

EVERTHON SILVA FONSECA

**WAVELETS, PREDIÇÃO LINEAR E LS-SVM APLICADOS NA
ANÁLISE E CLASSIFICAÇÃO DE SINAIS DE VOZES
PATOLÓGICAS**

Tese apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica.

Área de Concentração: Processamento de Sinais e Instrumentação.

Orientador: **Prof. Dr. José Carlos Pereira**

São Carlos
2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Dedicatória

Aos meus pais Luiz Roberto da Fonseca e Cleonice Silva da Fonseca.

AGRADECIMENTOS:

Primeiramente a Deus, pela vida, pelo amor, paz e força entregues a mim durante cada momento destes anos de trabalho.

Ao Prof. Dr. José Carlos Pereira, pela amizade, excelente orientação e por todo apoio técnico, acadêmico e pessoal.

Aos meus pais, Luiz e Cleonice; meus irmãos, Abel e Roger, meus avós Ilka e Abílio (*in memoriam*) e todos os familiares, tios(as), primos(as), pela presença constante e toda atenção e apoio prestados em cada passo de minha vida.

A todos os professores do Departamento de Engenharia Elétrica da Escola de Engenharia de São Carlos (EESC – USP), em especial ao Prof. Dr. Carlos Dias Maciel pelo apoio técnico, acadêmico e pela paciência de cada dia e também ao Professor Edison Gesualdo, paraninfo de minha turma da graduação, por todo apoio moral e todas as conversas.

A todos os funcionários do Departamento de Engenharia Elétrica, Marisa, Denise, Rosane, Paulo, José Carlos, Jussara, Aura, Cida, Vladimir, Xaraba e todos os demais, pelo excelente trabalho prestado durante todos estes anos na USP.

Aos amigos dos Laboratórios de Processamento de Sinais Biológicos (LPSB), Instrumentação Microeletrônica (LIM) e Telecomunicações do Departamento de Engenharia Elétrica da EESC – USP, Mori, Parê, Marcelo, Arlindo, Eugenia, Vinicius, Daiane, Rafael, Julian, Soledad, Edwin, Elton, Paulo, Benini, Marina, Betiol, Pizolato,

Helvécio e todos os demais cujos nomes não estiverem aqui, mas que muito me auxiliaram com cada uma das conversas, discussões, cumprimentos e aperto de mãos de todos os dias.

A todos os amigos de São José do Rio Preto e região, em especial ao Eduardo Meireles, por todas as conversas e forte apoio moral, e também ao Hugo, Helder, Matheus Gabarrão, Nelson, Daniel, Fernando, Thiago Vasconcelos e todos os demais cujos nomes não estiverem aqui, mas que, de alguma forma, deram-me força e muito me apoiaram durante este Doutorado.

À Professora Dra. Abeer Alwan e todos os amigos do *Speech Processing and Auditory Perception Laboratory* (SPAPL) do Departamento de Engenharia Elétrica da Universidade da Califórnia, Los Angeles (UCLA), Markus, Yen, Chigen, Jonas e Panchapagesan, pela importante participação neste projeto, no auxílio aos trabalhos referentes aos artigos publicados nos EUA e todo apoio prestado durante o período de pesquisa no exterior.

À USP e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por toda a assessoria e financiamento deste projeto, tendo me prestado um apoio de altíssimo nível para o aperfeiçoamento deste trabalho.

*“A verdade pode ser encontrada com pura lógica,
mesmo que as verdades lógicas e puros nos pareçam, muitas vezes,
inalcançáveis.”*

RESUMO

FONSECA, Everthon Silva (2008). *DWT, LPC e LS-SVM Aplicados na Análise e Classificação de Sinais de Vozes Patológicas*. 107 p. Tese de Doutorado. Escola de Engenharia de São Carlos, Universidade de São Paulo (EESC - USP), São Carlos, Brasil.

Neste trabalho, foram utilizadas as vantagens da ferramenta matemática de análise temporal e espectral, a Transformada Wavelet Discreta (DWT), além dos coeficientes de predição linear (LPC) e do algoritmo de inteligência artificial, *Least Squares Support Vector Machines* (LS-SVM), para aplicações em análise de sinais de voz e classificação de vozes patológicas. Inúmeros trabalhos na literatura têm demonstrado o grande interesse existente por ferramentas auxiliares ao diagnóstico de patologias da laringe. Os componentes da DWT forneceram parâmetros de medida para a análise e classificação das vozes patológicas, principalmente aquelas provenientes de pacientes com edema de Reinke e nódulo nas pregas vocais. O banco de dados com as vozes patológicas foi obtido do Departamento de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (FMRP-USP). Utilizando-se o algoritmo de reconhecimento de padrões, LS-SVM, mostrou-se que a combinação dos componentes da DWT de Daubechies com o filtro LP inverso levou a um classificador de bom desempenho alcançando mais de 90% de acerto na classificação das vozes patológicas.

Palavras-Chave: Transformada wavelet, filtro inverso de predição linear, vozes patológicas, classificador *support vector machines*.

ABSTRACT

FONSECA, Everthon Silva (2008). *DWT, LPC e LS-SVM Applied for Analysis and Classification of Pathological Voice Signals*. 107 p. PhD. Thesis. Escola de Engenharia de São Carlos, Universidade de São Paulo (EESC - USP), São Carlos, Brasil.

The main objective of this work was to use the advantages of the time-frequency analysis mathematical tool, Discrete Wavelet Transform (DWT), besides the linear prediction coefficients (LPC) and the artificial intelligence algorithm, *Least Squares Support Vector Machines* (LS-SVM), for applications in voice signal analysis and classification of pathological voices. A large number of works in the literature has been shown that there is a great interest for auxiliary tools to the diagnosis of laryngeal pathologies. DWT components gave measure parameters for the analysis and classification of pathological voices, mainly that ones from patients with Reinke's edema and nodule in the vocal folds. It was used a data bank with pathological voices from the Otolaryngology and the Head and Neck Surgery sector of the Clinical Hospital of the Faculty of Medicine at Ribeirão Preto, University of São Paulo (FMRP-USP), Brazil. Using the automatic learning algorithm applied in pattern recognition problems, LS-SVM, results have showed that the combination of Daubechies' DWT components and inverse LP filter leads to a classifier with good performance reaching more than 90% of accuracy in the classification of the pathological voices.

Keywords: Discrete Wavelet Transform, linear prediction inverse filter, pathological voices and support vector machines classifier.

LISTA DE FIGURAS

Figura 1 – Trato Vocal.....	30
Figura 2 – Banco de Filtros de Análise da DWT.....	51
Figura 3 – Diferentes tipos de análise de sinais.....	53
Figura 4 – Algumas funções wavelets.....	54
Figura 5 – Margem(d):- mínima distância possível entre qualquer ponto do treinamento e o hiperplano de classificação.....	60
Figura 6 – Banco de Filtros da DWT evidenciando os componentes utilizados no algoritmo de classificação deste trabalho.....	66
Figura 7 – Diagrama em Blocos simplificado com os passos do algoritmo utilizado para a identificação e classificação das vozes patológicas.....	69
Figura 8 – Componentes de detalhe D_2 da DWTdb10 normalizados pelo valor de pico dos sinais de voz normal e patológica do banco de dados.....	72
Figura 9 – Valores RMS do componente de detalhe D_2 da DWTdb10 normalizado pelo módulo de seu valor de pico para cada um dos 60 sinais de voz. Os pontos * e o correspondem às vozes patológicas e normais, respectivamente.....	76
Figura 10 – Valores RMS do componente de detalhe D_2 da DWTdb10 normalizado pelo módulo de seu valor de pico após o filtro LP inverso de ordem 100 para cada um dos 60 sinais de voz. Os pontos * e o correspondem às vozes patológicas e normais, respectivamente.....	76
Figura 11 – Densidades espectrais de potência (DSP) do componente de detalhe normalizado RnD_2 da DWTdb10 de 4 sinais de voz normal, antes ($D2n$) e após o filtro LP inverso ($D2fn$).....	78
Figura 12 – Densidades espectrais de potência (DSP) do componente de detalhe normalizado RnD_2 da DWTdb10 de 4 sinais de voz patológica, antes ($D2n$) e após o filtro LP inverso ($D2fn$).....	79

LISTA DE FIGURAS

- Figura 13 – Os dois primeiros gráficos correspondem aos componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes normais antes e após a aplicação do filtro digital inverso LP e os dois últimos correspondem aos componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes normais com ruído adicional ao sinal antes e após a aplicação do filtro digital inverso LP. 81
- Figura 14 – Componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes normais antes e após a aplicação do filtro digital inverso LP. .. 82
- Figura 15 – Componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes patológicas antes e após a aplicação do filtro digital inverso LP. 83
- Figura 16 – Médias dos valores RMS_{RD_2} de todas as vozes normais (traço pontilhado) e de todas as vozes patológicas (traço contínuo) de nosso banco de dados para cada ordem N do filtro LP inverso. 84
- Figura 17 – Gráfico com os valores RMS dos componentes D_2 da DWTdb10 normalizados pelo valor de pico após o uso do filtro LP inverso para as vozes normais e patológicas..... 85
- Figura 18 – Visualização gráfica do classificador LS-SVM com 2 atributos de entrada..... 86
- Figura 19 – Esboços gráficos da classificação com os valores RMS dos componentes já filtrados pelo LP inverso, RD_1 (eixo x) e RD_2 (eixo y), para os kernels linear (em (A)), MLP (em (B)) e RBF (em (C)) respectivamente. Os pontos * e □ foram usados para o treinamento do classificador e eles são provenientes, respectivamente, de vozes normais e patologicamente afetadas. Os pontos ○ e ● foram usados para o teste ou validação do classificador e são, respectivamente, provenientes de vozes normais e patologicamente afetadas. 89

LISTA DE FIGURAS

- Figura 20 – Esboços gráficos da classificação com os valores RMS dos componentes já filtrados pelo filtro LP inverso, RD_1 (eixo x) e RD_2 (eixo y), para vozes normais e vozes de pacientes com edema de Reike em (A) e para vozes normais e vozes de pacientes com nódulo nas pregas vocais em (B). Os pontos * e □ foram usados para o treinamento do classificador e eles são provenientes, respectivamente, de vozes normais e patologicamente afetadas. Os pontos ○ e ● foram usados para o teste ou validação do classificador e são, respectivamente, provenientes de vozes normais e patologicamente afetadas.92
- Figura 21 – Componentes reconstruídos de aproximação e detalhe nos níveis 5 e 6 obtidos da decomposição da DWTdb10 de um sinal de voz oriundo de um paciente com nódulo nas pregas vocais.....94
- Figura 22 – Valores de Jitter dos componentes wavelet RD_6 , RA_5 ou RA_6 de acordo com aquele cuja frequência fundamental fosse a mais similar possível com a frequência fundamental do respectivo sinal de voz decomposto...96

LISTA DE TABELAS

Tabela 1 – Patologias, suas origens e sintomas acústicos.....	30
Tabela 2 – Comparação entre os métodos propostos neste trabalho (primeira e antepenúltima linha) e outros com objetivos similares encontrados na literatura.....	36
Tabela 3 – Famílias Wavelet utilizadas (FW), tamanhos de suporte dos filtros correspondentes (TS) e as respectivas diferenças relativas percentuais entre a média aritmética dos valores RMS_{RD2} das vozes normais e das vozes patológicas (DRNP).....	73
Tabela 4 – Acertos na Classificação para cada ordem N do filtro LP inverso.....	75
Tabela 5 – Conjunto de teste escolhido aleatoriamente para o classificador.	87
Tabela 6 – Porcentagem de acerto nos dados de teste ou validação do classificador.	90
Tabela 7 – Acertos na Classificação do LS-SVM (Normal x Nódulo; Normal x Edema; Normal x Patológica).....	93
Tabela 8 – Acertos na Classificação do LS-SVM para a distinção entre as duas patologias vocais (Nódulo e Edema).	97

LISTA DE SIGLAS

APQ	- <i>Amplitude Perturbation Quotient</i>
AVI	- <i>Amplitude Variability Index</i>
BR	- <i>Breathiness Ratio</i>
CWT	- <i>Continuous Wavelet Transform</i>
dbN	- <i>Função wavelet de Daubechies de ordem N</i>
DPF	- <i>Directional Perturbation Factor</i>
DWT	- <i>Discrete Wavelet Transform</i>
DWTdbN	- <i>Daubechies' Discrete Wavelet Transform (order N)</i>
HFPR	- <i>High Frequency Power Relation</i>
HNR	- <i>Harmonics-to-Ration Ratio</i>
JF	- <i>Jitter Factor</i>
JM	- <i>Jitter Médio</i>
JR	- <i>Jitter Ratio</i>
LDB	- <i>Local Discriminant Basis</i>
LP	- <i>Linear Prediction</i>
LPC	- <i>Linear Prediction Coefficients</i>
LS-SVM	- <i>Least Squares Support Vector Machines</i>
LTAS	- <i>Long Time Average Spectrum</i>
NHR	- <i>Noise-to-Harmonics Ratio</i>
NNE	- <i>Normalized Noise Energy</i>
PFF	- <i>Perturbation Frequency Factor</i>
PPQ	- <i>Period Perturbation Quotient</i>
PVI	- <i>Period Variability Index</i>

LISTA DE SIGLAS

RAP - *Relative Average Perturbation*

SNL - *Spectral Noise Level*

STFT - *Short Time Fourier Transform*

SVM - *Support Vector Machines*

LISTA DE SÍMBOLOS

SÍMBOLOS GERAIS:

- \vec{A}_j - Componentes de aproximação (baixas frequências) da DWT
- \vec{D}_j - Componentes de detalhe (altas frequências) da DWT
- cA_j - Coeficientes de aproximação da DWT
- cD_j - Coeficientes de detalhe da DWT
- $L^2(\mathbb{R})$ - Espaço das funções que têm quadrado Lesbague integrável
- \mathbb{R} - Conjunto dos Números Reais
- \mathbb{Z} - Conjunto dos Números Inteiros

LISTA DE SÍMBOLOS

SÍMBOLOS GREGOS

ψ - Função wavelet ou wavelet mãe

ϕ - Função de escala ou wavelet pai

$\delta_{k,l}$ - Delta de Kronecker

SUMÁRIO

1. INTRODUÇÃO E OBJETIVOS	23
1.1. INTRODUÇÃO	23
1.2. OBJETIVOS	26
2. REVISÃO BIBLIOGRÁFICA	29
2.1. VOZES PATOLÓGICAS.....	29
2.2. CLASSIFICAÇÃO DE SINAIS DE VOZES PATOLÓGICAS.....	34
2.3. MEDIDAS DE PERTURBAÇÃO	37
2.4. MEDIDAS ESPECTRAIS DE RUÍDO.....	39
2.5. TÉCNICAS QUE UTILIZAM A TRANSFORMADA WAVELET	42
3. FUNDAMENTOS TEÓRICOS	45
3.1. PREDIÇÃO LINEAR	45
3.2. A TRANSFORMADA WAVELET	47
3.2.1. Relações Matemáticas da Decomposição Wavelet.....	47
3.2.2. A Análise Wavelet.....	51
3.2.3. As Famílias Wavelet.....	53
3.3. <i>SUPPORT VECTOR MACHINES</i> (SVM).....	58
3.3.1. SVM para Reconhecimento de Padrões.....	58
3.3.2. LS-SVM para Reconhecimento de Padrões	61
4. MATERIAIS E MÉTODOS	63
4.1. BANCO DE DADOS.....	64
4.2. METODOLOGIA E ALGORITMO IMPLEMENTADO	65
5. RESULTADOS E DISCUSSÕES	71
6. CONCLUSÃO E SUGESTÕES	99
7. REFERÊNCIAS BIBLIOGRÁFICAS	101

Capítulo 1

INTRODUÇÃO E OBJETIVOS

1.1. INTRODUÇÃO

Pode-se dizer que, sob o ponto de vista sistêmico, o sinal de voz carrega toda a informação intrínseca ao aparelho fonador, possibilitando-se obter a partir destes sinais cada uma das características relacionadas ao trato vocal do paciente. Sabe-se que é possível utilizar o processamento discreto no tempo de sinais de voz (DELLER JR.; PROAKIS; HANSEN, 1993; DENG; SHAUGHNESSY, 2003) com o objetivo de detectar características acústicas que diferenciam sinais de voz normal e patológica. As patologias do trato glotal são usualmente identificadas a partir da utilização de padrões acústicos perceptuais como a aspereza, a soprosidade e a rouquidão (BLESS, 1991; BAKEN, 1996; HECKER; KRUEL, 1970).

No entanto, devido à complexa estrutura de funcionamento do sistema produtor da fala, patologias com características soprosas podem ser confundidas com as perceptualmente definidas como roucas (ISSHIKI, 1969). A turbulência no fluxo glotal, resultante do mau funcionamento das pregas vocais, pode ser quantificada pelo ruído nos componentes espectrais da fala (ROSA; PEREIRA; GRELLET, 2000). Patologias causadas pelo incompleto fechamento da glote, como os nódulos nas pregas vocais, são usualmente associadas com ruído em alta

freqüência (GRANQVIST; HAMMARBERG, 2003; MARTINEZ; RUFFINER, 2000). Nesse trabalho, portanto, foi realizada uma análise dessas características em alta freqüência a partir dos componentes wavelet para distinguir as vozes normais daquelas patologicamente afetadas. .

Foram utilizadas, neste trabalho, duas ferramentas de grande aplicação em processamento de sinais, a Transformada Wavelet Discreta (DWT) (BRUCE, 1998; DAUBECHIES, 1992; CHUI, 1992; MALLAT, 1999; KAISER, 1994) e o filtro de predição linear (LP) inverso (ANANTHAPADMANABHA; YEGNANARAYANA, 1979; DENG; SHAUGHNESSY, 2003; GRAY; MARKEL, 1976). Utilizando-se valores RMS dos componentes de detalhe da DWT de Daubechies (DAUBECHIES; GILBERT, 1997) em determinada escala foram obtidos padrões que diferenciaram as vozes normais de vozes provenientes de pacientes com nódulo nas pregas vocais, identificando-as com até 85% de acerto ao utilizar um valor de limiar entre as duas classes (FONSECA et al., 2004).

Ao ampliar o banco de dados e aplicar o algoritmo para o reconhecimento automático de padrões *Least Squares Support Vector Machines* (LS-SVM) (VAPINIK, 1995; SUYKENS; VANDEWALLE, 1999; CRISTIANINI; SHAW-TAYLOR, 2000; BURGESS, 1999), mostrou-se a possibilidade de se treinar convenientemente o algoritmo com o kernel *Radial Basis Function* (RBF) e obter, para determinado grupo de teste, mais de 95% de acerto na classificação usando os mesmos valores RMS anteriores (FONSECA et al., 2005). Estudou-se ainda a utilização de diferentes funções kernel (Linear, *Radial Base Function* (RBF) e *Multi-layer Perceptron* (MLP)) no algoritmo LS-SVM para a identificação das vozes patológicas. Além disso, treinou-se também o classificador com todas as

combinações possíveis dos valores RMS de 4 diferentes componentes de detalhe D_1 , D_2 , D_3 e D_4 da Transformada Wavelet Discreta de Daubechies (DWTdb), especificando-se informações de diferentes escalas de frequência do sinal (com a taxa de amostragem de 22050 Hz, estes componentes analisam respectivamente as frequências de 5512,5 a 11025 Hz, 2756,25 a 5512,5Hz, 1378,125 a 2756,25 e de 689,06 a 1378,125Hz) para cada uma das 3 funções kernel do algoritmo LS-SVM comentadas anteriormente (FONSECA et al., 2007a).

Demonstra-se assim a possibilidade de se estabelecer um novo parâmetro de classificação, chamado DWT-RMS (FONSECA et al., 2007b), para identificar as vozes patológicas, principalmente aquelas provenientes de pacientes com nódulos nas pregas vocais. Mostrou-se também um estudo descritivo do ajuste da ordem estatisticamente mais adequada dos coeficientes deste filtro para a maior separabilidade entre as classes, bem como qual das funções wavelet e qual a escala frequencial são as mais convenientes. Isto pode ser aplicado em conjunto com outras técnicas existentes de modo a melhorar a classificação, como em (HADDJITODOROV; BOYANOV; TESTON, 2000), ou pode constituir unicamente um classificador. É importante salientar que este algoritmo possui uma ordem de complexidade computacional linear com relação ao comprimento do sinal e se baseia, para a análise dos sinais de voz normal e patológica, unicamente na combinação da DWT e do filtro LP inverso.

