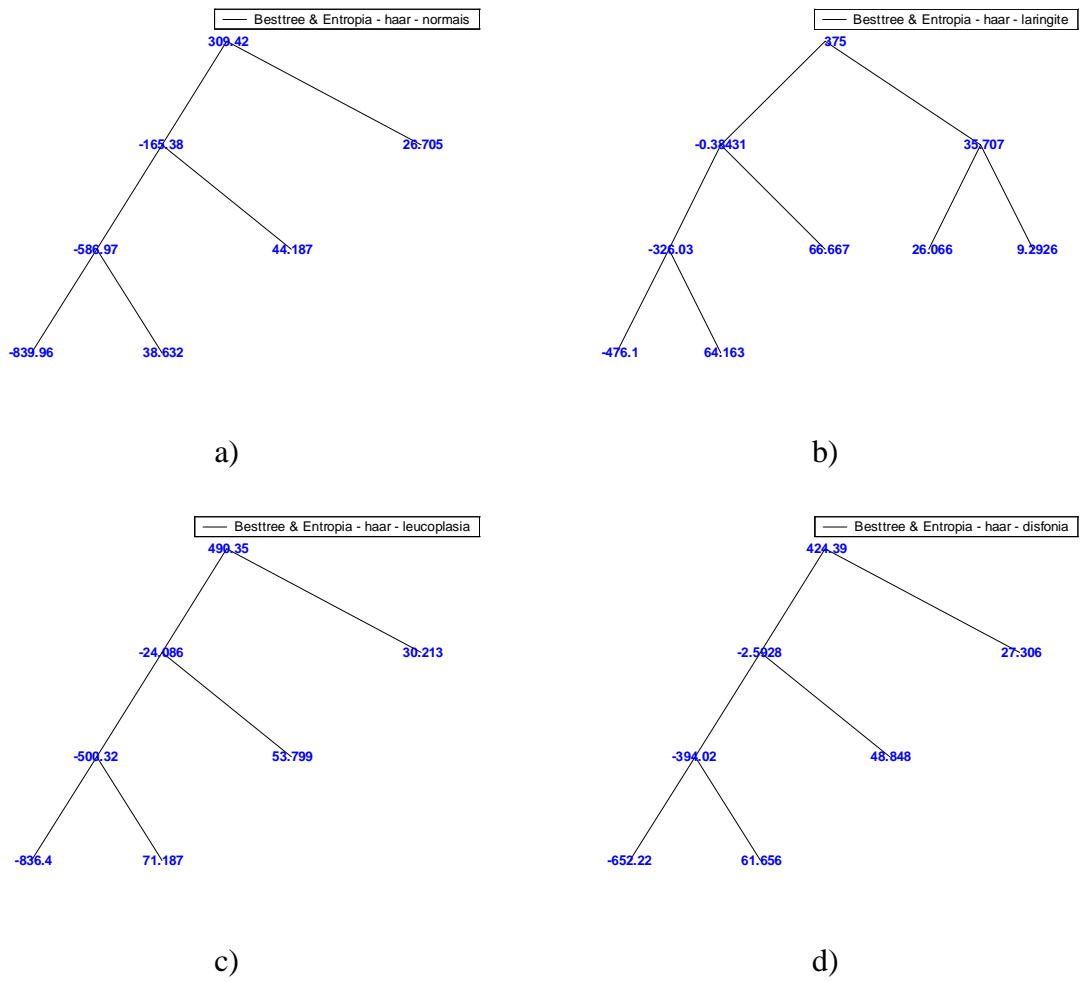


Figura A.33 Best Basis LPC patológicos com a função haar para a) normais b) laringite c) leucoplasia d) disfonia.