M. Rosa, J. C. Pereira e M. Grellet (2000) mostram que a melhor medida acústica indicada para a discriminação de patologias na voz é o *Jitter* (perturbação em torno da frequência fundamental) (SCHOENTGE; GUCHTENEERE, 1995). Dessa maneira, foram também encontrados parâmetros que utilizam a medida do

jitter dos componentes da DWT em escalas convenientes de frequência para a classificação entre duas diferentes patologias vocais, o edema de Reinke e o nódulo nas pregas vocais. Utilizando-se o classificador LS-SVM foi possível obter mais de 82% de acerto nesta classificação. Os resultados deste estudo foram também submetidos em artigo enviado para a análise dos revisores da revista IEEE – EMBS (*Engineering in Medicine and Biology Magazine*) com o título “*Wavelet Analysis and LS-SVM for Classification of Pathological Voice Signals from patients with Reinke’s Edema and Nodule in Vocal Folds*”.

Esta tese está organizada da seguinte maneira:

No subitem seguinte deste capítulo são apresentados os principais objetivos do trabalho. No Capítulo 2, são descritos os conceitos teóricos básicos e a pesquisa bibliográfica realizada. No Capítulo 3, será feita uma descrição dos fundamentos teóricos e das ferramentas matemáticas usadas nos algoritmos implementados neste trabalho. No Capítulo 4, são apresentados os materiais e os algoritmos utilizados, bem como a metodologia e a descrição dos testes realizados. No Capítulo 5, estão os resultados e a discussão dos mesmos. Posteriormente, no Capítulo 6, são apresentados os comentários finais e a conclusão com possíveis sugestões para trabalhos futuros. Finalmente, têm-se as referências bibliográficas utilizadas.

1.2. OBJETIVOS

Os principais objetivos deste trabalho são:

- Obter um parâmetro de análise do ruído das vozes patologicamente afetadas utilizando os componentes wavelets em alta frequência e o filtro LP inverso para a identificação de características específicas destes sinais que os

diferenciam dos sinais de voz normal;

- Utilizar a medida de perturbação do período, ou *jitter*, dos componentes wavelets próximos à faixa de frequência fundamental dos sinais para a classificação entre sinais de voz de pacientes com edema de Reinke e nódulo nas pregas vocais;
- Alcançar uma avaliação automática confiável das condições da laringe de um paciente como um auxílio a profissionais da saúde.

Capítulo 2

REVISÃO BIBLIOGRÁFICA

2.1. VOZES PATOLÓGICAS

Acusticamente, as patologias relacionadas ao trato vocal e suas várias partes (Figura 1) são diferenciadas em relação a padrões perceptuais como rouquidão, aspereza e soproiedade (BAKEN, 1996; BLESS, 1991; BOONE; MCFARLANE, 1994). A voz áspera se caracteriza por um elevado estiramento das pregas vocais causado pelos músculos que as aduzem. A soproiedade está presente quando ocorre escape de ar na região glotal, seja por mau funcionamento dos músculos adutores, seja por alguma alteração orgânica. A combinação da aspereza com a soproiedade produz a rouquidão.

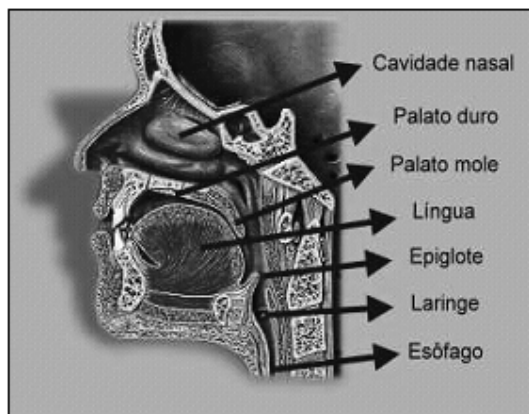


Figura 1 – Trato Vocal.

A avaliação acústica da voz em termos perceptuais pode levar ao bloco patológico que originou tal modificação acústica, embora possa ocorrer situação em que “se ouve” um determinado padrão acústico e nenhuma patologia é verificada em exames invasivos mais precisos como a videolaringoscopia e a videoestroboscopia. Isshiki (1969) alerta para o fato de que, devido à complexa estrutura de funcionamento do sistema produtor da fala, patologias com características soprosas podem ser confundidas com as perceptualmente definidas como roucas. As principais patologias de origem organofuncionais são apresentadas na Tabela 1 conforme a classificação de M. Behlau e P. Pontes (1995) (BEHLAU, M., 2001):

Tabela 1 – Algumas patologias de origem organofuncionais, localização anatômica e suas características vocais.

Patologias	Localização	Características Vocais
Edema de Reinke	Ao longo de toda a prega vocal.	Voz grave, fluida, rouca ou crepitante, no início agradável; virilizada.

continua ↓

Patologias	Localização	Características Vocais
Nódulo	Metade da porção vibratória; junção do 1/3 anterior e 1/3 médio da prega vocal.	Voz rouco-soprosa, grau discreto a extremo, ataques bruscos, ressonância baixa; alguns casos sem sinais vocais.
Pólipo	Borda livre da prega vocal.	Voz grave, rouca, rouco-soprosa, frequência dicrótica, incoordenação pfa, fadiga vocal.
Úlcera de contato	Porção cartilaginosa da prega vocal.	Voz grave-crepitante, disфония silente, odinofonia, odinofagia e otalgia.
Leucoplasia	Na mucosa, concentrada na borda livre.	Voz áspera, rouco-áspera, frequência aguda, fadiga vocal.

Um inadequado funcionamento de qualquer músculo intrínseco (principalmente) pode acarretar uma vocalização soprosa (ROSA; PEREIRA; GRELLET, 2000). Ao ocorrer uma superadução prolongada das pregas vocais, estas ficam passíveis de originarem uma fenda na região posterior, também causando escape de ar. Algumas interferências nas bordas de aproximação das pregas vocais podem resultar em disфония e ruído turbulento, ou seja, aparecimento de soprosidade na avaliação perceptual. Padrões e medidas acústicas capazes de diferenciar as diversas patologias vocais através métodos não invasivos têm sido razões para um grande número de pesquisas científicas. Além do desconforto, um problema encontrado no uso dos instrumentos de exame invasivo é uma ação de reflexo na cavidade supraglotal que pode levar a avaliações equivocadas sobre as condições da laringe do paciente. Desta maneira, além dos exames invasivos e da avaliação

perceptual, a utilização de um maior número de parâmetros de análise acústica pode auxiliar cada vez mais no diagnóstico mais preciso dos problemas vocais.

Atualmente, pesquisadores de diferentes áreas como fonoaudiólogos, engenheiros e matemáticos têm estudado diversas medidas de perturbação da voz. Diferentes tipos de parâmetros são utilizados para medidas de perturbação em torno da frequência fundamental (*Jitter*) (DAVIS, 1979; SCHOENTGE; GUCHTENEERE, 1995) e em torno da amplitude (*Shimmer*) (DELLER JR.; PROAKIS; HANSEN, 1993) do sinal vocal. Estas medidas são também denominadas como medidas da perturbação ciclo-a-ciclo. Como a voz humana é quase periódica, a duração de cada ciclo sofre pequenas variações no tempo. Dessa forma, define-se o *Jitter* como fenômeno de perturbação de período (ou frequência fundamental) presente em todos os sons vocálicos, tanto normais quanto patológicos. Alterações involuntárias ou provenientes de patologias produzem padrões errôneos de vibração nas pregas vocais, que são detectados por medidas baseadas na alteração ciclo-a-ciclo da amplitude ou da frequência fundamental. Alterações do *Shimmer*, por exemplo, são encontradas na presença de lesões de massa nas pregas vocais como nódulos, pólipos, edemas ou carcinomas, enquanto alterações do *Jitter* podem significar indício de doenças neurológicas ou dificuldades psicológicas.

Podem-se citar como exemplos de parâmetros de medida do *Jitter*, o fator de perturbação de frequência (PFF), o fator de perturbação direcional (DPF), a perturbação média relativa (RAP), o JR (*Jitter Ratio*), o JF (*Jitter Factor*) e o PPQ (*Period Perturbation Quotient*) (DAVIS, 1979). Já para a medida do *Shimmer*, são exemplos o quociente de perturbação de amplitude (APQ) e o índice de variabilidade de amplitude (AVI). A soproidade é facilmente detectada no espectro de

freqüências do sinal devido ao aumento das magnitudes em alta freqüência decorrentes do fluxo de ar turbulento na glote. Alguns parâmetros propostos na literatura são: LTAS (*Long Term Average Spectra*), HNR (*Harmonic to Noise Ratio*), RH (*Relative Harmonic*), BR (*Breathiness Ratio*), SNL (*Spectral Noise Level*), NNE (*Normalized Noise Energy*). Posteriormente, na seção 2.3, serão mais bem detalhados alguns destes parâmetros acústicos utilizados na análise de sinais de voz. O objetivo de se chegar a resultados mais eficientes na classificação de patologias do trato vocal, a partir de correlações entre os parâmetros acústicos, tem sido motivo para a realização de um número cada vez maior de pesquisas. É de enorme interesse científico a análise acústica de sinais, permitindo uma avaliação não invasiva e uma quantificação de parâmetros objetivos que venham a permitir o acompanhamento das condições do paciente no decorrer de uma terapia.

Duas patologias vocais, cujas características acústicas podem ser bastante similares em alguns casos, são mais tratadas neste trabalho: o edema de Reinke e o nódulo nas pregas vocais. Os nódulos, que são um dos problemas mais comuns na laringe, consistem em uma lesão benigna, geralmente imóvel durante a fonação (HUGH-MUNIER et al., 1997). São produzidos pelo uso contínuo e abusivo das pregas vocais, comuns em professores e cantores. São caracterizados por uma protuberância esbranquiçada na borda glótica das pregas vocais, geralmente bilateral, localizados na porção média da membrana vocal, o local de máxima força muscular e aerodinâmica durante a fonação. Eles provocam uma vibração assimétrica e o incompleto fechamento das pregas vocais, gerando soprosidade e resultando em variações no período de *pitch* e amplitude da voz, bem como adicionando ruído em alguns componentes espectrais. O estudo de parâmetros

acústicos que meçam estas variações pode, portanto, indicar a presença da patologia.

O edema de Reinke, também chamado de degeneração polipoidal, é descrito como um aumento de massa bilateral nas pregas vocais, causado por edema no corpo gelatinoso de Reinke, logo abaixo da mucosa (MURRY; ABITBOL; HERSAN, 1999). A patologia é usualmente associada a pacientes fumantes há longo período de tempo, bem como àqueles que fazem uso intensivo da voz ou apresentam algum fonotrauma (YATES; DEDO, 1984). É também considerada uma lesão benigna na laringe, podendo ser caracterizada por um aumento de fluido no espaço de Reinke, resultando em mudanças no movimento das pregas vocais. A vibração das pregas vocais apresenta assimetria, irregularidades e mudanças no *jitter* (variação no período de *pitch*). Com a elevação no tamanho das pregas vocais, a frequência fundamental é reduzida, tornando a voz mais grave e instável, além de ser também acompanhada de sopro. Esta conjunção de efeitos pode produzir uma percepção de rouquidão na voz do paciente.

2.2. CLASSIFICAÇÃO DE SINAIS DE VOZES PATOLÓGICAS

Uma série de técnicas vem sendo utilizada para a classificação de sinais de vozes patológicas com o uso de diferentes parâmetros e medidas acústicas, bem como, a partir da aplicação de inúmeras ferramentas para o reconhecimento de padrões e a classificação automática dos sinais. Fazendo-se uma análise comparativa dos algoritmos de detecção de patologias da laringe descritos na literatura, tem-se que a maior parte deles são baseados em wavelets, fractais ou redes neurais (HADDJITODOROV; BOYANOV; TESTON, 2000).

Nota-se que as redes neurais cobrem a maior parte das técnicas existentes, alcançando acertos próximos a 100% em alguns casos, mas com uma ordem de complexidade computacional relativamente elevada por aplicar uma grande quantidade de parâmetros como atributos de entrada do classificador. Usualmente, estes classificadores agrupam as vozes levando em consideração os seguintes parâmetros: frequências formantes, período de *pitch* e suas variações, estabilidade do período de *pitch* durante a fonação, grau de similaridade do formato do *pitch*, espectro médio de longo tempo (*long time average spectrum (LTAS)*), razão ruído-harmônicas (*noise-to-harmonics ratio (NHR)*) e razão harmônicas-ruído (*harmonics-to-noise ratio (HNR)*). Classificadores baseados em fractais podem alcançar cerca de 90% de acerto, mas, usualmente, detectam somente alguma patologia particular, como, por exemplo, a ataxia de Friedreich (ACCARDO; MUMOLO, 1998; PEAN et al., 2000).

Classificadores usando *wavelet packet* (algoritmo *best-basis*) (MALLAT, 1999) associados a redes neurais produzem cerca de 85% de acerto (SCHUCK JR.; GUIMARÃES; WISBECK, 2003). Um trabalho com a decomposição *wavelet packet* (MALLAT, 1999) de frases pronunciadas em inglês e um classificador utilizando LDB (*Local Discriminant Basis*) mostrou ser possível obter até 96% de acerto na classificação em dois grupos (sinais de fala normal e sinais de fala patológica) e até 74% na classificação em 4 grupos (sinais de fala masculina normal, feminina normal, masculina patológica e masculina normal) (UMAPATHY; KRISHNAN, 2005).

A DWT e o algoritmo LS-SVM estabeleceram um resultado adequado de generalização para distinguir entre as vozes normais e as patologicamente afetadas, principalmente aquelas provenientes de pessoas com edema de Reinke e nódulo

nas pregas vocais. No item 2.5 deste documento será ainda realizada uma análise mais detalhada das técnicas de classificação nas quais a Transformada Wavelet é utilizada. A Tabela a seguir mostra alguns resultados obtidos utilizando diferentes tipos de classificadores encontrados na literatura, bem como também com o uso de outros tipos de análise e bancos de dados diferentes dos utilizados neste trabalho, sendo também com diferentes casos de patologias e grupos de classificação.

Tabela 2 – Comparação entre os métodos propostos neste trabalho (primeira e antepenúltima linha) e outros com objetivos similares encontrados na literatura.

Métodos	Acertos (%)
DWT-db do sinal vocálico /a/, LPC e valores RMS (FONSECA et al., 2004)	80
Wavelet packet (<i>best basis</i>) + Redes Neurais (SCHUCK Jr., A.; GUIMARÃES, L. V.; WISBECK, J. O., 2003)	85
Fractais (ACCARDO; MUMOLO, 1998; PEAN et al., 2000)	90
DWT-db do sinal vocálico /a/, LPC, valores RMS e LS-SVM (FONSECA et al., 2007)	92
Wavelet packet de frases faladas em inglês + LDB (UMAPATHY; KRISHNAN, 2005)	96
Parâmetros acústicos variados + Redes Neurais (HADDJITODOROV; BOYANOV; TESTON, 2000)	próximo de 100

Deve-se enfatizar que os valores percentuais de acertos obtidos são extremamente relativos já que as características dos bancos de dados utilizados em cada trabalho são bastante diferentes. Embora alguns algoritmos utilizando as redes neurais possam alcançar valores de acerto próximos a 100% para o específico banco de dados destes trabalhos, isto ocorre, geralmente, com a utilização de um grande número de parâmetros acústicos, o que pode aumentar consideravelmente a complexidade computacional exigida.

2.3. MEDIDAS DE PERTURBAÇÃO

A seguir, serão brevemente tratadas as principais medidas de perturbação e parâmetros acústicos utilizados na análise de sinais de voz normal e patológica. Neste trabalho, é proposto um novo parâmetro de análise que apresenta várias relações com os já existentes, principalmente os relacionados à análise tempo-freqüência, quantificação de ruído e o *jitter*.

- Fator de Perturbação de Freqüência (PFF) - Representa a freqüência relativa das perturbações do período de *pitch* (ERÇELEBI, 2002) (P_i) maiores que 0,5 ms. Outros nomes relacionados a este parâmetro são: Fator de Perturbação de *Pitch* e Fator de Perturbação de Período. Estudos demonstram ser possível afirmar que vozes patológicas possuem um valor maior de PFF que vozes normais e que o PFF é sensível ao tamanho e locação da laringe. Para N amostras de um sinal, este parâmetro é dado por (LIEBERMAN, 1963):

$$PFF = \frac{\sum_{i=2}^N |P_i - P_{i-1}| > 0,5ms}{\sum_{i=1}^N P_i} \quad (1)$$

- Fator de Perturbação Direcional (DPF) – O DPF considera a direção das diferenças dos períodos, isto é, a porcentagem de alteração do sinal algébrico (aumento ou diminuição do período do pulso glotal) (HECKER; KRUEL, 1970). Este estudo utilizou seguimentos de vozes provenientes de frases, o qual demonstrou ser inadequado por Horii (HORII, 1979) que sugere a utilização de vogais sustentadas, verificando a mudança da direção do Jitter para 11 ciclos dos sinais. A contagem da mudança dos ciclos é realizada seguindo-se o seguinte procedimento: o primeiro período é considerado como período referência; se o segundo ciclo apresentar um

período menor, então l_{he} é atribuído um sinal negativo, caso contrário, um sinal positivo. A partir deste ponto, considera-se o segundo sinal como referência e se verifica o comprimento do posterior. O processo é repetido até o último ciclo. Através desse conjunto de sinais, é verificada a quantidade de mudanças ocorridas. Dessa maneira, o parâmetro é computado como se segue:

$$DPF = \frac{NMS}{n-1} \times 100\% . \quad (2)$$

em que :

NMS : Número de Mudanças algébricas do sinal;

n : número de ciclos contínuos.

- Perturbação Média Relativa (RAP) - Parâmetro proposto por Koike (KOIKE; CANLCATERRA, 1977) que, analogamente ao PFF, mede a flutuação dos períodos de *pitch*, porém de uma forma mais sofisticada. Sendo N o número de períodos analisados, tem-se, para uma janela de três amostras:

$$RAP = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| \frac{P(i-1) + P(i) + P(i+1)}{3} - P(i) \right|}{\frac{1}{N} \sum_{i=1}^N P(i)} . \quad (3)$$

- Jitter Médio (JM) - Este foi o primeiro parâmetro relacionado ao *jitter* (HORII, 1979),

$$JM = \frac{\sum_{i=1}^{N-1} |P_i - P_{i+1}|}{N-1} . \quad (4)$$

- Razão de Jitter (JR) - Fornece uma relação entre a média das perturbações do período a partir do período médio. O JR foi proposto para realçar a redução do jitter para freqüências fundamentais relativamente altas. Algebricamente esta relação é dada por:

$$JR = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |P_i - P_{i+1}|}{\frac{1}{N} \sum P_i} \cdot 10^3. \quad (5)$$

- Fator de Jitter (JF) - Este parâmetro é análogo ao JR, porém definido no domínio da freqüência como se segue:

$$JF = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |F_i - F_{i+1}|}{\frac{1}{N} \sum F_i} \cdot 10^2. \quad (6)$$

- Índice de variabilidade de período (PVI) e Índice de variabilidade de amplitude (AVI) - Basicamente estes parâmetros derivam do coeficiente de variação aplicado às N_1 amostras de um sinal (KOIKE et al., 1973):

$$CV = \frac{\frac{1}{N} \sum_{i=1}^{N_1} (X_i - \bar{X})^2}{\bar{X}^2}. \quad (7)$$

2.4. MEDIDAS ESPECTRAIS DE RUÍDO

- Espectro Médio de Tempo Longo (LTAS) - Basicamente constitui a densidade espectral de potência para um grande número de amostras.
- Relação de Potência em Altas Freqüências (HFPR) - Proposto for Shoji (1992),

este parâmetro em última análise é semelhante ao LTAS, porém calcula a média do logaritmo da densidade espectral de potência em intervalos de frequência bem definidos. É calculado por:

$$HPFR_i = \sum_{w=f_i}^{f_h} \overline{PSD_w} . \quad (8)$$

- Relação Harmônicas/Ruído (HNR)¹ - Esta relação é definida como (RODRIGUES; BEHLAU; PONTES, 1994):

$$HNR = \frac{N \sum_{t=1}^{T_a} S_a^2(t)}{\sum_{i=1}^N \sum_{t=1}^{T_i} [s_i(t) - s_a(t)]^2} , \quad (9)$$

onde $s_a(t)$ é a estimativa livre de ruído, $s_i(t)$ é o i -ésimo trecho do sinal, T_a é o comprimento da estimativa livre de ruído e T_i é o comprimento do i -ésimo trecho do sinal.

- Intensidade Relativa de Harmônicas - Mede a energia contida na segunda harmônica com relação à energia total do sinal. Este parâmetro estabelece um limite para pessoa com voz normal de 67,5%.
- Harmônicas Relativas - É a relação entre as componentes harmônicas e as não harmônicas calculadas em um segmento de três períodos:

¹ Na literatura especializada, é também utilizado o parâmetro inverso que mede a Relação Ruído/Harmônicas (NHR).

$$R = 10 \log \left[\frac{\sum_{i=1}^N S_i}{\sum_{i=1}^N N_i} \right], \quad (10)$$

em que S_i são as componentes harmônicas e N_i são as componentes não harmônicas.

- Corretores de Erro - O modo mais comum para realizar a correção de erros é aproximar os valores obtidos ao valor da mediana dos períodos. Mas, no caso de vozes patológicas, esse tipo de corretor altera as características de *jitter* do sinal original, invalidando as medidas de período realizadas. Sendo assim, até o momento na pesquisa bibliográfica realizada, não foi encontrado nenhum corretor de erro específico para vozes patológicas.

- Nível Espectral de Ruído (SNL) - Este parâmetro analisa as intensidades espectrais em intervalos de frequência, sendo útil na quantificação do som.

- Energia de Ruído Normalizada (NNE) - Considera o sinal composto de componentes periódicas mais um ruído aditivo. Para os valores amostrais do espectro do sinal analisado das frequências w de N_L até N_H , este parâmetro pode ser definido como:

$$NNE = 10 \log \left[\frac{\frac{1}{L} \sum_{w=N_L}^{N_H} \sum_{m=1}^L |\hat{N}_m(w)|^2}{\frac{1}{L} \sum_{w=N_L}^{N_H} \sum_{m=1}^L |S_m(w)|^2} \right] \quad (11)$$

2.5. TÉCNICAS QUE UTILIZAM A TRANSFORMADA WAVELET

Já é bem conhecido o fato das Transformadas Wavelet serem capazes de detectar temporal e espectralmente eventos transientes do sinal. Um dos trabalhos de pesquisa mais realizados em processamento de sinais de voz é a detecção e estimação do período de *pitch* do sinal que está diretamente relacionado ao fechamento do pulso glotal e que determina a frequência fundamental (ERÇELEBI, 2002). Para se determinar o período de *pitch* de um sinal de voz, inúmeras dificuldades são encontradas por causa dos efeitos da presença do ruído e também da própria perturbação de período existente no sinal de voz devido ao chamado *jitter*. Desta forma, pesquisadores procuram por algoritmos e ferramentas matemáticas eficientes que levem à robustez na detecção do *pitch*.

A informação do período de *pitch* é necessária para várias aplicações em processamento digital de sinais de fala, tais como, identificação e verificação de voz, análise e síntese de sinais de fala, aquisição e reconhecimento fonético, diagnóstico de patologias na voz, entre outras. Segundo S. Kadambe e G. F. Boudreaux-Bartels (1992), a transformada wavelet discreta pode ser utilizada para a detecção robusta do período de *pitch* de sinais de voz. Alguns exemplos demonstram inclusive uma performance superior do detector de *pitch* baseado na transformada wavelet discreta do que aqueles baseados em métodos mais convencionais a esta aplicação. Com a saída da decomposição wavelet é possível detectar o fechamento do pulso glotal utilizando a *wavelet cubic spline dyadic* e a de haar (KADAMBE; BOUDREAUX-BARTELS, 1992). A DWT de Daubechies também já foi utilizada por pesquisadores da USP – São Carlos para quantificar o nível de sopro da voz interpretada como ruído branco que acompanha a modulação do sinal (CAVALLINI JR.;

PEREIRA, 2001).

Alguns estudos já foram realizados na busca pela diferenciação entre sinais patológicos e não-patológicos, tendo demonstrado resultados vantajosos da wavelet packet (algoritmo *best-basis*) (MALLAT, 1999) com a base denominada Symlet de ordem 5 (PARRAGA, 2002) que, em associação com as Redes Neurais, podem identificar as vozes patológicas com até 85% de acerto (SCHUCK; GUIMARÃES; WISBECK, 2003). Outro estudo proposto ainda mais recentemente por Karthikeyan Umaphy e Sridhar Krishnan (2005) utiliza sinais de fala contendo frases em inglês pronunciadas por pacientes com diferentes tipos de patologias de origens orgânicas, neurológicas, traumáticas ou psicogênicas. Treinando e testando um classificador utilizando o algoritmo LDB (*“Local Discriminant Bases”*) com decomposições wavelet packet (algoritmo *best-basis*) foi possível obter deste estudo até 96% de acerto na classificação em dois grupos, os sinais de fala normal e sinais de fala patológica, e até 74% para a classificação em 4 grupos, sinais de fala masculina normal, feminina normal, masculina patológica e masculina normal (UMAPATHY; KRISHNAN, 2005). No entanto, ainda existe uma ordem de complexidade computacional relativamente elevada na implementação destes algoritmos *best-basis* com o LDB e naqueles utilizando ainda redes neurais. Portanto, pode ser de grande valia uma técnica computacionalmente mais simples em aplicações como, por exemplo, equipamentos eletrônicos móveis para a aquisição dos sinais de voz e para a avaliação automática dos pacientes no decorrer de uma terapia, dando aos médicos a possibilidade de receber periodicamente os sinais de voz destes pacientes pelo sistema eletrônico, de modo a acompanhar a evolução do tratamento.

Capítulo 3

FUNDAMENTOS TEÓRICOS

3.1. PREDIÇÃO LINEAR

A análise de sinais de fala baseada em termos dos coeficientes de predição linear (LPCs) (GRAY; MARKEL, 1976) tem atingido resultados satisfatórios há mais de 30 anos de pesquisa (DENG; SHAUGHNESSY, 2003, p. 41-42). Estudos demonstram que um sinal de voz, $s(t)$, pode ser visto como a saída de um filtro digital tudo-pólo IIR (Resposta ao Impulso Infinita) cuja função de transferência é $1/A(z)$, excitado por uma seqüência de impulsos que corresponde ao erro LPC do sinal $err(n)$, com transformada $E(z)$. Ou seja, a transformada z do sinal de voz $s(t)$, é dada por:

$$S(z) = \frac{E(z)}{A(z)} \quad (12)$$

O filtro $A(z)$ com coeficientes até ordem p é dado por:

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k}, \quad (13)$$

e é chamado de filtro digital inverso (ANANTHAPADMANABHA; YEGNANARAYANA, 1979).

De acordo com esse modelo, a n ésima amostra do sinal de voz pode ser aproximada por uma combinação linear das p amostras anteriores. A diferença do valor real e do valor aproximado corresponde ao erro de predição linear do sinal ou LP Residual. A energia do erro de predição do sinal é minimizada para determinar os pesos chamados coeficientes LP (LPCs). Dessa forma, para o sinal de voz $\{s(nT)\}$, o valor predito pela n ésima amostra é dado por:

$$\hat{s}(nT) = -\sum_{k=1}^p a_k s(nT - kT), \quad (14)$$

onde a_k são os LPC's. Estes coeficientes a_k do filtro são calculados de maneira a minimizar a soma dos quadrados dos erros:

$$err(n) = s(nT) - \hat{s}(nT). \quad (15)$$

Idealmente, a saída do filtro inverso digital de sinais de voz deveria constituir impulsos separados pelos respectivos períodos de *pitch*. Contudo, notam-se desvios do LP residual da seqüência de impulsos esperada. A análise LP assume este modelo tudo-pólo para representar o efeito combinado da resposta ao impulso do sistema do trato vocal e do formato do pulso glotal. O modelo tudo-pólo assume implicitamente uma característica de fase mínima para o sinal de voz (ANANTHAPADMANABHA; YEGNANARAYANA, 1979). Se isto não for válido, a resposta de fase do sistema do trato vocal não será compensada exatamente pelo filtro digital inverso. A compensação de fase será também afetada quando os formantes e suas larguras de banda não forem exatamente estimados, o que ocorre na prática. Além disso, o filtro não compensa os zeros que podem ser introduzidos devido à duração finita do pulso glotal.

3.2. A TRANSFORMADA WAVELET

Nesta seção são mostrados os fundamentos teóricos sobre a análise proposta pela DWT, as famílias das diversas funções utilizadas no algoritmo de decomposição desta transformada, suas características e propriedades matemáticas, bem como a relação destas com suas respectivas aplicações no processamento de sinais (MALLAT, 1989).

3.2.1. RELAÇÕES MATEMÁTICAS DA DECOMPOSIÇÃO WAVELET

Visando detalhar melhor como funciona matematicamente o algoritmo de decomposição da Transformada Wavelet, serão sintetizadas aqui as fórmulas mais importantes utilizadas nesta análise. Os sinais $s(t)$ podem ser representados pela análise wavelet a partir de séries do tipo:

$$s(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_k^j \psi_k^j(t) \quad (16)$$

onde,

$$\psi_k^j(t) = \psi(2^j t - k) \quad (17)$$

e

$$d_k^j = 2^j \int_{-\infty}^{\infty} s(t) \psi(2^j t - k) dt \quad (18)$$

A função $\psi(t)$ corresponde à função base da decomposição denominada wavelet, que pode ser de inúmeros tipos como será discutido posteriormente, e o termo 2^j fornecerá as compressões e dilatações necessárias, enquanto k deslocará a função convenientemente para varrer todo o sinal. O termo $d_{j,k}$ corresponde aos coeficientes do filtro (STRANG; NGUYEN, 1996; MALLAT, 1989; MISITI et al., 1997) utilizado na decomposição wavelet do sinal. Claramente, como foram utilizados sinais discretos, a integral da Equação 18 será substituída no algoritmo pela somatória e seus intervalos de integração pelo número de amostras do sinal a ser

decomposto $s(n)$.

Existe uma função $\phi \in L^2(\mathbb{R})$ tal que a família de funções $\phi_{j,k}(u) = \phi(2^j u - k)$, com j e $k \in \mathbb{Z}$, é base ortogonal do subespaço V_j (KAISER, 1994). É possível verificar que o espaço V_j está ligado aos componentes de aproximação (baixas freqüências) gerados na decomposição wavelet e também que a projeção de um sinal decomposto pela DWT em um determinado subespaço vetorial denominado W_j corresponde aos componentes de detalhe (altas freqüências) da DWT. Em termos matemáticos, têm-se as seguintes relações entre estes subespaços vetoriais:

$$\{0\} \leftarrow \dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots \rightarrow L^2 ; \quad (19)$$

$$V_{j+1} = V_j \oplus W_j. \quad (20)$$

Em termos de *banco de filtros* (GOMES; VELHO; GOLDENSTEIN, 1997; MALLAT, 1999; STRANG; NGUYEN, 1996) e observando agora os componentes de aproximação (A_j) e detalhe (D_j) para um sinal discreto $\vec{S}(i)$, a decomposição dada ao sinal pela Transformada Wavelet pode ser vista da seguinte maneira:

$$\vec{S}(i) = \sum_{k=0}^{2^j-1} \vec{H}_{J,k}(i) \phi_{J,k}(i) + \sum_{l=1}^J \sum_{k=1}^{2^{j-l}-1} \vec{G}_{l,k}(i) \psi_{l,k}(i), \quad (21)$$

$$\vec{S} = \vec{A}_J + \sum_{i=1}^J \vec{D}_i, \quad (22)$$

$$\vec{D}_j(t) = \sum_k d_k^j \psi_{j,k}(t) \quad \text{e} \quad \vec{A}_J = \sum_{j>J} D_j, \quad (23)$$

$$\phi[n] = \sum_k h_n \phi[2n-k] \quad \text{e} \quad \psi[n] = \sum_k g_n \phi[2n-k], \quad (24)$$

$$H_{j,k}(n) = \langle S, \phi_{j,k}(n) \rangle \quad \text{e} \quad G_{l,k}(n) = \langle S, \psi_{l,k} \rangle, \quad (25)$$

sendo $\psi(t)$ a função wavelet (ou wavelet mãe), $\phi(t)$ a função de escala (ou wavelet pai) que serão melhor explanadas ainda neste capítulo, g_k são os coeficientes de um filtro passa-altas que dão origem aos componentes de detalhe \vec{D}_j (Equações 22 e 23) correspondentes à projeção de \vec{S} no subespaço vetorial W_j com a base contendo $n/2^j$ vetores, e h_k são os coeficientes de um filtro passa-baixas que dão origem aos componentes de aproximação \vec{A}_j (Equações 22 e 23) correspondentes à projeção de \vec{S} no subespaço vetorial V_j com a base contendo $n/2^j$ vetores (MALLAT, 1989; FONSECA et al., 2006); j, k e $J \in \mathbb{Z}$, sendo J um nível de referência que define dois tipos de detalhes, aqueles associados com índices $j \leq J$ que são os detalhes finos e os outros correspondentes a $j > J$ correspondentes aos detalhes mais grosseiros que, quando agrupados, dão origem aos componentes de aproximação (MISITI et al., 1997). As funções $\psi(t)$ e $\phi(t)$ correspondem à convolução discreta no tempo de \vec{S} com o filtro passa-baixas $h[]$ e também com o filtro passa-altas $g[]$ (Equação 24) (STRANG; NGUYEN, 1996).

Na Figura 2 é apresentado o Banco de Filtros da decomposição wavelet, também relatado nas referências (GOMES; VELHO; GOLDENSTEIN, 1997; KAISER, 1994; MALLAT, 1989; MALLAT, 1999; MISITI et al., 1997; STRANG; NGUYEN, 1996), onde S é o sinal a ser analisado, sendo que os filtros passa-baixas e passa-altas correspondem respectivamente aos desenhos da direita e da esquerda inseridos no segundo retângulo de cima para baixo desta figura. Quanto menor o nível j nesta mesma Figura 2, maior as frequências e faixas destas

analisadas, ou seja, uma menor precisão de análise frequencial. Esta faixa de freqüências diminui pela metade ao passar de um nível a outro (2^j), aumentando assim esta precisão. Já com relação ao tempo, ocorre o inverso, quanto menor o nível j , menor também a janela temporal utilizada, ou seja, uma maior precisão na análise temporal é estabelecida (STRANG; NGUYEN, 1996). Para se obter os componentes de análise de aproximação e detalhe da Figura 2 (cA_i e cD_i) aplica-se ainda um operador de “downsampling” (GOMES; VELHO; GOLDENSTEIN, 1997) logo após os filtros. No banco de filtros de reconstrução (ou síntese) utilizado para a obtenção dos componentes de aproximação e detalhe das Equações 22 e 23, o projeto é realizado com a utilização dos filtros de síntese passa-altas e passa-baixas e o operador de “upsampling” (MALLAT, 1989; STRANG; NGUYEN, 1996).

As wavelets mãe (funções base na decomposição) e pai (funções de escala) presentes na literatura, $\psi(t)$ e $\phi(t)$ respectivamente, aumentam cada vez mais o número de integrantes nesta família. De acordo com a família wavelet têm-se algumas propriedades como, por exemplo, a existência ou não da função de escala $\phi(t)$ e de expressões de definição explícita para $\psi(t)$, simetria, regularidade, número de momentos nulos, suporte compacto, ortogonalidade ou biortogonalidade da análise resultante, se é ou não uma aproximação baseada em filtros do tipo FIR (Resposta ao Impulso Finita), entre outras. Estas propriedades estão muitas vezes relacionadas à aplicação dada à transformada como o processamento digital dos sinais (ou imagens) específicas, características de suavidade destes sinais reconstruídos e suas relações com a função estimada em análise de regressão não linear (MALLAT, 1999), propósitos de compressão e compactação, velocidade de convergência destas funções quando o tempo ou a freqüência tende ao infinito, a

familiaridade com o uso de determinadas funções, entre outras.

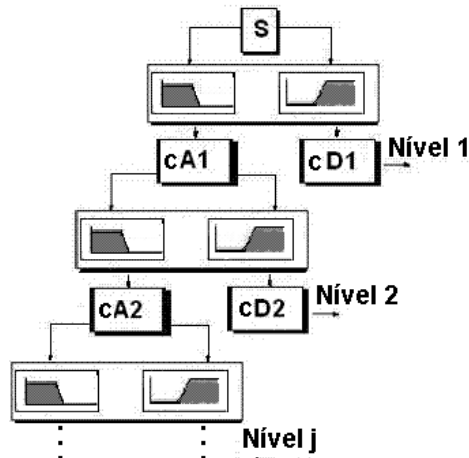


Figura 2 – Banco de Filtros de Análise da DWT.

3.2.2. A ANÁLISE WAVELET

Sabe-se que a Transformada de Fourier com Janela (*Short Time Fourier Transform* (STFT)) pode apresentar problemas na análise de um determinado sinal quando os detalhes deste sinal são muito menores que a largura da janela utilizada (falta de localização no domínio do tempo) e quando estes detalhes são maiores que a largura da janela (falta de localização no domínio da frequência). Estes problemas relacionados à largura da janela estão ligados ao conceito do *Princípio da Incerteza* (GOMES; VELHO; GOLDENSTEIN, 1997) que, em termos simples, afirma que é impossível obter precisões simultaneamente no domínio do tempo e no domínio da frequência. Pode-se concluir isto de maneira intuitiva já que, para se obter determinada precisão de localização no domínio do tempo é necessária a análise do sinal em curtos períodos, enquanto para se analisar precisamente as frequências deste sinal, deve-se observar o sinal por longos períodos de tempo.

Foi exatamente na tentativa de atenuação destes problemas que

pesquisadores de diversas áreas criaram a Transformada Wavelet, que apresenta estudos teóricos desde 1909 com a criação do algoritmo do sistema de Haar. Diversos outros trabalhos foram feitos na década de 30 e pesquisas mais aprofundadas com o desenvolvimento da teoria da *decomposição atômica* (CHUI, 1992) foram realizadas de 1960 a 1980. No final da década de 70, um engenheiro geofísico, J. Morlet, cria um método alternativo a Fourier com Janela para *análise de multiresolução* (MALLAT, 1999) de um sinal e em 1985 se dá o advento no estudo em processamento de sinais com Stephane Mallat e aplicações específicas das wavelets neste assunto com a pesquisadora Ingrid Daubechies.

Esta ferramenta matemática decompõe sinais em termos de diferentes tipos de funções base com a utilização de uma largura de janela variável de acordo com uma determinada escala pré-especificada em vários níveis, correspondendo a compressões e dilatações destas funções. Em outras palavras, pode-se dizer que a Transformada Wavelet é capaz de solucionar os problemas da Transformada de Fourier na obtenção de precisões requeridas tanto no domínio do tempo quanto no domínio da frequência (ver Figura 3 a seguir), constituindo-se uma ferramenta matemática utilizada para decompor, analisar e sintetizar funções com ênfase na localização tempo-frequência (DAUBECHIES; GILBERT, 1997; KAISER, 1994; GOMES; VELHO; GOLDENSTEIN, 1997; MALLAT, 1999), o que vem a ser de enorme importância em processamento digital de diversos tipos de sinais, inclusive os de voz normal e também patológica.