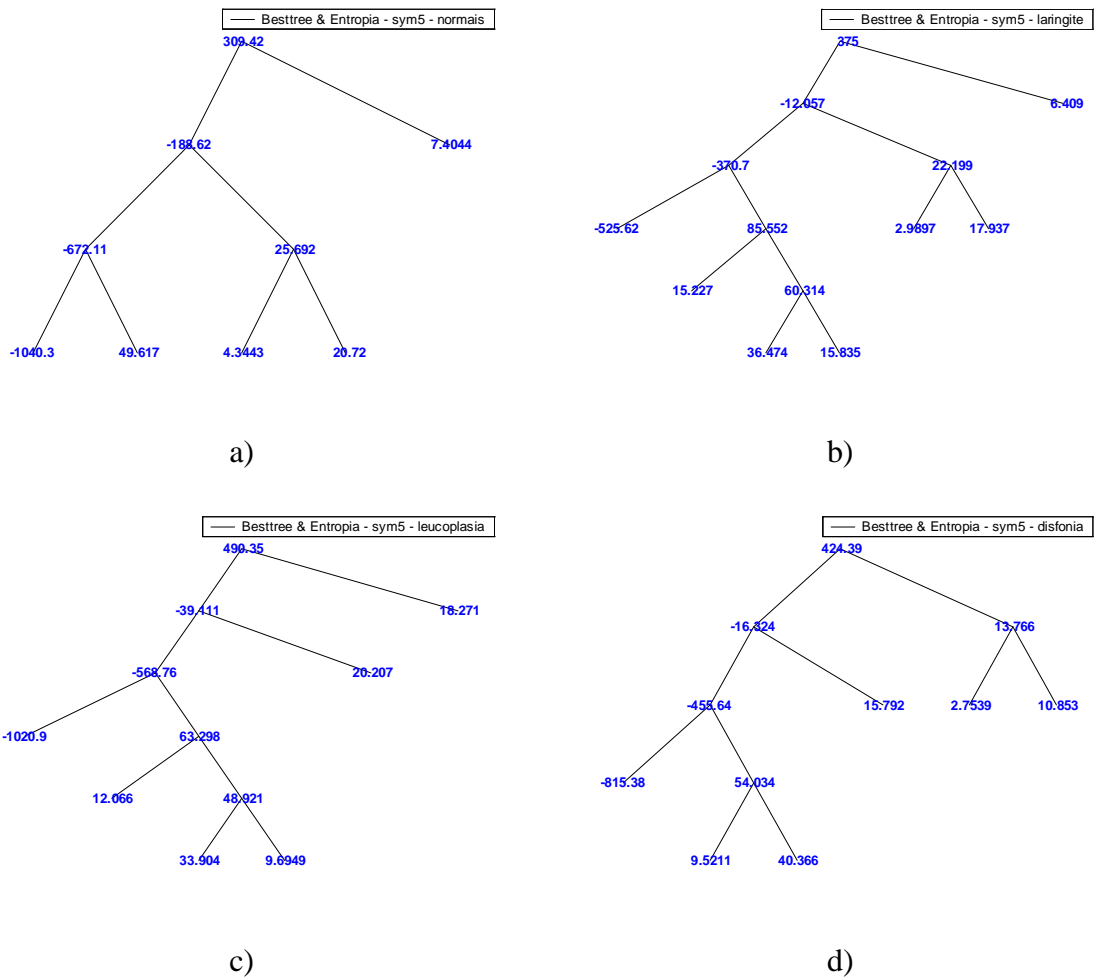


Figura A.34 Best Basis LPC patológicos com a função *sym5* para a) normais b) laringite c) leucoplasia d) disfonia.