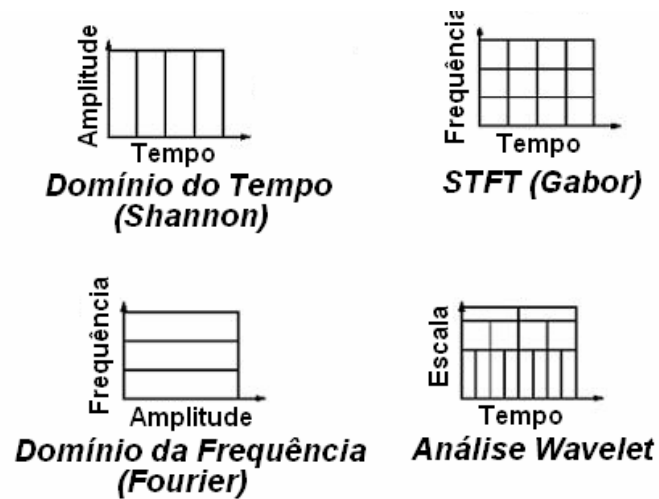


Figura 3 – Diferentes tipos de análise de sinais.

3.2.3. AS FAMÍLIAS WAVELET

Existem diferentes tipos de bases ortonormais e não ortogonais (Haar, Daubechies (db_s), Symlet (sym_s), Biortogonais (bior_s), Coiflet, Mexican Hat, B-splines, entre várias outras) pertencentes ao espaço quadrático das funções lebesgue-integráveis em \mathbb{R} ($L^2(\mathbb{R})$) utilizadas na construção das funções wavelet (KAISER, 1994) (Figura 4). A obtenção de melhores resultados em determinadas aplicações se tornou fundamental para a escolha destas bases. Para aplicações em computação gráfica como em compressão e edição de imagens, por exemplo, a base de Haar torna-se a mais adequada devida sua descontinuidade (STOLNITZ; DEROSE; SALESIN, 1996). Já em processamento digital de sinais, sabe-se que as wavelets de Daubechies possuem características especiais que as tornam mais utilizadas, trazendo resultados de grande importância científica (MALLAT, 1999).

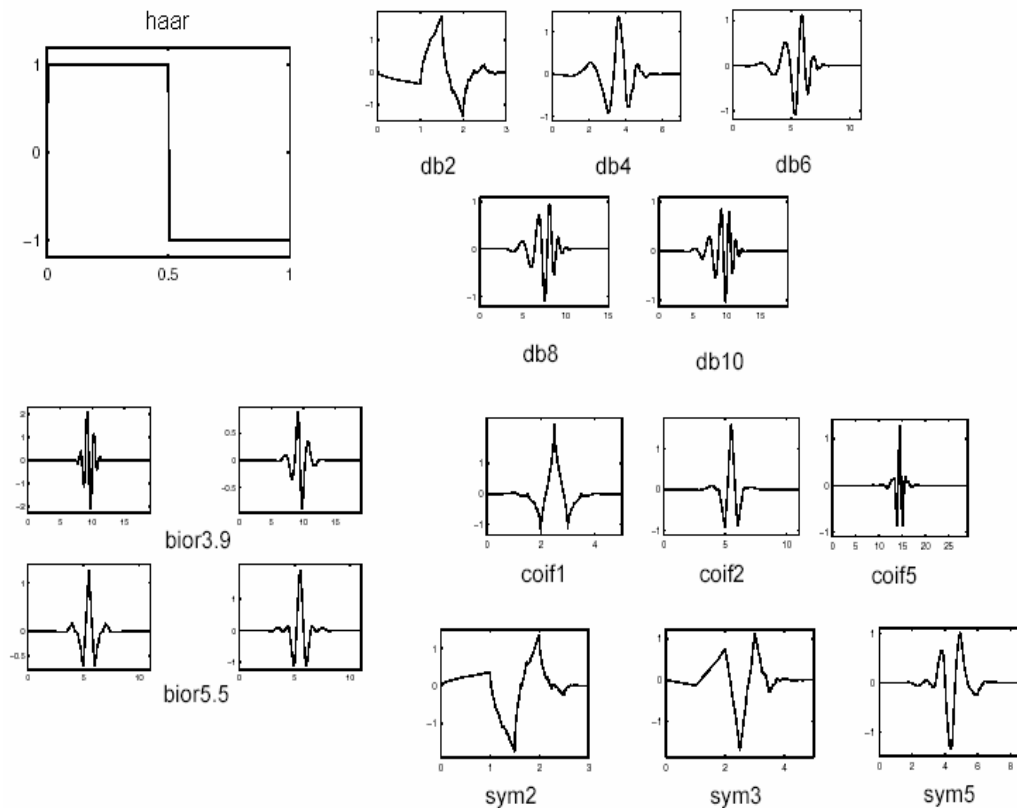


Figura 4 – Algumas funções wavelets.

A escolha da função wavelet adequada está, portanto, intrinsecamente relacionada a que são aplicados os resultados da decomposição nas várias escalas de frequência de um sinal qualquer. Os resultados obtidos desta decomposição, por sua vez, se relacionam às similaridades das características do sinal decomposto pela transformada em comparação à função wavelet base utilizada.

A Transformada Wavelet possui aplicações em vários campos de atuação da Engenharia Elétrica, como, por exemplo, na área de Bioengenharia em imagens médicas, processamento de sinais de voz e no projeto genoma, bem como na área de Telecomunicações e Análise de Sinais. As Wavelets, portanto, podem ser aplicadas em todas as áreas que utilizam o tratamento de imagens e sinais e sempre

que houver a necessidade de se manipular uma quantidade muito grande de dados. As wavelets já vêm sendo utilizadas com sucesso em inúmeros algoritmos relacionados a processamento de sinais como em filtros para ruído de grande eficiência (“de-noise”), compressão de sinais de som, estimação e detecção do período de *pitch* em sinais de voz, análise de qualidade da energia elétrica e de sinais de transmissão e recepção em telecomunicações, entre outros.

Fazem parte da família das wavelets ortogonais e com suporte compacto (GOMES; VELHO; GOLDENSTEIN, 1997; KAISER, 1994) as de Haar, as de Daubechies (dbN), as Symlets (symN) e as Coiflets (coifN) (MISITI et al., 1997). Uma função de suporte compacto é uma função com valor diferente de zero em um intervalo finito e igual a zero fora desse intervalo. Para estas wavelets, tanto $\phi(t)$ como $\psi(t)$ têm suporte compacto (DAUBECHIES; GILBERT, 1997), $\phi(t)$ existe e a análise é ortogonal, $\psi(t)$ tem um número determinado de momentos nulos (MALLAT, 1999), é possível usar para estas funções a análise da Transformada Wavelet Contínua (CWT) (DAUBECHIES, 1992) e também o algoritmo da Transformada Rápida de Fourier (FWT) (MISITI et al., 1997) para a DWT, bem como a análise utilizando filtros do tipo FIR. Uma desvantagem desta família está justamente na característica de pobre regularidade destas funções (MALLAT, 1999). Tem-se também que as wavelets de Daubechies são as únicas totalmente assimétricas para todas as dbNs. O número N corresponde à ordem da função wavelet sendo estritamente inteiro positivo, e está relacionado à quantidade de momentos nulos de $\psi(t)$, sendo que alguns autores consideram o número de momentos nulos igual a N e outros igual a 2N (MISITI et al., 1997). A família das wavelets de I. C. Daubechies são as mais comumente utilizadas e estão bem descritas nas referências de

Daubechies (1992) e Daubechies e Gilbert (1997). Estas wavelets possuem os coeficientes de seus filtros passa-baixas h_k calculados com a solução do seguinte sistema de equações:

$$\begin{cases} \sum_{k=0}^{N-1} (-1)^k h_k k^b = 0 \\ \sum_{k=0}^{N-1} h_k = \sqrt{2} \\ \sum_{k=0}^{N-1} h_k h_{k+2l} = 2\delta_{0,l} \end{cases} \quad (30)$$

onde N é o suporte do filtro, $b = 0, 1, \dots, (N/2)-1$, $\delta_{0,l}$ é o delta de Kronecker e $l \in \mathbb{Z}$. A primeira equação é responsável por produzir $N/2$ momentos nulos, a segunda vem do fato de que a área sob a função escala é unitária: $\int \phi(x) dx = 1$ e a terceira assegura a ortogonalidade entre os pares dos vetores da função escala.

A wavelet de Haar, bastante utilizada na compressão, compactação, edição e algoritmos de busca de imagens, é a db1 desta família possuindo apenas um momento nulo para $\psi(t)$ e considerada a mais simples das wavelets valendo +1 de zero a 1/2, -1 de 1/2 a 1 e zero caso contrário. Já as symNs apresentam alguma simetria (PARRAGA, 2002), enquanto a família das coifNs, além disso também possuem momentos nulos tanto para $\phi(t)$ quanto para $\psi(t)$. Mais detalhes sobre a construção das Symlets e das Coiflets podem ser encontrados em Daubechies (1992, p. 194, 254-259).

São exemplos da família das wavelets com as propriedades mínimas de existência a função gaussiana, a morlet e a chapéu mexicano (“*mexican hat*”, chamada assim devido o formato da função $\psi(t)$) (MISITI et al., 1997). São propriedades destas transformadas a não existência de $\phi(t)$, a análise não é

ortogonal, $\psi(t)$ não possui suporte compacto, a propriedade da reconstrução perfeita não está assegurada (MALLAT, 1999), possibilita a decomposição utilizando a CWT, as funções possuem simetria e $\psi(t)$ apresenta uma expressão matemática explícita. As principais dificuldades destas wavelets estão no fato do algoritmo FWT não poder ser aplicado e da reconstrução perfeita não estar disponível.

As B-splines apresentam ambas as funções $\phi(t)$ e $\psi(t)$ para a decomposição e para a reconstrução com suporte compacto e a análise é biortogonal (DAUBECHIES; GILBERT, 1997; MALLAT, 1999). Para estas funções $\phi(t)$ e $\psi(t)$ possuem momentos nulos na decomposição e reconstrução, além de apresentarem conhecida regularidade. A análise pode ser feita tanto pela CWT, como também pela DWT utilizando o algoritmo FWT. Elas também apresentam propriedades de simetria utilizando filtros FIR e suas desejáveis propriedades podem ser convenientemente alocadas na análise ou síntese (MISITI et al., 1997). Mais detalhes sobre a teoria e aplicações das wavelets biortogonais podem ser encontrados nas páginas 259, 262-285 da referência Daubechies (1992) e também em Cohen (1992).

Nos últimos anos, diversos estudos têm sido realizados buscando a construção de wavelets específicas à aplicação realizada pela decomposição ou síntese de determinados sinais. Isto ocorre por exemplo nos estudos das wavelets de Malvar, bastante utilizadas em sinais de fala e voz devido às características especiais dadas a esta função (RIS; FONTAINE; LEICH, 1995), também ocorrendo em estudos como na detecção em tempo real de *spikes* (GUIDO et al., 2006) de diversos tipos de sinais e inclusive para a identificação de sinais de voz patológica utilizando uma função wavelet não-ortogonal com fase praticamente linear, estável e causal com filtro do tipo FIR (*Finite Impulse Response*). Dessa maneira, é possível

afirmar que para a escolha da família wavelet mais adequada, a melhor alternativa seja atender às necessidades da aplicação que será dada aos resultados dos bancos de filtros de análise e/ou síntese.

3.3. SUPPORT VECTOR MACHINES (SVM)

As *Support Vector Machines* (SVM) constituem um algoritmo universal de aprendizado proposto por Vapnik (1995) sendo aplicado tanto para regressão quanto para reconhecimento de padrões (CORTES; VAPNIK, 1995). Os resultados da aplicação desta técnica são comparáveis aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais e, em algumas tarefas, têm se mostrado superiores, tal como na detecção de faces em imagens, na categorização de textos e em aplicações em bioinformática. A rotina de treinamento do SVM consiste em resolver um problema de otimização quadrática forçada. O SVM é aplicado para problemas de ordem linear e não-linear e usa uma função denominada *kernel* para mapear os dados de entrada em um espaço de característica ultradimensional no qual o problema se torna linearmente separável para o classificador (BURGES, 1999). A boa capacidade de generalização, robustez em espaços de grande dimensão, convexidade da função objetivo e uma teoria de aprendizado bem definida matemática e estatisticamente são importantes vantagens do classificador SVM (CRISTIANINI; SHAW-TAYLOR, 2000).

3.3.1. SVM PARA RECONHECIMENTO DE PADRÕES

Em problemas de reconhecimento de padrões, a função de classificação é um hiperplano, $f(x) = wx + b$, capaz de separar linearmente as classes, onde os parâmetros w e b podem ser otimizados no treinamento SVM de maneira a

maximizar a separação existente entre as classes. O treinamento do classificador SVM consiste na solução de um problema de otimização quadrática. Seja (x_i, y_i) , $x_i \in R_i$ e $y_i \in [-1, +1]$, $i = 1, \dots, n$ um conjunto de dados de treinamento, onde x_i é o i -ésimo elemento do vetor de entrada com uma específica saída binária 1 ou -1, então o classificador original satisfaz as seguintes condições (ZHANG; ZHOU; JIAO, 2004):

$$w^T \tau(x_i) + b \geq 1 \quad \text{se } y_i = 1, \quad (26)$$

$$w^T \tau(x_i) + b \geq -1 \quad \text{se } y_i = -1. \quad (27)$$

ou equivalentemente:

$$y_i [w^T \tau(x_i) + b] - 1 \geq 0 \quad \text{para } i = 1, 2, \dots, N, \quad (28)$$

onde $\tau : R^n \rightarrow R^m$ é a função que mapeia o espaço de entrada, geralmente um espaço de característica ultradimensional no qual os pontos dos dados de entrada se tornam linearmente separáveis por um hiperplano definido pelo par $(w \in R^m$ e $b \in R)$ (ZHENG; LIU; TIAN, 2004). Então a função de classificação é:

$$f(x) = \text{sign}[y_i w^T \tau(x_i) + b]. \quad (29)$$

De maneira a permitir uma violação da Equação 28, são introduzidas as variáveis ξ_i , tal que (VAPNIK, 1995; ZHENG; LIU; TIAN, 2004),

$$y_i [w^T \tau(x_i) + b] - 1 \geq 1 - \xi_i; i = 1, \dots, N \text{ e } \xi_i > 0. \quad (30)$$

Para encontrar um bom classificador linear, estas variáveis ξ_i devem ser minimizadas para obter menores erros no treinamento e, para uma melhor generalização, a margem entre as classes deve ser então maximizada. Estas

variáveis podem ser então combinadas no seguinte problema de minimização (ZHANG; ZHOU; JIAO, 2004; ZHENG; LIU; TIAN, 2004):

$$\min_{w,b,\xi} J(w,b,\xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i, \quad (31)$$

com

$$y_i [w^T \tau(x_i) + b] - 1 \geq 1 - \xi_i;$$

$$\xi_i \geq 0; i = 1, 2, \dots, N, \text{ e } C > 0,$$

onde C é um parâmetro constante usado para controlar o *trade-off* entre o erro de treinamento e a margem. A figura a seguir mostra um classificador linear com a demarcação da margem:

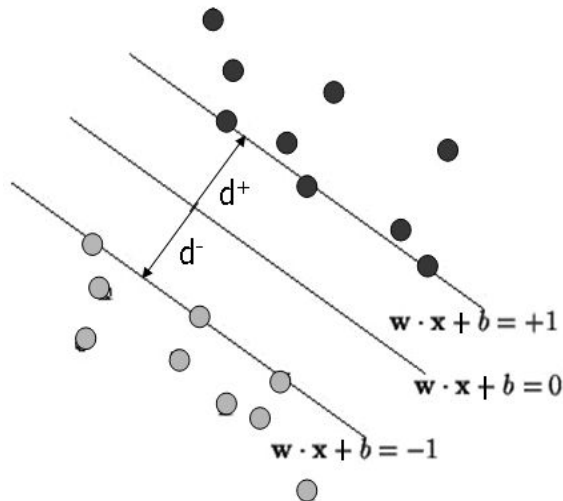


Figura 5 – Margem(d):- mínima distância possível entre qualquer ponto do treinamento e o hiperplano de classificação.

Utilizando técnicas que fazem uso dos multiplicadores de Lagrange, a minimização da Equação 31 nos leva ao seguinte problema:

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (32)$$

sujeito a:

$$\sum_{i=1}^n y_i \alpha_i = 0; \quad 0 \leq \alpha_s \leq C; \quad e \quad s = 1, \dots, n.$$

A função $K(x_i, x_j)$ é a chamada função *kernel* do classificador SVM (SV kernel) se satisfizer algumas das condições estabelecidas pelo Teorema de Mercer (BERTSEKAS, 1995). Uma função SV kernel é dada por:

$$K(x_i, x_j) = \tau(x_i)^T \tau(x_j) \quad (33)$$

Mais freqüentemente, em problemas nos quais o espaço de entrada é não-linear, a função $\tau(x)$ é muito complexa e as funções kernel se tornam excelentes peças deste algoritmo. A linear, a polinomial, a base radial e a função sigmoideal são os kernels mais comumente utilizados.

3.3.2. LS-SVM PARA RECONHECIMENTO DE PADRÕES

No algoritmo LS-SVM (Least Squares Support Vector Machines) para reconhecimento de padrões (SUYKENS; VANDEWALLE, 1999; SUYKENS et al., 2002), o padrão do classificador SVM de Vapnik foi modificado para a seguinte formulação:

$$\min_{w, b, \xi} J(w, b, \xi) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^n e_k^2 \quad (34)$$

sujeita às seguintes condições:

$$y_i [w^T \tau(x_i) + b] = 1 - e_k \quad e \quad k = 1, 2, \dots, n.$$

Pode-se notar que a passagem da Equação 31 para 34 envolve a substituição das constantes da inequação por constantes de uma equação com o termo de erro quadrático (portanto, mínimos quadrados (*least squares*)) similar ao método de regressão (*ridge regression*) (ZHENG; LIU; TIAN, 2004). A correspondente de Lagrange para a Equação 34 é a seguinte:

$$L(w, b, e, \alpha) = J(w, b, e) - \sum_{k=1}^n \alpha_k \{y_k [w^T \tau(x_k) + b] - 1 + e_k\}, \quad (35)$$

α_k sendo o multiplicador de Lagrange mostrado na referência (CRISTIANINI; SHAW-TAYLOR, 2000). A condição ótima nos leva ao seguinte $[(N + 1) \times (N + 1)]$ sistema linear:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad (36)$$

onde

$$\begin{aligned} Z &= [\tau(x_1)^T y_1, \dots, \tau(x_n)^T y_n]; \\ Y &= [y_1, \dots, y_n] \quad e \quad \alpha = [\alpha_1, \dots, \alpha_n]. \end{aligned} \quad (37)$$

O algoritmo LS-SVM utiliza os mínimos quadrados, portanto, apenas para modificar a formulação do SVM proposta inicialmente por Vapnik (1995) de modo a simplificar sua implementação. A aplicação proposta para reconhecimento de padrões é mantida da mesma maneira.

Capítulo 4

MATERIAIS E MÉTODOS

Este trabalho foi realizado utilizando os equipamentos do Laboratório de Processamento de Sinais Biológicos (LPSB) e de Instrumentação Microeletrônica (LIM) da Escola de Engenharia de São Carlos da Universidade de São Paulo (EESC – USP). O LPSB possui uma série de diferentes bancos de dados e trabalhos na área de modelagem biomecânica da laringe e, a partir de 1989, houve um forte impulso na área da análise acústica de voz com o intuito de detectar disfunções da laringe. Também foi utilizado o *Speech Processing and Auditory Perception Laboratory* (SPAPL) do Departamento de Engenharia Elétrica da Universidade da Califórnia, Los Angeles (UCLA), com o auxílio dos pesquisadores deste laboratório. Além disso, deve-se enfatizar o auxílio dos profissionais do Departamento de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP), de onde foi possível obter o banco de dados com as vozes patológicas utilizadas neste trabalho.