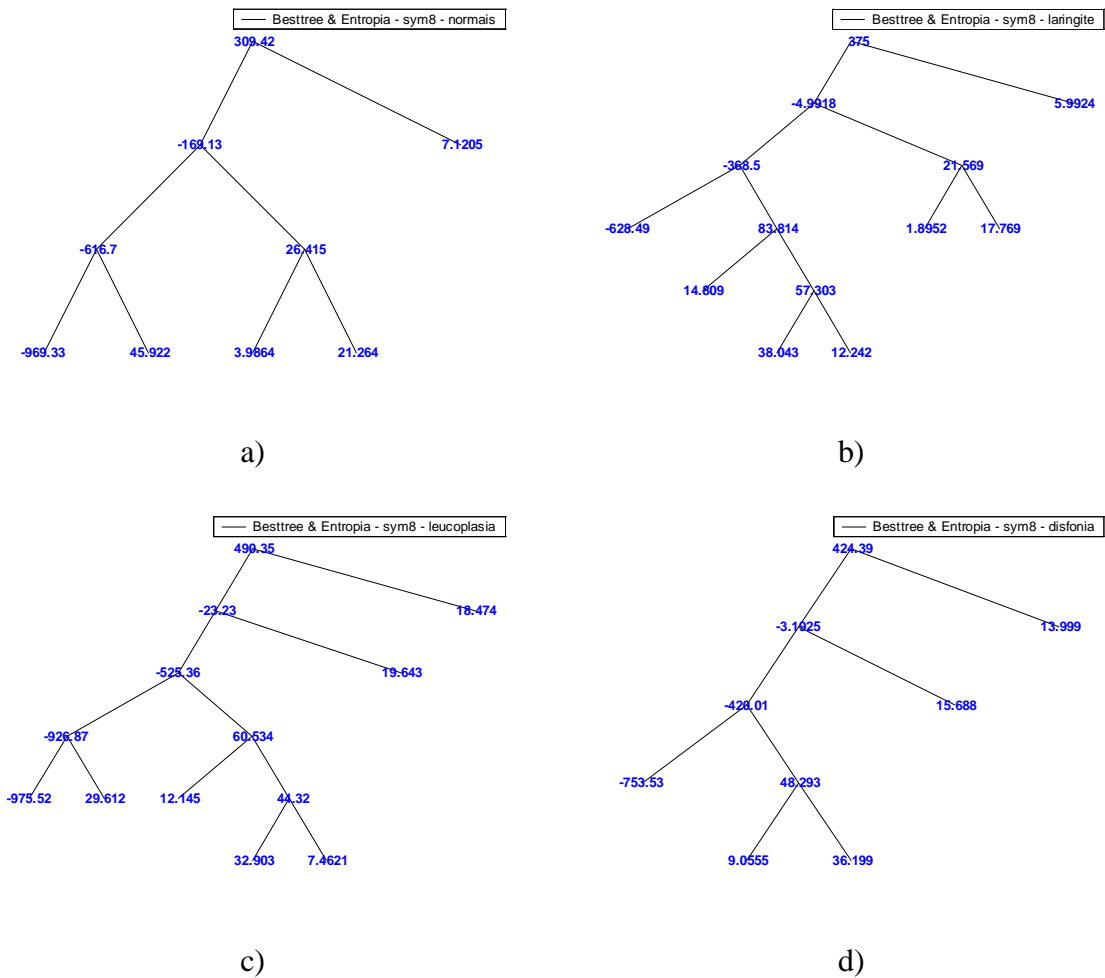


Figura A.35 Best Basis LPC patológicos com a função *sym8* para a) normais b) laringite c) leucoplasia d) disfonia.

Resultados dos Sinais Sintéticos com *Jitter* e *Shimmer*

Nas figuras A.36 a A.39, encontram-se as *Best Basis* dos LPC dos homens e mulheres com *Jitter* e *Shimmer* geradas conforme o item 4.2.3, com a função Wavelet *sym5* e a função custo entropia.