4.1. BANCO DE DADOS

Um banco de dados com sinais de voz contendo o fonema do português brasileiro /a/ foi extraído de 76 pessoas, 30 normais, sem nenhuma patologia na voz, 30 patologicamente afetados com nódulo nas pregas vocais e 16 afetados com edema de Reinke. Foram usados sinais provenientes de homens (43.33%) e mulheres (56,66%) com idades variando de 4 a 72 anos (90% destes de 18 a 54). Um /a/ sustentado em *pitch* e altura confortáveis é considerado o melhor som vocálico para extrair e medir parâmetros da voz, bem como para discriminar vozes provenientes de pacientes com diferentes tipos de patologias no trato vocal (ROSA; PEREIRA; GRELLET, 2000). Os sinais foram gravados em arquivos wave PCM (BOSI; GOLDBERG, 2003) com um software profissional amostrando as vozes em 22050Hz, 16 bits, mono-canal. Para a gravação foi utilizado um microfone de banda larga unidirecional (LeSon SM58), com controle dinâmico de ganho. Foi pedido para cada pessoa, tanto aquelas com voz normal quanto as com vozes patológicas, para sustentar este som vocálico por 5 segundos, em níveis confortáveis de magnitude e *pitch*, conforme a conversação natural de cada um/uma. Este sinal foi então janelado para que este dado fosse analisado com 2500 amostras, correspondendo de 20 a 40 períodos do sinal de voz, de acordo com sua frequência fundamental.

A voz de pessoas saudáveis foi usada como grupo de controle neste experimento. Todos os indivíduos foram diagnosticados por médicos do setor de otorrinolaringologia e cirurgia da cabeça e pescoço do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP), Brasil (<http://www.fmrp.usp.br>), usando videolaringoscopia e videoestroboscopia para a confirmação. O grupo de controle, não fumantes, também foram submetidos a

procedimentos clínicos para confirmação da ausência de qualquer origem orgânica ou funcional. Se o paciente não pudesse suportar o intervalo de fonação, então lhe era pedido que parasse a gravação e reiniciasse. Neste caso, muito cuidado foi tomado para prevenir um aumento inerente da frequência fundamental e uma estabilidade artificial da produção da fala, devido o aumento do fluxo de ar através da glote.

4.2. METODOLOGIA E ALGORITMO IMPLEMENTADO

Após uma série de testes, utilizando diferentes bancos de dados como será discutido na próxima seção, definiu-se o seguinte algoritmo cujos passos estão descritos a seguir. Cada um dos 76 sinais do banco de dados foi normalizado em termos da amplitude, de tal modo que o maior dos valores de amplitude do sinal, positivo ou negativo, tornou-se 1. A decomposição DWT, usando os filtros de análise, $h[]$ e $g[]$, e os operadores de downsampling (MALLAT, 1999) foi então aplicada. Somente os coeficientes de detalhe D_1 , D_2 , D_3 , D_4 , D_6 e os coeficientes de aproximação A_5 e A_6 foram usados nos próximos passos deste algoritmo. Os componentes D_1 , D_2 , D_3 e D_4 foram usados para a extração do ruído em alta frequência identificada em vozes patológicas (GRANQVIST; HAMMARBERG, 2003; MARTINEZ; RUFFINER, 2000) a partir do uso da análise temporal e frequencial da DWT (CAVALLINI; PEREIRA, 2001). Já os parâmetros A_5 , A_6 e D_6 foram extraídos por estarem mais próximos da faixa de frequência que contêm as frequências fundamentais dos sinais de voz para se realizar a medida do *jitter*. Levando-se em consideração a taxa de amostragem de 22050 Hz utilizada na gravação dos sinais e o processo de multiresolução da DWT, estão sendo analisados os componentes de frequência de 5512,5 a 11025 Hz em D_1 , de 2756,25 a 5512,5 em D_2 , de 1378,125 a

2756,25 em D_3 , de 689,06 a 1378,125 em D_4 , de 0 a 344,5 Hz em A_5 , de 0 a 172,25 Hz em A_6 e de 172,25 a 344,5 Hz em D_6 .

A Figura a seguir mostra o banco de filtros de análise da DWT utilizando os operadores de downsampling, bem como os filtros passa-altas ($g[]$) e passa-baixas ($h[]$) (MALLAT, 1999) para a obtenção dos coeficientes de aproximação e detalhe específicos que foram utilizados neste trabalho para a análise dos sinais de voz.

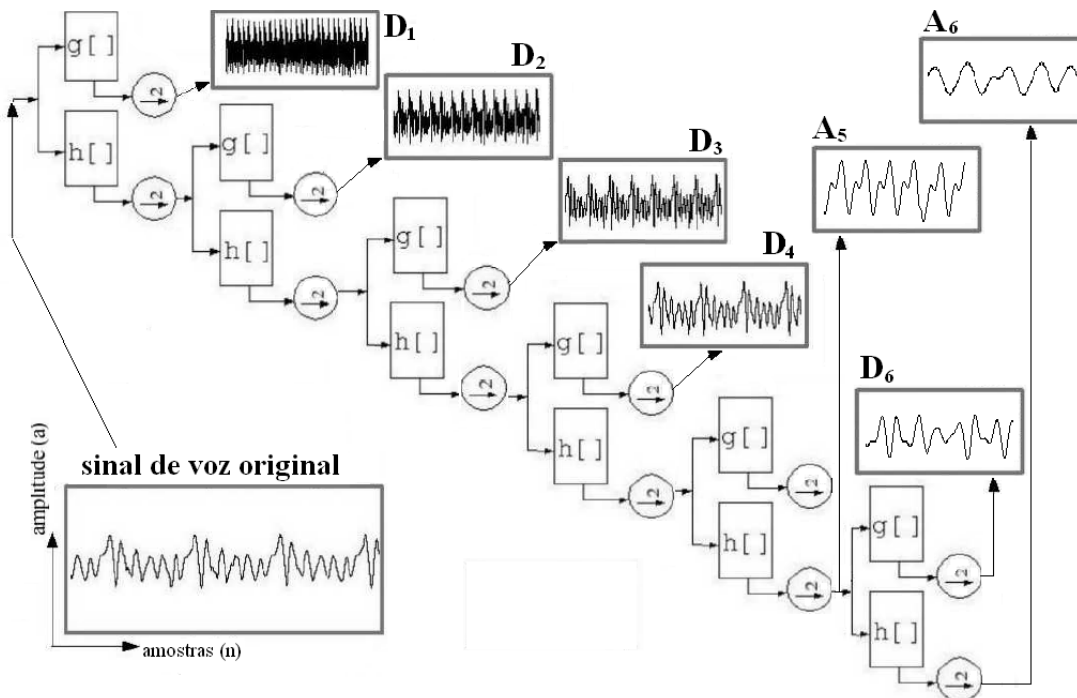


Figura 6 – Banco de Filtros da DWT evidenciando os componentes utilizados no algoritmo de classificação deste trabalho.

Depois disto, os filtros de síntese e os operadores de upsampling (STRANG; NGUYEN, 1996) foram aplicados para reconstruir o sinal original somente a partir do coeficiente D_1 , obtendo RD_1 . O mesmo processo foi repetido para obter RD_2 , RD_3 , RD_4 , RD_6 , RA_5 e RA_6 , respectivamente a partir dos coeficientes D_2 , D_3 , D_4 , D_6 , A_5 e A_6 . Somente os coeficientes reconstruídos em escala de altas frequências (RD_1 ,

RD₂, RD₃ e RD₄) foram normalizados em termos da amplitude. Foram utilizados os coeficientes de predição linear (LPCs) (GRAY; MARKEL, 1976), buscando enfatizar ainda mais a separabilidade entre as classes, conforme será mostrado na próxima seção. Deve-se lembrar que os LPCs têm atingido resultados satisfatórios há mais de 30 anos na análise de sinais de fala (DENG; SHAUGHNESSY, 2003, p. 41-42). O filtro LP inverso, ordem 100, foi aplicado aos sinais reconstruídos RD₁, RD₂, RD₃ e RD₄, de modo a aumentar ainda mais a diferença entre os sinais de voz normal e patológica, como evidenciado em (FONSECA et al., 2007a).

Os valores RMS dos componentes wavelet em alta frequência (RD₁, RD₂, RD₃ e RD₄) foram calculados, obtendo RMS_{RD1}, RMS_{RD2}, RMS_{RD3} e RMS_{RD4}. É importante lembrar que o valor RMS (*Root Mean Square*), ou valor eficaz, é correlato à energia de um sinal, sendo uma medida da magnitude de uma quantidade variável. Ele pode ser calculado para uma série de valores discretos ou para uma função variável contínua. O nome deriva do fato de ser a raiz quadrada da média aritmética dos quadrados dos valores. Para um sinal discreto de N valores {x₁, x₂, ..., x_N} ele é calculado da seguinte maneira:

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (38)$$

Para a validação dos resultados e a obtenção de um hiperplano de classificação, utilizou-se então o algoritmo de inteligência artificial de reconhecimento de padrões discutido anteriormente, o LS-SVM. Foram utilizadas todas as combinações possíveis dos valores RMS dos 4 diferentes componentes de

detalhe D_1 , D_2 , D_3 e D_4 da Transformada Wavelet Discreta de Daubechies (DWTdb), especificando-se informações de diferentes escalas de frequência do sinal para 3 diferentes funções kernel do algoritmo LS-SVM mais utilizadas, a função Linear, a *Radial Base Function (RBF)* e a *Multi-layer Perceptron (MLP)*. Desta maneira, foram obtidos os resultados de classificação entre os sinais de voz normal e sinais de voz patológica (nódulo e edema separadamente).

Para a classificação entre as duas diferentes patologias vocais analisadas (edema de Reinke e nódulo nas pregas vocais), utilizou-se o software comercial *Análise de Voz*, versão 6.0 (SCALASSARA et al., 2007), para se calcular o valor de *jitter* dos componentes wavelet na faixa das frequências fundamentais dos sinais de voz a partir do quinto nível de decomposição, RD_6 , RA_5 e RA_6 . Deve-se lembrar que M. Rosa, J. C. Pereira e M. Grellet (2000) mostram que a melhor medida acústica indicada para a discriminação de patologias na voz é o *Jitter* (perturbação em torno da frequência fundamental) (SCHOENTGE; GUCHTENEERE, 1995). Dessa maneira, foram procurados parâmetros que utilizam a medida do *jitter* dos coeficientes da DWTdb10 em escalas convenientes de frequência para a classificação entre as duas diferentes patologias vocais (nódulo e edema de Reinke). Os valores de *jitter* dos componentes wavelet na faixa da frequência fundamental dos sinais (RD_6 , RA_5 e RA_6) foram então utilizados como atributos de entrada no classificador LS-SVM para distinguir entre os grupos de sinais oriundos de pacientes com as 2 diferentes patologias vocais. Estes valores de *jitter* dos componentes wavelet RD_6 , RA_5 ou RA_6 (de acordo com a similaridade entre a frequência fundamental destes componentes com a frequência fundamental dos respectivos sinais de voz) ofereceram resultados significativos para classificação entre as duas

patologias. Os limiares definidos por estes classificadores e os acertos obtidos na classificação serão discutidos na próxima seção.

Um diagrama em blocos simplificado com cada um dos passos do algoritmo utilizado é mostrado na Figura 7:

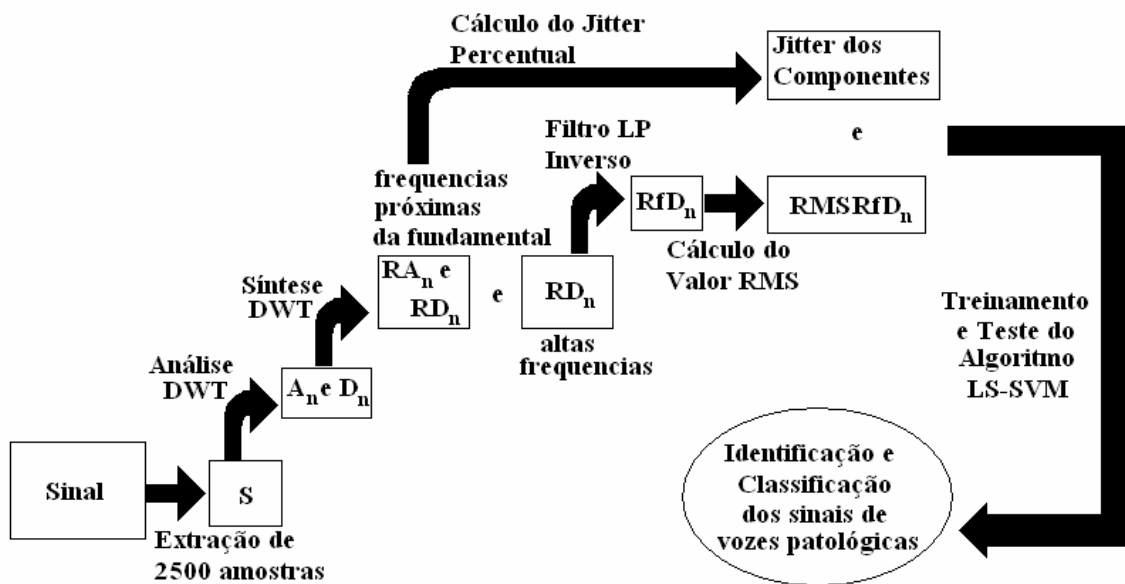


Figura 7 – Diagrama em Blocos simplificado com os passos do algoritmo utilizado para a identificação e classificação das vozes patológicas.

Capítulo 5

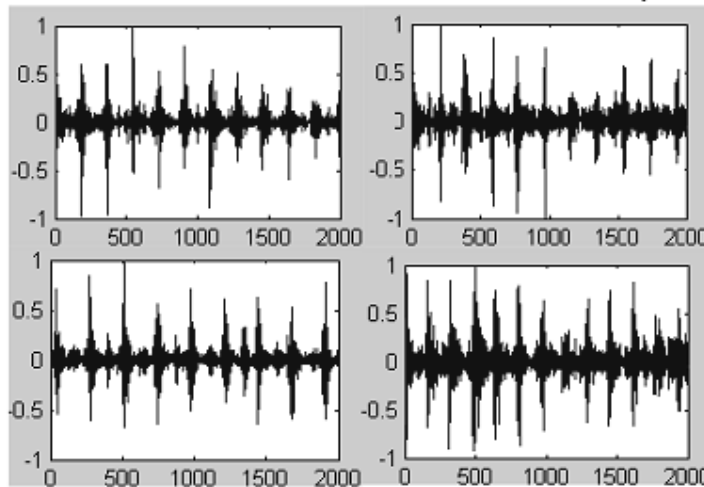
RESULTADOS E DISCUSSÕES

Neste capítulo são detalhados os resultados obtidos, bem como a discussão dos mesmos. Comparações com outros trabalhos na literatura também são realizadas, lembrando-se sempre que se deve levar em consideração diferenças entre os bancos de dados utilizados, assim como, diferentes maneiras de obtenção dos mesmos, os equipamentos utilizados e as condições de gravação.

A seguir são mostrados os componentes de detalhe (D_2) do segundo nível de decomposição da DWT de Daubechies (análise de ruído em altas frequências (2756,25 a 5512,5)) para trinta sinais de vozes normais e trinta sinais de vozes de pacientes com nódulos nas pregas vocais. Deve-se lembrar que estes sinais foram reconstruídos utilizando-se somente as informações desta escala de frequência associada. Foram também normalizados pelos seus respectivos valores de pico, obtendo como resultado sinais com padrões similares aos expostos na figura a seguir. Devido a soproidade existente nos sinais oriundos de pacientes com nódulo nas pregas vocais, os gráficos mostram a presença acentuada de ruído em alta frequência (correspondente à faixa frequencial deste componente) nos sinais de vozes patológicas que parecem impedir o surgimento aproximadamente periódico de picos de amplitude que ocorrem nos sinais oriundos de vozes normais (FONSECA et

al., 2004). Foi também analisado que estes picos ocorrem a uma distância equivalente ao período de *pitch* do sinal de voz, podendo tornar estes sinais úteis em estudos de aplicação das wavelets em detecção de *pitch* (KADAMBE; BOUDREAUX, 1992), assunto já abordado por vários pesquisadores da área.

**Componentes D2 da DWTdb10 normalizados
pelo valor de pico (vozes normais)**



**Componentes D2 da DWTdb10 normalizados
pelo valor de pico (vozes patológicas)**

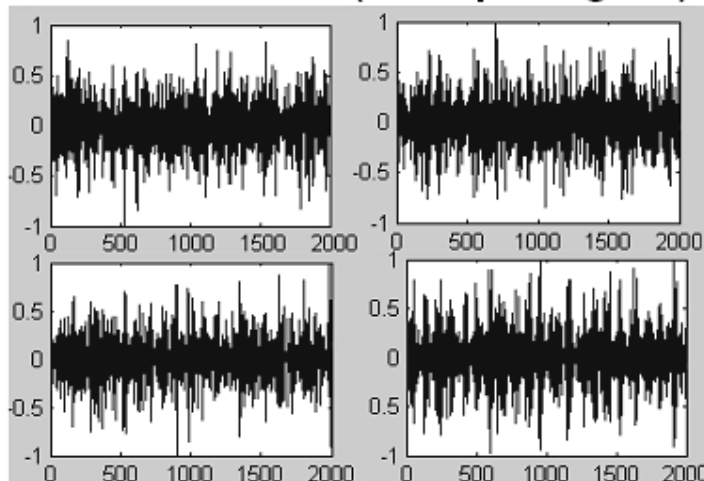


Figura 8 – Componentes de detalhe D_2 da DWTdb10 normalizados pelo valor de pico dos sinais de voz normal e patológica do banco de dados.

A Tabela a seguir sumariza os resultados das relações percentuais obtidas entre as médias aritméticas dos valores RMS_{RD2} das vozes normais com as das vozes patológicas (Diferença Relativa normal/patológica (DRNP)) com as diferentes famílias wavelet utilizadas.

Tabela 3 – Famílias Wavelet utilizadas (FW), tamanhos de suporte dos filtros correspondentes (TS) e as respectivas diferenças relativas percentuais entre a média aritmética dos valores RMS_{RD2} das vozes normais e das vozes patológicas (DRNP).

FW	TS	DRNP (%)
Haar	2	17.6
Daubechies	4	17.6
Daubechies	8	20.8
Daubechies	12	20.8
Daubechies	16	20.8
Daubechies	20	14.74
Daubechies	40	13.46
Coiflet	6	13.1
Coiflet	12	12
Coiflet	18	12
Coiflet	30	13.1
Symmlet	8	16
Symmlet	16	17.3
Symmlet	24	14.2

Os resultados da tabela anterior mostram que as wavelets de Daubechies apresentam melhores resultados para classificação, já que, para estas, obteve-se as maiores diferenças relativas entre os valores RMS das duas classes. As vozes normais apresentaram um valor mais baixo de RMS nesta escala corresponde à alta frequência em comparação aos valores RMS desta mesma escala para as vozes patológicas. Certamente, isto ocorre devido à maior presença de ruído em alta frequência provenientes deste tipo de patologia (GRANQVIST; HAMMARBERG,

2003; MARTINEZ; RUFFINER, 2000), detectado pela DWT-db8 (CAVALLINI; PEREIRA, 2001) e utilizado agora como importante informação para identificação de desordens na voz ou vozes provenientes de pacientes com patologias deste tipo, onde há um incompleto fechamento das pregas vocais. Os resultados são bastante encorajadores, dados que as amostras contêm homens e mulheres, bem como crianças e adultos. Uma interessante discussão ainda em aberto é que para diferentes famílias wavelet e para tamanhos de suporte acima de 8 desta mesma família, mesmo com as respostas em frequência dos filtros de Daubechies estando mais perto da resposta ideal, isto não significou uma melhoria na classificação, o que mostrou que o melhor filtro em termos de seletividade frequencial não corresponde ao melhor filtro para a classificação e vice-versa.

Ao aplicar diretamente os valores RMS dos componentes de detalhe reconstruídos RD_2 sem passar pelo filtro LP inverso e utilizando simplesmente um valor de limiar de classificação adequado, obtém-se um acerto de **71,67%** na classificação entre as vozes normais e patológicas. Aplicando o filtro LP inverso com diferentes ordens M , obtém-se então os seguintes resultados de classificação, conforme tabela a seguir:

Tabela 4 – Acertos na Classificação para cada ordem N do filtro LP inverso.

<i>Ordem do filtro LP inverso</i>	<i>Acerto na Classificação</i>
5	73,33
10	75
20	76,67
40	78,33
60	81,67
80	83,33
100	85
150	85
200	85
300	85

Dessa maneira, mostra-se um aumento significativo no acerto da classificação ao aplicar o filtro LP inverso chegando a **85%** a partir da ordem $M = 100$. Os gráficos a seguir mostram como os valores RMS dos componentes de detalhe se distribuem sem a aplicação do filtro LP inverso e após a aplicação do filtro LP inverso de ordem 100.