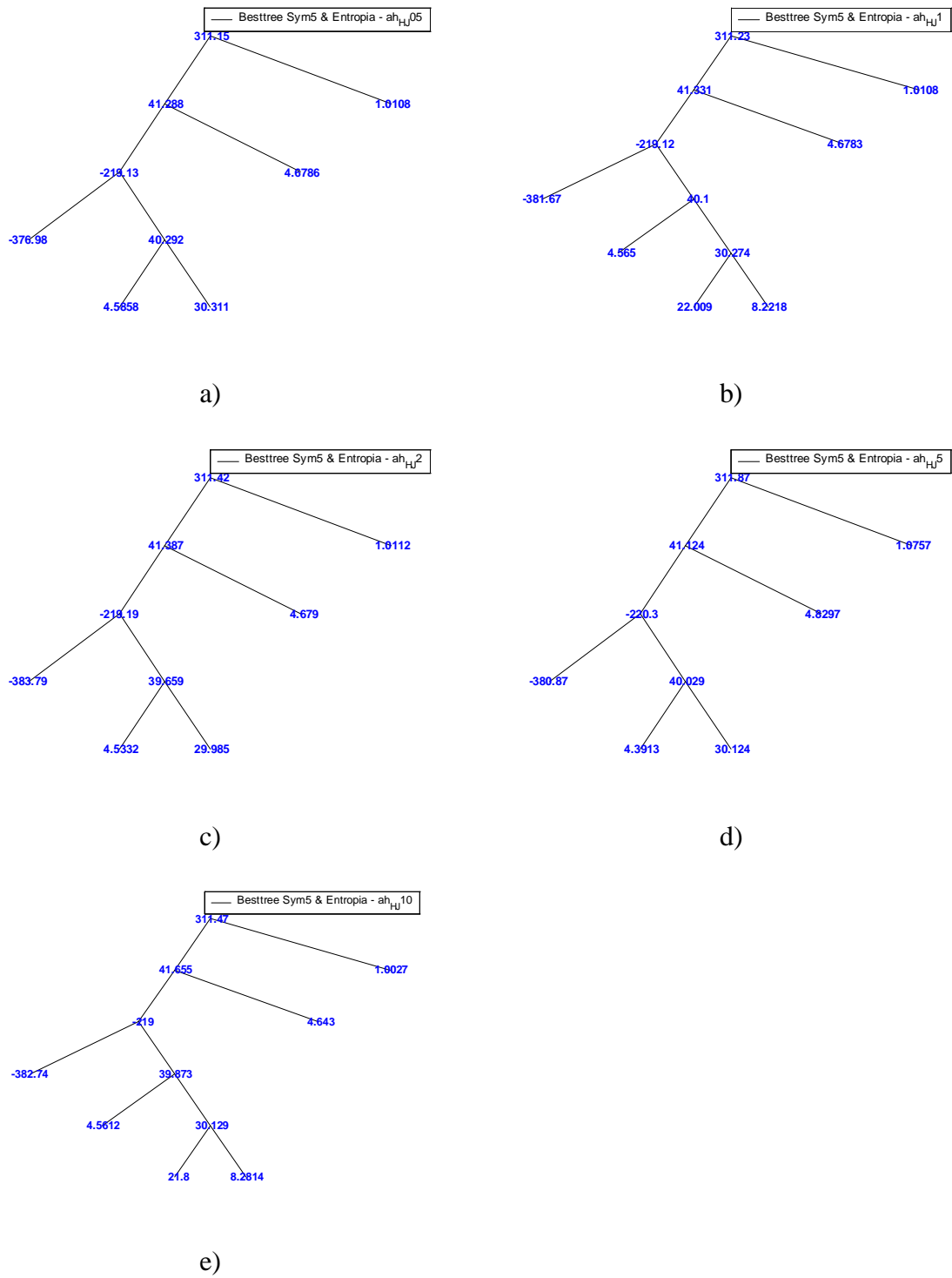


Figura A.36 Best Basis com função sym5 para LPC Homem com Jitter de a) 0,5% b) 1% c) 2% d) 5% e) 10%.

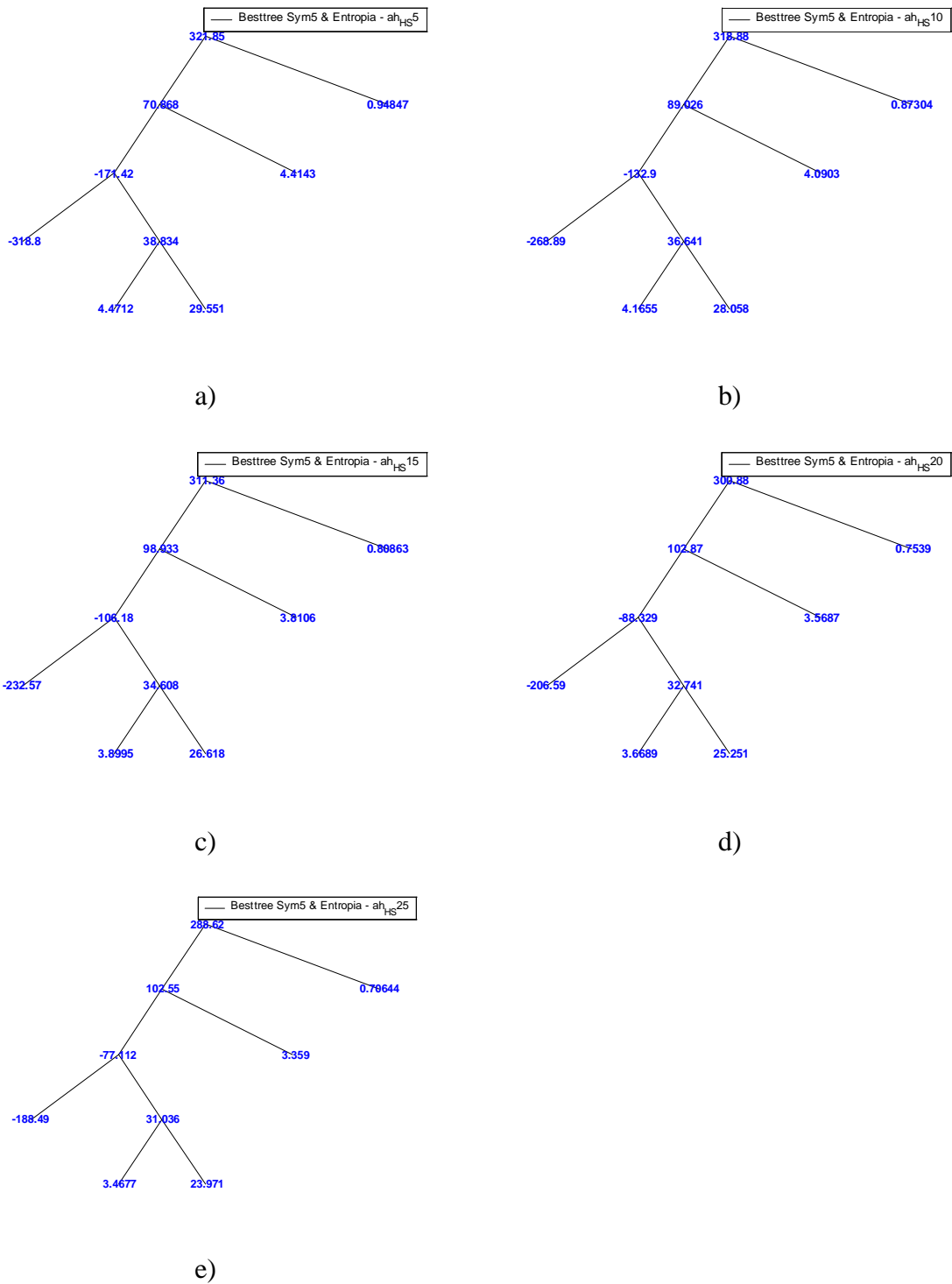


Figura A.37 Best Basis com função sym5 para LPC Homem com Shimmer de

a) 5% b) 10% c) 15% d) 20% e) 25%.