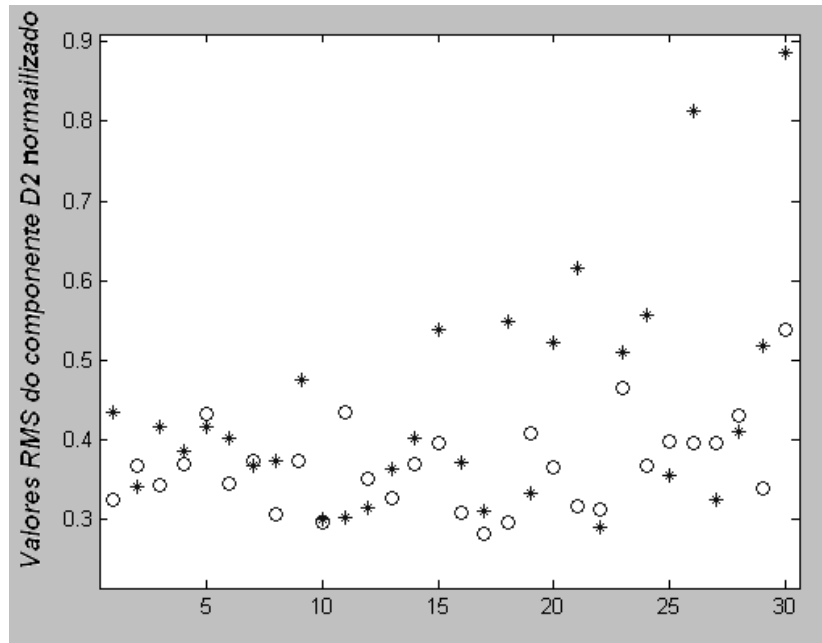


Figura 9 – Valores RMS do componente de detalhe D_2 da DWTdb10 normalizado pelo módulo de seu valor de pico para cada um dos 60 sinais de voz. Os pontos * e o correspondem às vozes patológicas e normais, respectivamente.

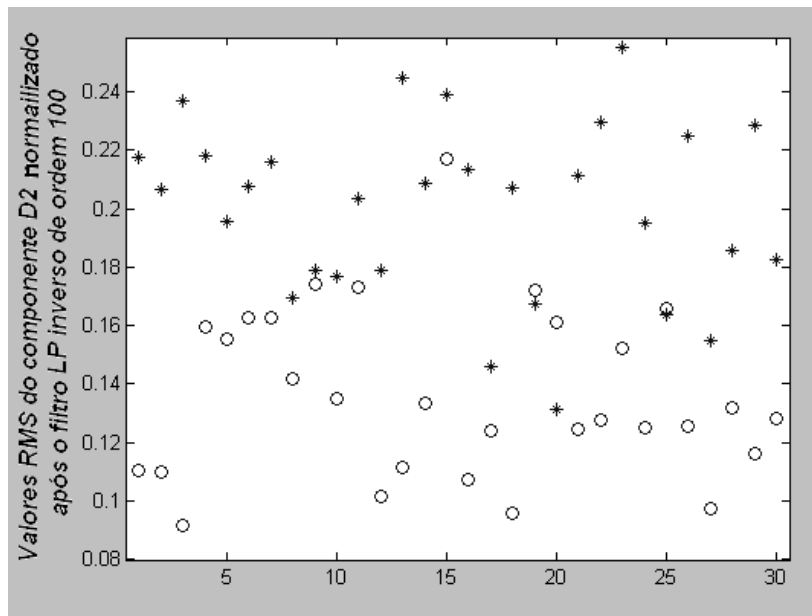


Figura 10 – Valores RMS do componente de detalhe D_2 da DWTdb10 normalizado pelo módulo de seu valor de pico após o filtro LP inverso de ordem 100 para cada um dos 60 sinais de voz. Os pontos * e o correspondem às vozes patológicas e normais, respectivamente.

É possível notar nos gráficos anteriores que a separação entre as classes (vozes normais e patológicas) se torna bem mais evidente após a utilização do filtro LP inverso de ordem $M = 100$.

Foram ainda obtidas as densidades espectrais de potência (*DSP – Density Power Spectrum*) dos componentes normalizados *RnD2* da DWTdb10 antes e após o filtro LP inverso para os 30 sinais de voz normal e 30 patológicas, de maneira a analisar a atuação do filtro. A seguir, encontram-se os gráficos com as densidades espectrais dos componentes *RnD2* (antes e após o filtro) de 4 sinais de voz normal (Figura 11) e 4 sinais de voz patológica (Figura 12).

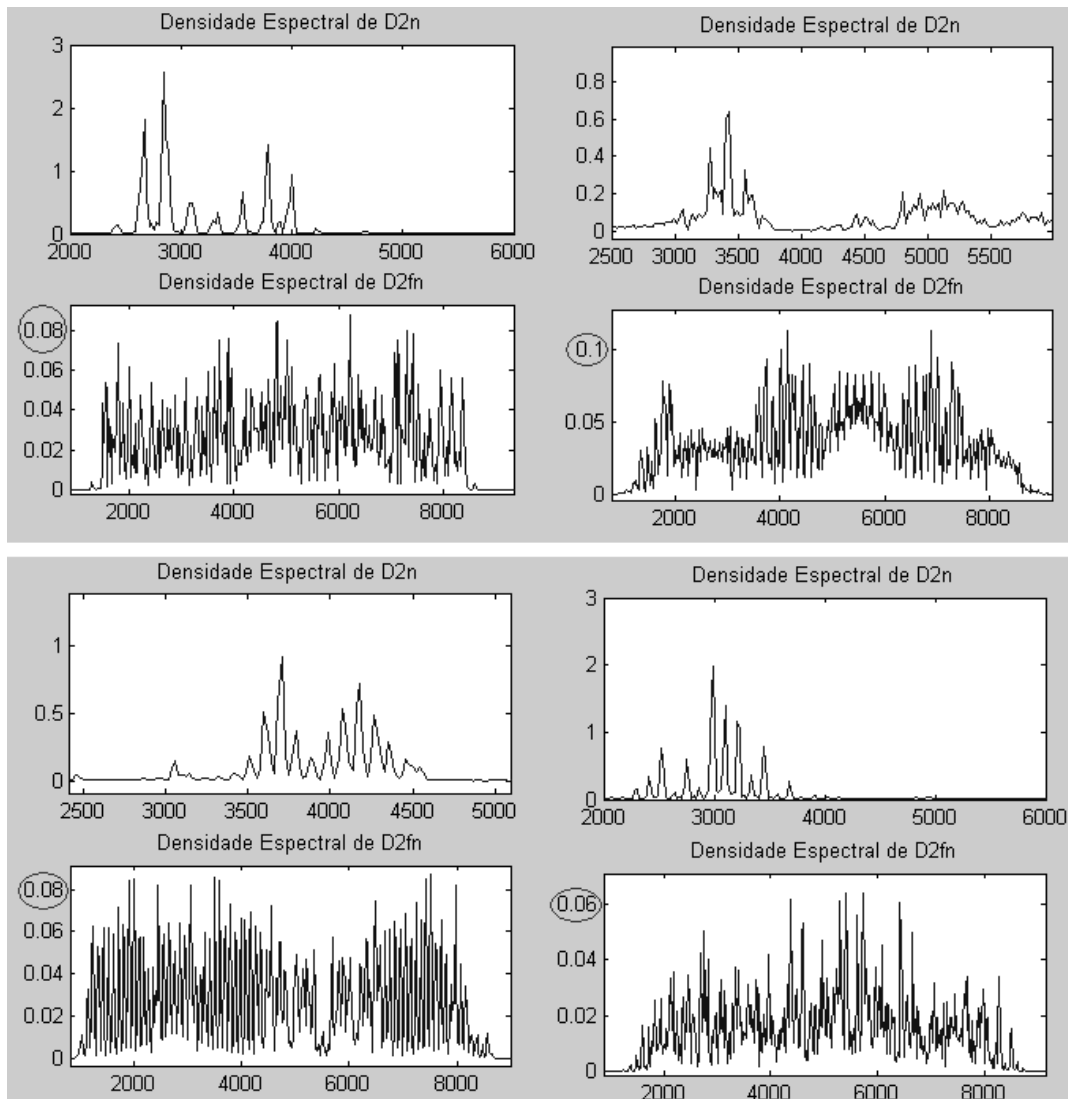


Figura 11 – Densidades espectrais de potência (DSP) do componente de detalhe normalizado RnD_2 da DWTdb10 de 4 sinais de voz normal, antes (D2n) e após o filtro LP inverso (D2fn).

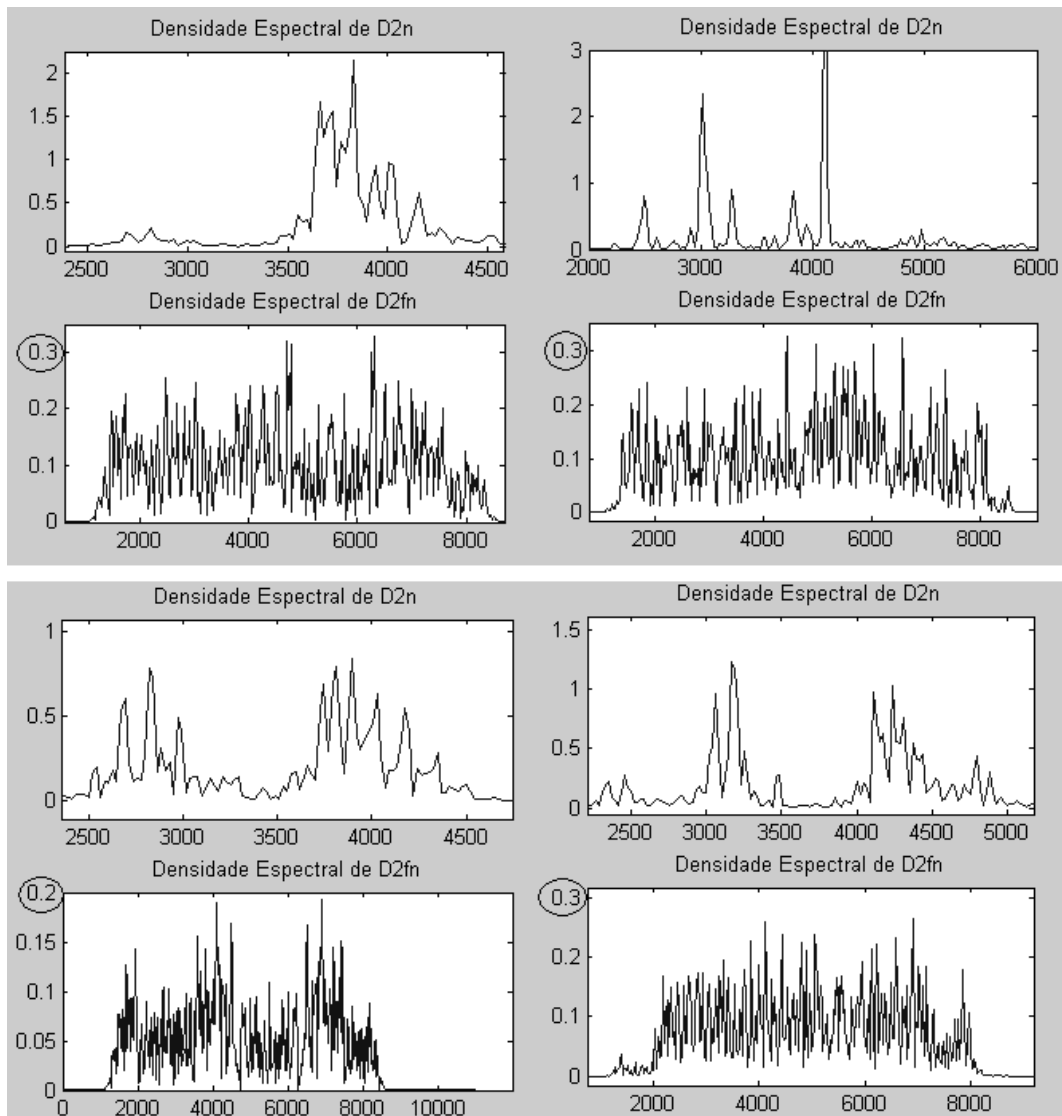


Figura 12 – Densidades espectrais de potência (DSP) do componente de detalhe normalizado RnD_2 da $DWTdb_{10}$ de 4 sinais de voz patológica, antes ($D2n$) e após o filtro LP inverso ($D2fn$).

Como já esperado e conforme as figuras anteriores, com a frequência de amostragem de 22050Hz dos sinais deste banco de dados, os componentes de detalhe D_2 nos traz informações nas frequências na faixa de 2756,25 a 5512,5Hz (MALLAT, 1989). Também de acordo com os gráficos anteriores, ficam evidentes os maiores valores RMS dos componentes $D2fn$ oriundos de vozes patológicas (picos

em torno de 0,3) se comparados com os oriundos de vozes normais (picos em torno de 0,08). Este resultado corrobora o que já era esperado segundo os gráficos das figuras anteriores que mostram que os valores RMS das vozes patológicas (pontos *) são realmente superiores aos das vozes normais (pontos o). Isto certamente é uma consequência da turbulência no fluxo glotal, resultante do mau funcionamento das pregas vocais devido à presença do nódulo, o que pode ser quantificado pelo ruído nos componentes espectrais da fala (ROSA; PEREIRA; GRELLET, 2000). Patologias causadas pelo incompleto fechamento da glote, como os nódulos nas pregas vocais, são usualmente associadas com ruído em alta frequência (GRANQVIST; HAMMARBERG, 2003; MARTINEZ; RUFFINER, 2000). Assim, é encontrado mais ruído no componente de alta frequência D_2 das vozes patológicas do que no mesmo componente das vozes normais. O filtro LP inverso foi usado, portanto, de forma a diminuir ainda mais o ruído existente no componente D_2 oriundo das vozes normais que é inferior ao ruído presente no mesmo componente das vozes patológicas. Deve-se salientar que a normalização destes sinais foi importante para a obtenção dos maiores valores RMS dos sinais de voz patológica, já que o sinal ruidoso normalizado apresenta maior valor RMS que o sinal menos ruidoso. Dessa maneira, fica estabelecida a possibilidade de se utilizar o parâmetro DWT-RMS (FONSECA, 2006) na identificação de sinais vozes patológicas que é baseado na combinação do uso da DWT com o filtro LP inverso.

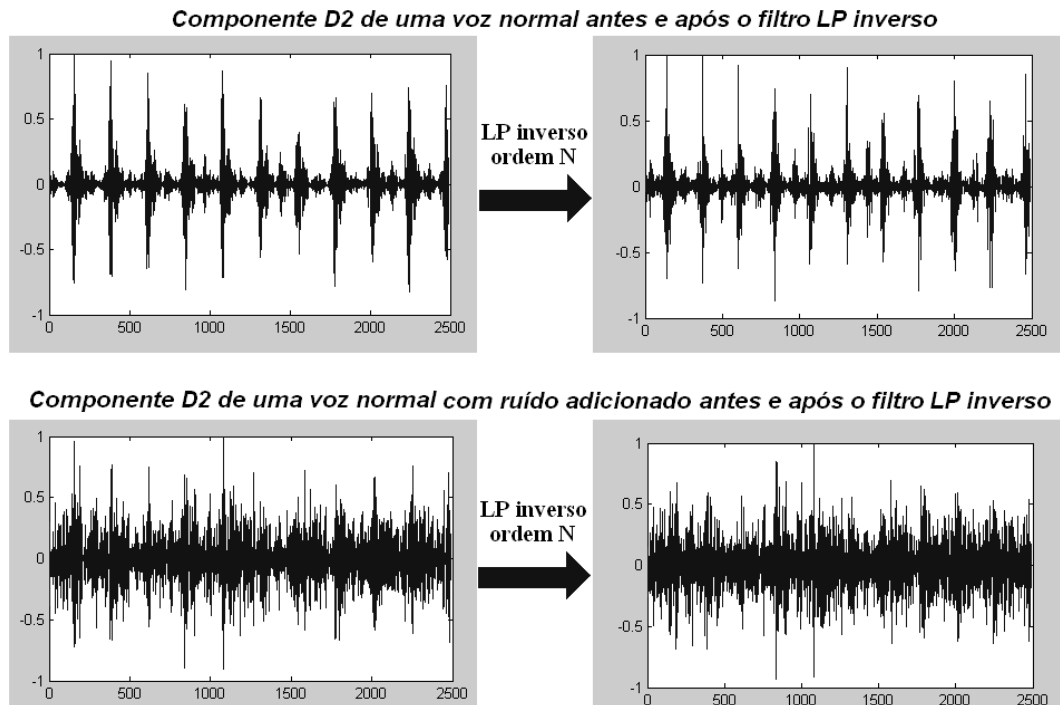


Figura 13 – Os dois primeiros gráficos correspondem aos componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes normais antes e após a aplicação do filtro digital inverso LP e os dois últimos correspondem aos componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes normais com ruído adicionado ao sinal antes e após a aplicação do filtro digital inverso LP.

Os sinais da figura 13 mostram que o ruído presente no componente D_2 da voz normal foi subtraído com o uso do filtro LP inverso. É possível ainda concluir destes sinais que, ao se adicionar uma maior quantidade de ruído branco a este componente, o mesmo filtro LP inverso não é capaz de subtrair o ruído presente na mesma proporção. Desta maneira, conclui-se que a diminuição do valor RMS dos componentes D_2 oriundos das vozes normais nada mais é do que uma diminuição do ruído branco presente neste componente.

As Figuras 14 e 15 mostram, na mesma escala, três componentes de detalhe D_2 normalizados por seus valores de pico provenientes de vozes normais e patológicas respectivamente. Inicialmente são mostrados sinais de voz depois da

aplicação dos passos 1, 2 e 3 comentados no capítulo anterior e, logo após a seta, são mostrados os sinais após a aplicação do filtro LP inverso. Nos 3 exemplos mostrados nas figuras a seguir, houve uma diminuição relativamente grande dos valores RMS para os sinais provenientes de vozes normais após o uso do filtro LP inverso, o que não ocorre para os sinais provenientes de vozes patológicas. Dessa maneira, este filtro LP aumentou consideravelmente a margem de diferenças entre as classes.

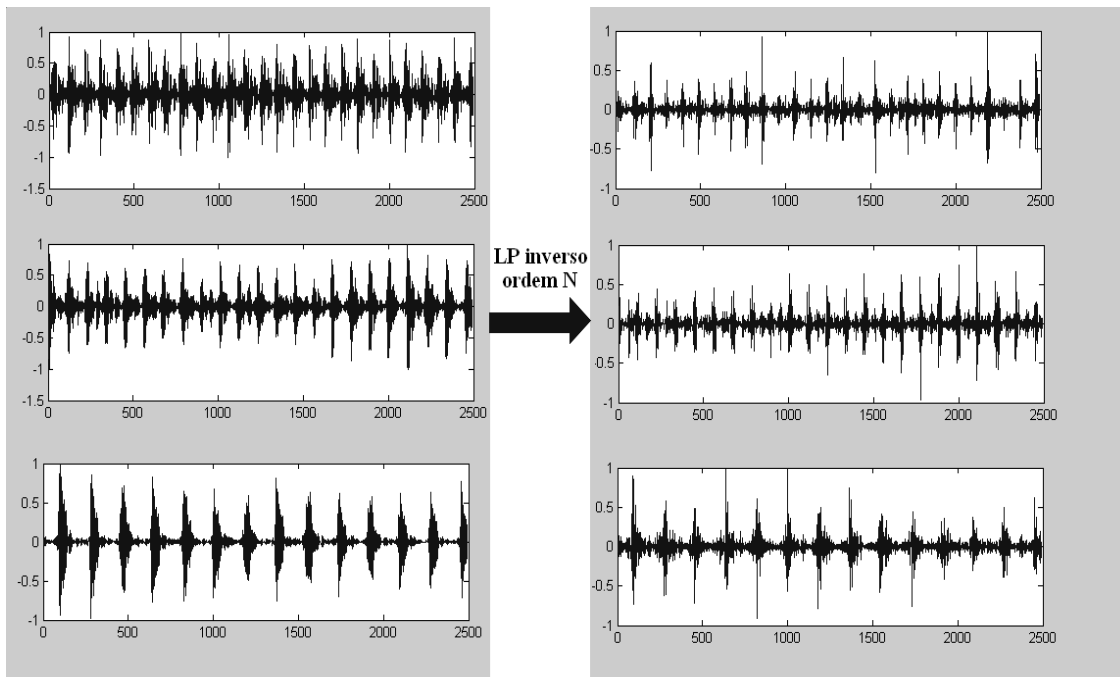


Figura 14 – Componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes normais antes e após a aplicação do filtro digital inverso LP.