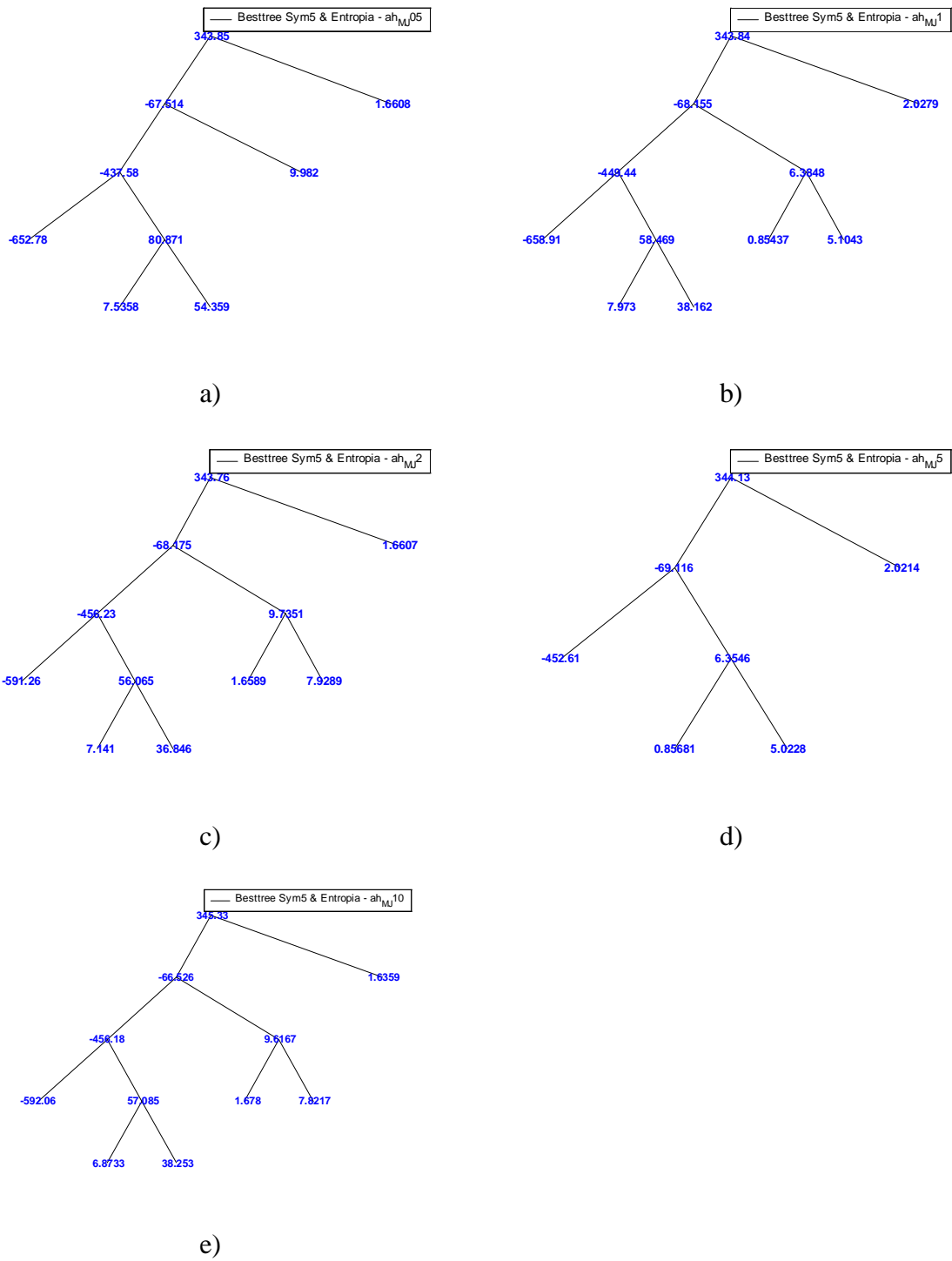


Figura A.38 Best Basis com função sym5 para LPC Mulher com Jitter de a) 0,5% b) 1% c) 2% d) 5% e) 10%.

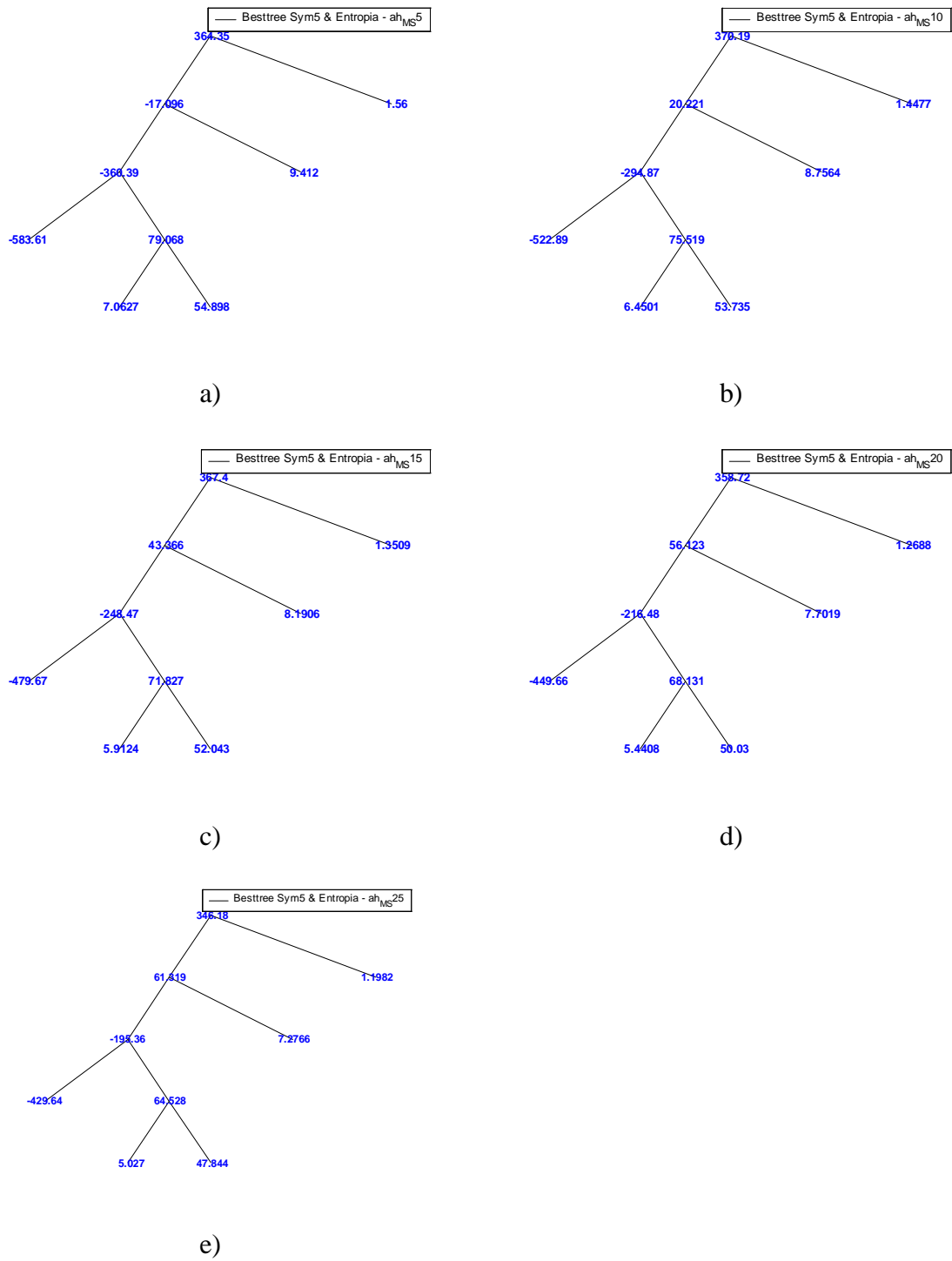


Figura A.39 Best Basis com função sym5 para LPC Mulher com Shimmer de a) 5% b) 10% c) 15% d) 20% e) 25%.