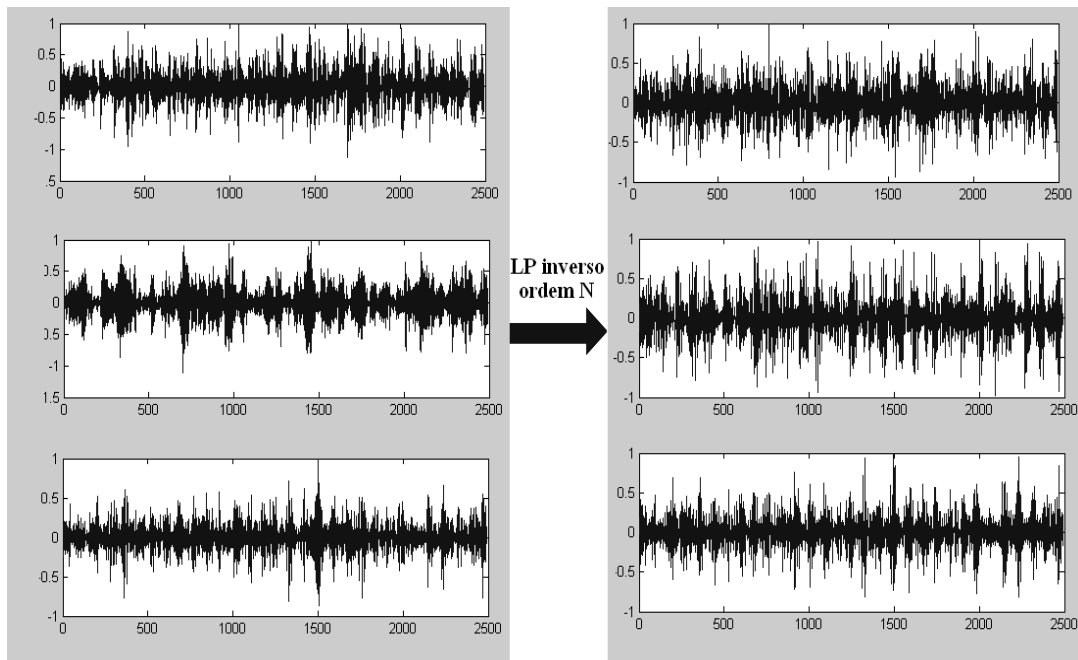


Figura 15 – Componentes de detalhe D_2 normalizados da DWT-db10 provenientes de vozes patológicas antes e após a aplicação do filtro digital inverso LP.

Particularmente para as wavelets de Daubechies com suporte 8 (maior DRNP obtida (Tabela 3) junto as Daubechies com suportes 12 e 16), a Figura 16 mostra as médias aritméticas dos valores RMS_{RD2} de todos os sinais de vozes normais e patológicas de nosso banco de dados em função da ordem N do filtro LP inverso. A partir desta figura é possível notar que a diferença relativa entre os valores RMS_{RD2} provenientes de vozes normais e patológicas se torna maior com o aumento da ordem. Aplicando-se o filtro LP inverso, a diferença relativa entre os RMS_{RD2} das vozes normais e patológicas subiu de 20,8% (Tabela 3) para 36% (Figura 16). Pode-se notar que, para as vozes normais, a média dos RMS_{RD2} segue uma curva com decaimento tendendo a uma exponencial decrescente, enquanto para as vozes patológicas, mesmo com o aumento da ordem N , as médias se mantiveram praticamente em um valor constante. Dessa maneira, é possível afirmar que com o valor de diferença entre as médias das classes bastante superior após a utilização

do filtro LP inverso com esta ordem N específica, existe então um melhor intervalo de confiança para que os valores RMS sejam estatisticamente diferentes (ZAR, 1984). Ordens superiores não melhoram este resultado, como é também possível notar através da figura. Estes valores indicam quanto o modelo LPC filtra o sinal RD_2 normalizado. Na Figura 16, baixos valores RMS indicam uma melhor filtragem, enquanto altos valores indicam uma filtragem ruim. Com o aumento da ordem os valores RMS_{RD_2} provenientes das vozes normais decresceram cerca de 20% enquanto os provenientes das vozes patológicas se mantiveram praticamente constante. Como o RMS dos sinais de voz normal é menor que o RMS dos sinais de voz patológica, o filtro LP inverso dos sinais RD_2 normalizados provenientes de vozes normais pode ser mais bem modelado pelo LPC que os RD_2 normalizados de vozes patológicas (FONSECA, 2007b).

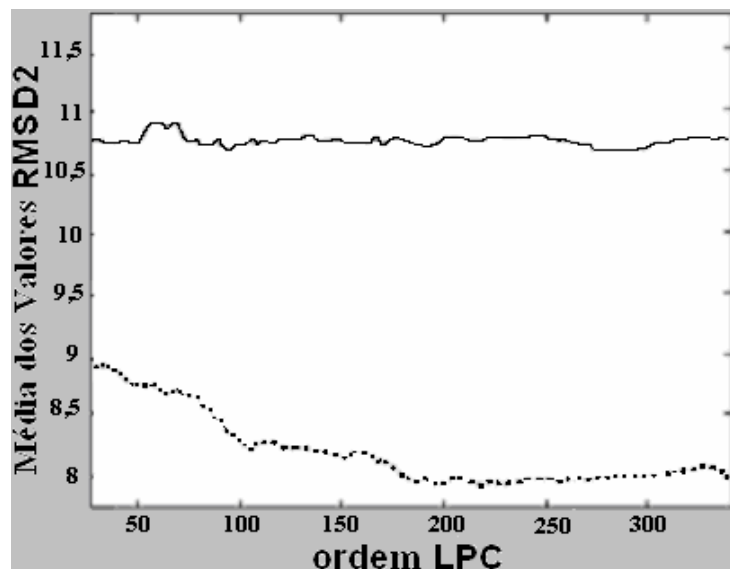


Figura 16 – Médias dos valores RMS_{RD_2} de todas as vozes normais (traço pontilhado) e de todas as vozes patológicas (traço contínuo) de nosso banco de dados para cada ordem N do filtro LP inverso.

Aplicando-se então o filtro LP inverso e utilizando-se o valor médio quadrático (RMS) dos sinais resultantes deste algoritmo, pode-se criar um eixo horizontal com

valores de 1 a 30 (número total dos sinais de voz utilizados) apenas com o objetivo de melhor visualizar a distribuição dos pontos, obteve-se então o seguinte gráfico na figura a seguir:

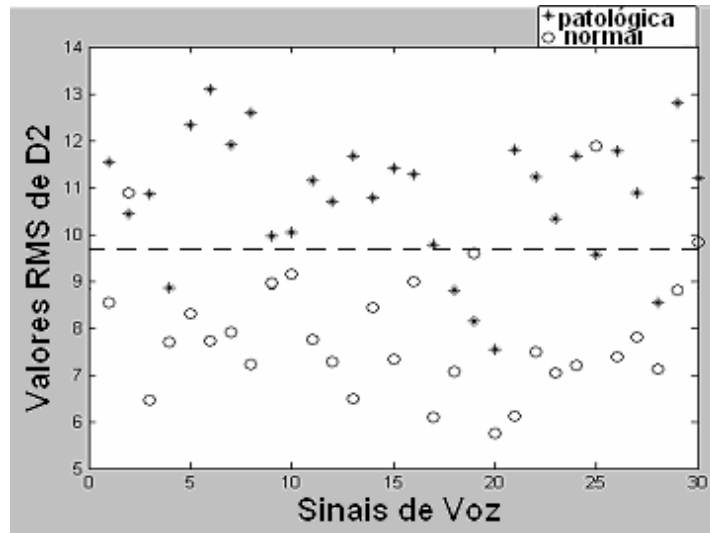


Figura 17 – Gráfico com os valores RMS^2 dos componentes D_2 da DWTdb10 normalizados pelo valor de pico após o uso do filtro LP inverso para as vozes normais e patológicas.

A fim de se obter uma melhor classificação entre os sinais patológicos e não-patológicos, utilizou-se o algoritmo de reconhecimento de padrões, LS-SVM com o kernel RBF. Dessa maneira, uma função não-linear foi obtida de tal forma a separar as classes. Foram utilizados os 60 sinais de voz (30 normais e 30 de pacientes com nódulos nas pregas vocais) e os respectivos valores médios quadráticos provenientes dos coeficientes da segunda escala (D_2) da DWT-db10 constituíram as entradas para o classificador LS-SVM. Cinquenta (50) valores foram utilizados para o treinamento da classificação e dez (10) para o teste do algoritmo. Apenas com o objetivo de se visualizar graficamente da mesma forma da figura anterior, podem ser

² Os valores RMS que aparecem no gráfico estão normalizados pelo inverso da raiz quadrada do número de amostras deste componente (0,02).

inseridos dois atributos de entrada utilizando um número para cada par de sinais patológico/não-patológico com os respectivos valores médios quadráticos provenientes da DWTdb10. Chega-se então ao seguinte gráfico para a classificação dos 50 sinais utilizados no treinamento do LS-SVM:

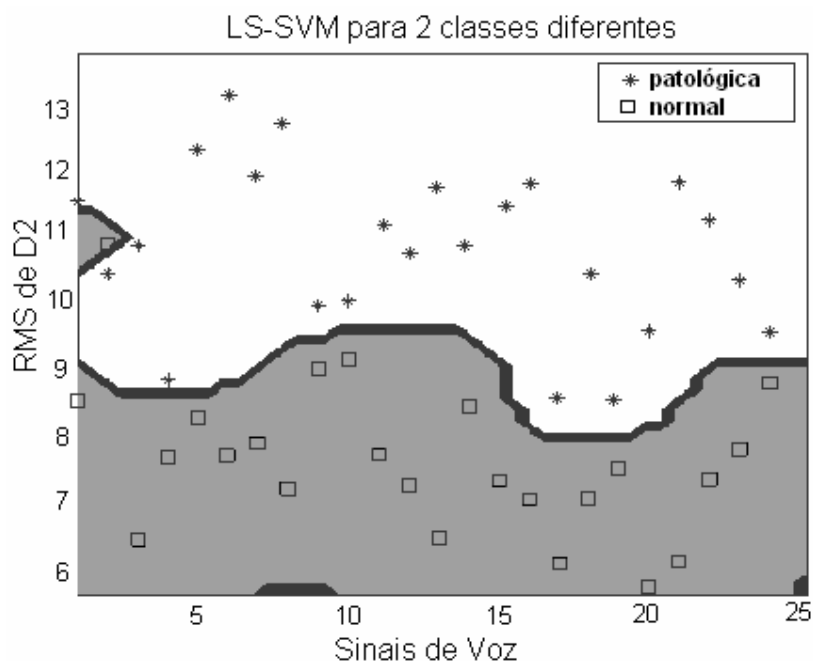


Figura 18 – Visualização gráfica do classificador LS-SVM com 2 atributos de entrada³.

Na tabela seguinte, são apresentados os 10 dados utilizados no conjunto de teste ou validação, escolhidos de maneira aleatória, entre os 60 valores totais utilizados no classificador, somando-se os dados para o treinamento e os dados para validação.

³ Deve-se enfatizar que este gráfico foi criado apenas para uma visualização bidimensional da distribuição dos pontos, onde o eixo horizontal são valores de 1 a 25 (número total de sinais de voz utilizados para o Treinamento). No entanto ele não corresponde ao treinamento e teste deste classificador que, para este caso, é unidimensional (apenas um atributo de entrada).

Tabela 5 – Conjunto de teste escolhido aleatoriamente para o classificador.

<i>Sinais de Voz</i>	<i>Valores médios quadráticos⁴ do componente D_2 normalizado da DWT-db10</i>
1	9.5778
2	11.8062
3	11.2360
4	10.3381
5	9.5637
6	7.7134
7	7.8989
8	7.2158
9	8.9659
10	9.1515

Aplicando simplesmente um limiar de separação entre as classes com valor 8,85, obteve-se 82% de identificação correta das vozes patológicas para os 50 sinais utilizados para estabelecer este limiar e 80% de acerto para o conjunto de teste. Utilizando agora o classificador LS-SVM, tendo como entrada unidimensional os valores de energia provenientes da DWT-db10, pôde-se obter 92% de identificação correta das vozes patológicas para os 50 sinais usados no treinamento do classificador e 100% de acerto para o conjunto de teste (FONSECA et al., 2005).

Utilizando agora o classificador LS-SVM para obter o melhor hiperplano de classificação com 3 diferentes funções kernels, assim como utilizando todas as combinações possíveis dos valores RMS_{RD1} , RMS_{RD2} , RMS_{RD3} e RMS_{RD4} foram obtidas então as seguintes percentagens de acerto mostradas na Tabela 6 a seguir

⁴ Os valores RMS que aparecem na tabela estão normalizados pela raiz quadrada do número de amostras.

para os dados de teste ou validação do classificador. Diferente da função kernel linear, a RBF e a MLP podem ser utilizadas com ótimas porcentagens de acerto em problemas que envolvem um espaço de entrada não-linear. O MLP constitui uma função tangente hiperbólica (FRONTZEK; ECKMILLER, 2001) também bastante utilizada nos algoritmos de redes neurais *multi-layer perceptron* (HAYKINV, 1999). Já o kernel RBF é dado por:

$$f(x_i, x_j) = e^{\left(\frac{1}{2\lambda^2} \|x_i - x_j\|^2 \right)}, \quad (42)$$

onde x_i e x_j são os pontos do banco de dados e λ é um parâmetro utilizado no treinamento da classificação de acordo com a referência (SUYKENS; VANDEWALLE, 1999). Deve-se salientar que se utilizou 20% dos dados (12 pontos, 6 de vozes normais e 6 de vozes patológicas) para esta validação, e 80% (48 pontos, 24 de vozes normais e 24 de vozes patológicas) para o treinamento e obtenção do melhor hiperplano de classificação.

A Figura 19 mostra os valores dos RMS_{RD1} e RMS_{RD2} usados para o treinamento e teste do classificador LS-SVM.

A técnica proposta mostra que os sinais reconstruídos, $RD_1 - RD_4$, distinguem claramente entre vozes normais e patologicamente afetadas de acordo com o treinamento dado ao classificador. Particularmente, sinais de voz normal apresentam menores valores RMS que os mesmos componentes para sinais de voz provenientes de pacientes com nódulos nas pregas vocais.

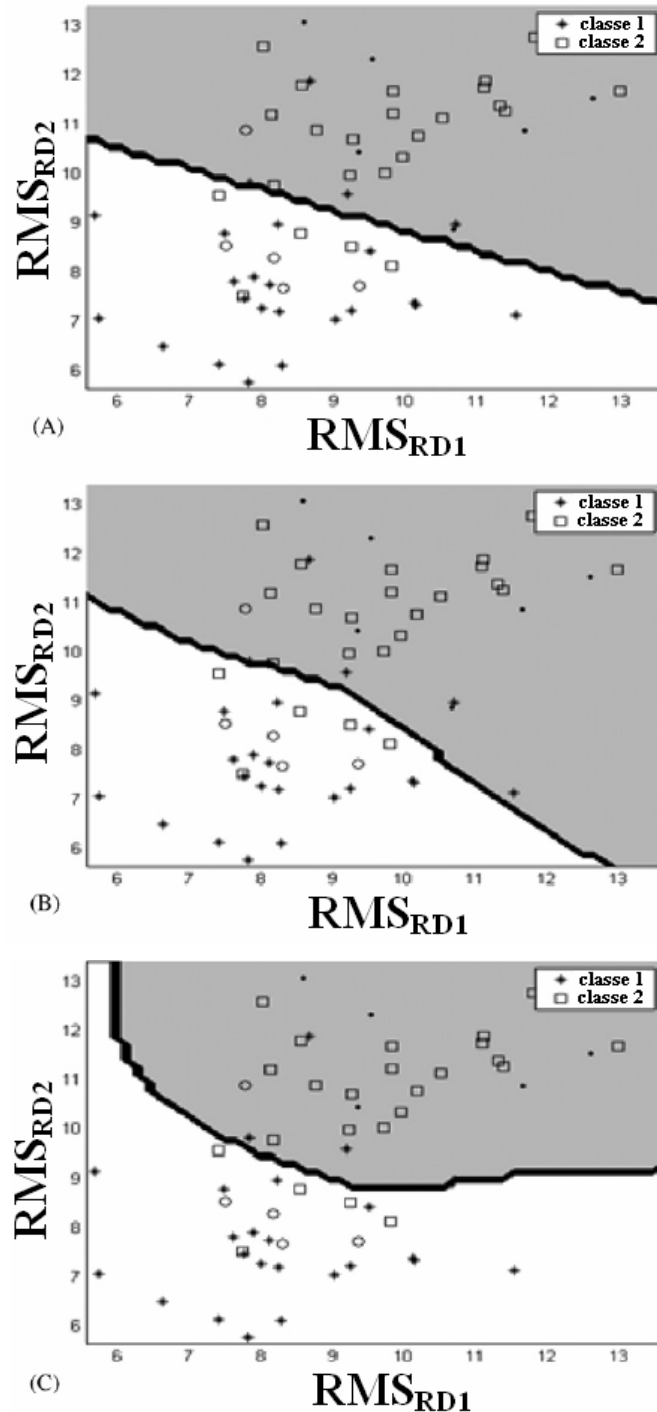


Figura 19 – Esboços gráficos da classificação com os valores RMS dos componentes já filtrados pelo LP inverso, RD_1 (eixo x) e RD_2 (eixo y), para os kernels linear (em (A)), MLP (em (B)) e RBF (em (C)) respectivamente. Os pontos * e □ foram usados para o treinamento do classificador e eles são provenientes, respectivamente, de vozes normais e patologicamente afetadas. Os pontos ○ e ● foram usados para o teste ou validação do classificador e são, respectivamente, provenientes de vozes normais e patologicamente afetadas.

Tabela 6 – Porcentagem de acerto nos dados de teste ou validação do classificador.

Kernels/combinacões	Linear	RBF	MLP
RMS_{RD1}	83,33	83,33	91,67
RMS_{RD2}	83,33	83,33	83,33
RMS_{RD3}	75	75	75
RMS_{RD4}	75	75	75
$RMS_{RD1} - RMS_{RD2}$	91,67	83,33	91,67
$RMS_{RD1} - RMS_{RD3}$	83,33	83,33	83,33
$RMS_{RD1} - RMS_{RD4}$	83,33	83,33	75
$RMS_{RD2} - RMS_{RD3}$	83,33	83,33	83,33
$RMS_{RD2} - RMS_{RD4}$	91,67	91,67	83,33
$RMS_{RD3} - RMS_{RD4}$	75	75	75
$RMS_{RD1} - RMS_{RD2} - RMS_{RD3}$	83,33	83,33	83,33
$RMS_{RD1} - RMS_{RD2} - RMS_{RD4}$	91,67	83,33	91,67
$RMS_{RD1} - RMS_{RD3} - RMS_{RD4}$	83,33	83,33	83,33
$RMS_{RD2} - RMS_{RD3} - RMS_{RD4}$	83,33	83,33	83,33
$RMS_{RD1} - RMS_{RD2} - RMS_{RD3} - RMS_{RD4}$	83,33	83,33	75

É importante notar que, para este algoritmo, foi utilizada a DWTdb10, não sendo a função wavelet desta família que apresentou os melhores resultados estatísticos de confiança na classificação, conforme Tabela 3 obtida após este estudo. O classificador LS-SVM com 3 diferentes funções kernel (linear, RBF e MLP) levou a um adequado classificador de voz patológica com mais de 90% de classificação correta, melhorando bastante o resultado anterior com o mesmo banco de dados e sem o uso de nenhum algoritmo de inteligência artificial para o reconhecimento de padrões (FONSECA et al., 2006). A função kernel linear levou a 91,67% de acerto para as combinações de entrada $RMS_{RD1} - RMS_{RD2}$, $RMS_{RD1} - RMS_{RD4}$ e $RMS_{RD1} - RMS_{RD2} - RMS_{RD4}$, o kernel RBF levou a 91,67% de acerto para as combinações $RMS_{RD2} - RMS_{RD4}$ e o MLP levou a 91,67% de acerto para RMS_{RD1} , $RMS_{RD1} - RMS_{RD2}$ e $RMS_{RD1} - RMS_{RD2} - RMS_{RD4}$. Os valores dos parâmetros

usados para o treinamento do classificador LS-SVM foram os melhores para evitar o processo de “*over-fitting*”, situação na qual o classificador se torna muito especializado para o banco de dados especificamente utilizado durante o treinamento (HADDJITODOROV; BOYANOV; TESTON, 2000).

Utilizando-se agora os sinais de voz oriundos de pacientes com edema de Reinke, além daqueles oriundos de pacientes com nódulo nas pregas vocais foram obtidos diferentes hiperplanos entre as classes. Nesta análise foram utilizados 25 sinais de pacientes com nódulo, 16 sinais de pacientes com edema de Reinke e 30 sinais de pessoas sem nenhuma patologia vocal. A Figura 20 (A) mostra o limiar obtido no classificador LS-SVM treinado com os valores RMS dos componentes RD_1 e RD_2 para distinguir entre as vozes normais e as vozes de pacientes com edema de Reinke. A Figura 20 (B) mostra o limiar obtido no classificador LS-SVM treinado com os valores RMS dos componentes RD_1 e RD_2 para distinguir entre as vozes normais e de pacientes com nódulo nas pregas vocais. As figuras mostram também um dos grupos de teste utilizados para a validação destes classificadores. Foram usados 6 pontos para o teste e os demais pontos de cada classe foram usados para o treinamento.

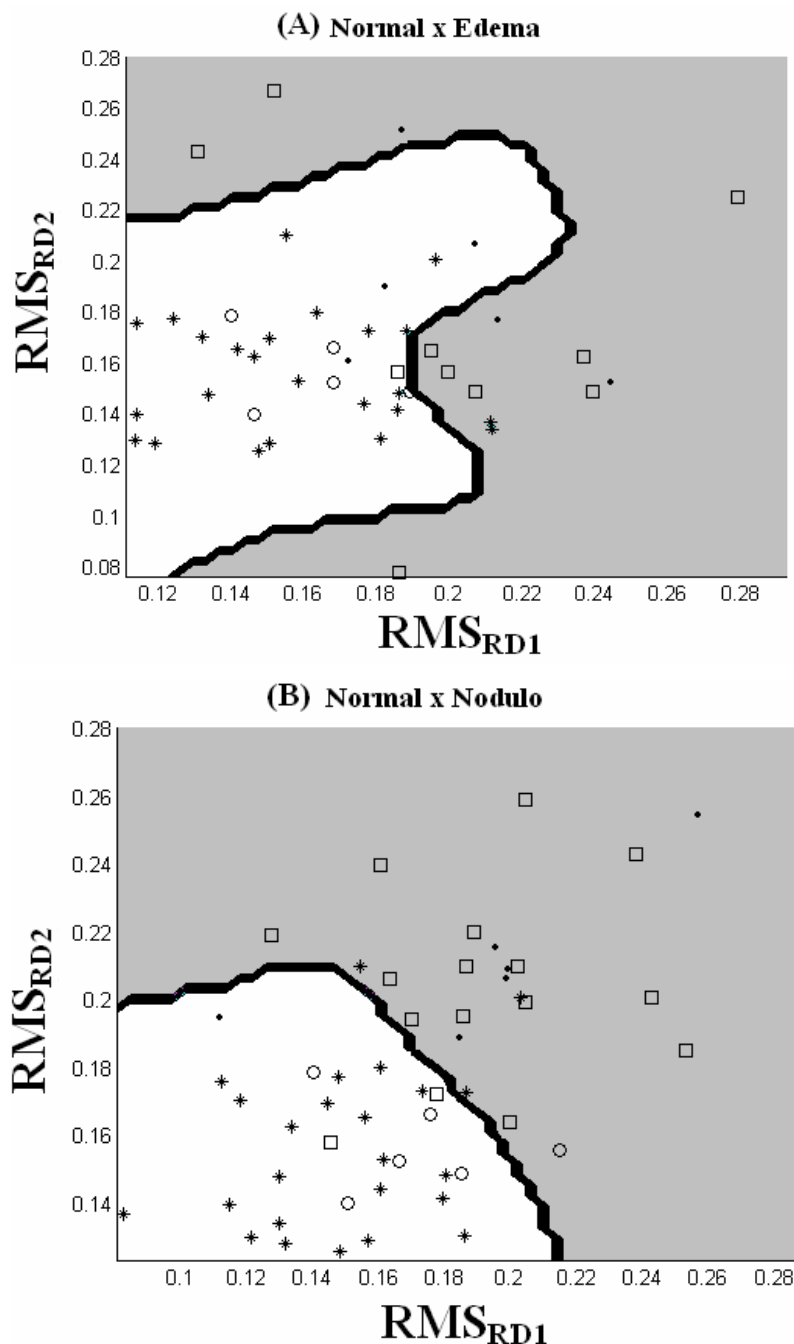


Figura 20 – Esboços gráficos da classificação com os valores RMS dos componentes já filtrados pelo filtro LP inverso, RD_1 (eixo x) e RD_2 (eixo y), para vozes normais e vozes de pacientes com edema de Reike em (A) e para vozes normais e vozes de pacientes com nódulo nas pregas vocais em (B). Os pontos * e □ foram usados para o treinamento do classificador e eles são provenientes, respectivamente, de vozes normais e patologicamente afetadas. Os pontos o e ● foram usados para o teste ou validação do classificador e são, respectivamente, provenientes de vozes normais e patologicamente afetadas.

A Tabela a seguir mostra os acertos obtidos na classificação após a validação cruzada com o uso de cada um dos atributos de entrada (valores RMS dos componentes RD_1 , RD_2 e RD_3). Os valores RMS destes parâmetros identificaram o ruído em alta frequência presentes nas duas patologias e que não foi identificado nos sinais de voz normal. Dessa forma, ao utilizar estes mesmos valores como atributos de entrada para a classificação entre as duas diferentes patologias, nenhum acerto significativo de classificação foi obtido, já que o ruído encontra-se presente nos dois grupos de sinais.

Tabela 7 – Acertos na Classificação do LS-SVM (Normal x Nódulo; Normal x Edema; Normal x Patológica).

Atributos de Entrada do LS-SVM	Normal x Nódulo (%)	Normal x Edema (%)	Normal x Patológica (%)
RMS_{RD1}	84.3	78.6	82.9
RMS_{RD2}	90.1	85.3	88.2
RMS_{RD3}	73	67.8	70.1

A validação cruzada dos valores RMS do componente wavelet RD_2 apresentou o melhor resultado de classificação. Estes resultados mostraram que os sinais reconstruídos de detalhe nos primeiro, segundo e terceiro níveis de decomposição da DWTdb10, correspondentes aos componentes de mais alta frequência do sinal, distinguiram claramente entre os sinais de voz normal e aqueles provenientes de pacientes com edema de Reinke e nódulo nas pregas vocais deste banco de dados. Pode-se notar que os sinais de voz normal apresentaram menores valores RMS que os sinais de voz patológica, como pode ser visualizado na Figura 20. Certamente, isto é uma consequência da maior presença de ruído em alta frequência detectado pela DWTdb10 nos sinais de voz patológica. É importante afirmar que a ordem 20 foi a melhor função wavelet de Daubechies em termos da

diferença entre os valores RMS dos dois grupos. Ordens menores que esta foram também testadas e atingiram diferenças inferiores entre os valores RMS dos grupos.

Utilizando-se o software comercial *Análise de Voz*, versão 6.0 (SCALASSARA et al, 2007), foi calculado o valor de *jitter* dos componentes wavelet na faixa das freqüências fundamentais dos sinais de voz a partir do quinto nível de decomposição, RD_6 , RA_5 e RA_6 . A figura a seguir mostra estes componentes reconstruídos para um sinal de voz oriundo de um paciente com nódulo nas pregas vocais.

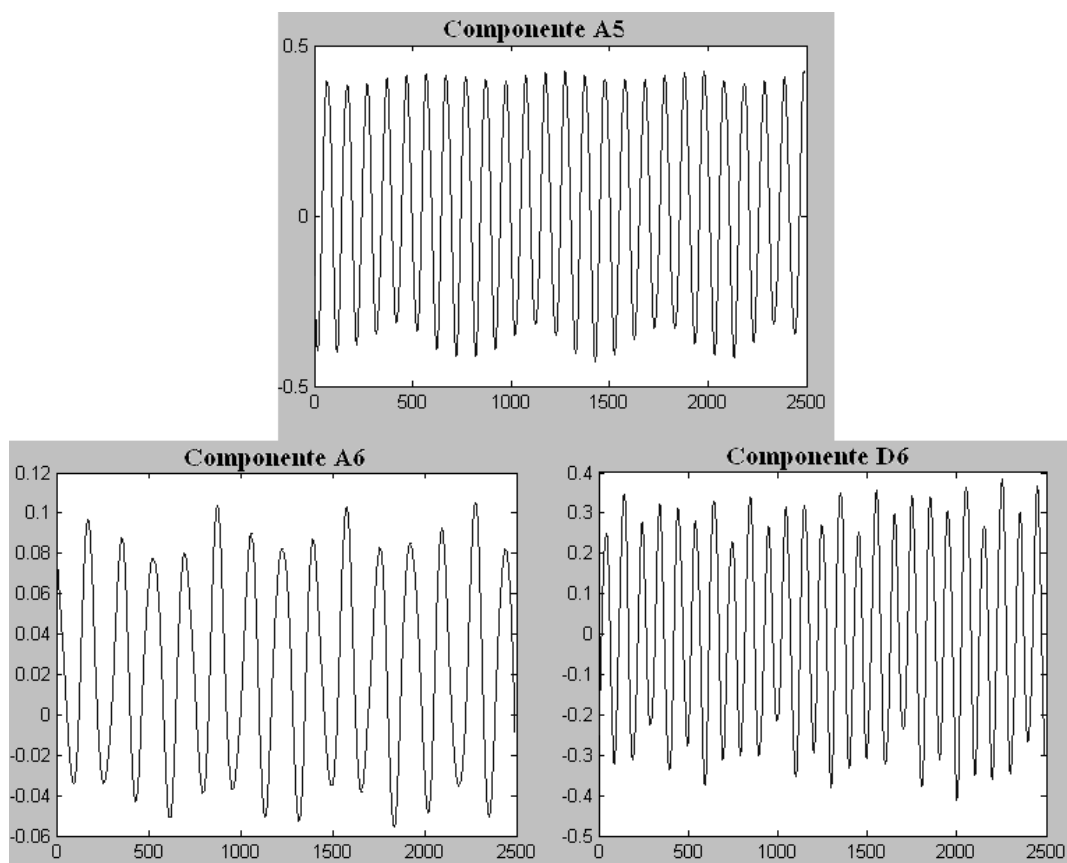


Figura 21 – Componentes reconstruídos de aproximação e detalhe nos níveis 5 e 6 obtidos da decomposição da DWTdb10 de um sinal de voz oriundo de um paciente com nódulo nas pregas vocais.

Os valores de *jitter* dos componentes wavelet na faixa da frequência fundamental dos sinais (RD_6 , RA_5 e RA_6) foram usados como atributos de entrada no classificador LS-SVM para distinguir entre os grupos de sinais oriundos de pacientes com as 2 diferentes patologias vocais. Estes valores de *jitter* dos componentes wavelet RD_6 , RA_5 ou RA_6 (de acordo com a similaridade entre a frequência fundamental destes componentes com a frequência fundamental dos respectivos sinais de voz) ofereceram resultados significativos para classificação entre as duas patologias. Os limiares definidos por estes classificadores e os acertos obtidos na classificação serão discutidos posteriormente.

A Figura a seguir mostra os valores de *jitter* dos componentes wavelet obtidos para as 16 vozes de pacientes com edema de Reinke e as 25 vozes de pacientes com nódulo nas pregas vocais. O componente wavelet usado para cada sinal foi RD_6 , RA_5 ou RA_6 , de acordo com aquele cuja frequência fundamental deste componente fosse a mais similar possível com a frequência fundamental do respectivo sinal analisado. Deve-se notar que, mesmo que RD_6 ou RA_6 pudessem atingir mais precisamente a faixa de frequências que contém a frequência fundamental dos sinais, RA_5 apresentou, em alguns casos, uma frequência fundamental mais similar à frequência fundamental do respectivo sinal de voz do que os demais componentes.

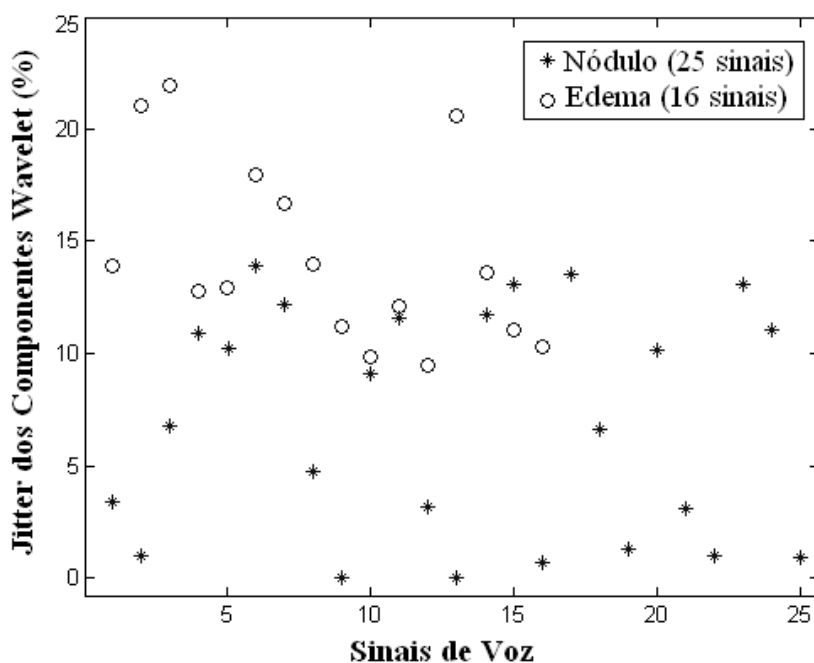


Figura 22 – Valores de Jitter dos componentes wavelet RD_6 , RA_5 ou RA_6 de acordo com aquele cuja frequência fundamental fosse a mais similar possível com a frequência fundamental do respectivo sinal de voz decomposto.

Deve-se lembrar que, nestes níveis de decomposição (quinto e sexto), os componentes espectrais de alta frequência (presentes em ambas as patologias devido a soprosidade) foram filtrados pela decomposição. Desta maneira, pode-se concluir que o *jitter* presente em diferentes intensidades nos componentes wavelet pôde ser mais bem avaliado, distinguindo as duas diferentes patologias após esta filtragem. A maior instabilidade presente nos sinais oriundos de pacientes com edema de Reinke pode ser uma das razões que levaram a valores maiores de *jitter*, conforme é observado na figura anterior, do que os encontrados em pacientes com nódulo nas pregas vocais.

A Tabela a seguir mostra os acertos obtidos no classificador LS-SVM após a validação cruzada utilizando como atributos de entrada os valores de *jitter* de cada

um dos componentes do quinto e sexto nível de decomposição (RD_6 , RA_5 ou RA_6).

Tabela 8 – Acertos na Classificação do LS-SVM para a distinção entre as duas patologias vocais (Nódulo e Edema).

Atributos de Entrada do LS-SVM	Nódulo x Edema (%)
Jitter dos Componentes Wavelet RA_5	72
Jitter dos Componentes Wavelet RA_6	68.7
Jitter dos Componentes Wavelet RD_6	62.3
Jitter dos Componentes Wavelet (RA_5 , RA_6 ou RD_6 dependendo da similaridade com a frequência fundamental do respectivo sinal de voz decomposto)	82.4

Desta maneira, pode-se concluir que o LS-SVM levou a um classificador adequado com mais de 90% de acerto entre os sinais de voz normal e de pacientes com nódulo nas pregas vocais (Tabela 7), mais de 85% de acerto entre as vozes normais e de pacientes com edema de Reinke (Tabela 7), e acima de 82% entre as duas diferentes patologias (Tabela 8). Os valores dos parâmetros de ajuste para o treinamento do classificador foram os melhores obtidos para evitar o processo de *over-fitting*, situação na qual o classificador se torna demasiadamente especializado naquele grupo utilizado para o treinamento, resultando em uma falha nos testes de validação. Estes resultados de classificação utilizando os sinais oriundos de pacientes com nódulo e edema de Reinke foram submetidos em artigo para revista IEEE – EMBM (*Engineering in Medicine and Biology Magazine*).

Capítulo 6

CONCLUSÃO E SUGESTÕES

Conclui-se deste trabalho que os algoritmos de análise tempo-freqüência da DWT associados ao filtro LP inverso da vogal sustentada /a/ podem levar a um parâmetro de classificação adequada para a distinção de vozes patológicas, principalmente as de pacientes com edema de Reinke e nódulo nas pregas vocais. Com a medida do *jitter* em determinados componentes é também possível classificar as duas diferentes patologias. Nota-se uma importante relação existente entre a escala frequencial dos componentes de detalhe em altas freqüências e a ordem utilizada para o filtro LP inverso, bem como melhorias obtidas na classificação com o método de inteligência artificial LS-SVM levando a mais de 90% de acerto entre os sinais de voz normal e de pacientes com nódulo nas pregas vocais, mais de 85% de acerto entre as vozes normais e de pacientes com edema de Reinke, e acima de 82% entre as duas diferentes patologias.

Diferentes funções wavelets podem ainda ser testadas para melhoria do algoritmo. Torna-se também interessante um estudo aprofundado na obtenção de características particulares para uma função wavelet própria conveniente a este propósito com variações de ordens e parâmetros matemáticos em sua construção.

As propriedades do banco de filtros podem, dessa forma, serem modificadas objetivando o maior acerto possível na identificação das vozes oriundas de pacientes com diferentes patologias no trato glotal, inclusive aquelas diferentes do edema de Reinke e dos nódulos nas pregas vocais. Outros tipos de árvores de decomposição como as utilizadas pela Transformada Wavelet Packet podem também ser testadas para se buscar novos parâmetros de classificação.

Além disso, estudos mais aprofundados relacionados às relações existentes entre os valores de acerto na classificação e os suportes dos filtros wavelet, a ordem do filtro LP inverso, bem como, outros métodos de inteligência artificial aplicados para o reconhecimento de padrões como os *Hidden Markov Models*, algoritmos genéticos, redes neurais podem ainda ser usados para a comparação com os resultados já obtidos com o LS-SVM.

REFERÊNCIAS BIBLIOGRÁFICAS

- ACCARDO, A. P.; MUMOLO, E. (1998). An algorithm for the automatic differentiation between the speech of normal and patients with Friedreich's ataxia based on the short-time fractal dimension. **Comp. Biol. Med.**, v. 28, n. 1, p. 75-89.
- ANANTHAPADMANABHA, T.V.; YEGNANARAYANA, B. (1979). Epoch extraction from linear prediction residual for identification of closed glottis interval. **IEEE Trans. on Acoustic, Speech and Signal Processing**, v. 27, n. 4.
- BAKEN, R.J. (1996). Clinical measurement of speech and voice. **Singular Publishing Group Inc.**
- BEHLAU, M. (2001). **Voz O Livro do Especialista**. v.1, ed. Revinter, Rio de Janeiro.
- BENEDAT, J. S. E PIERSON, A. G. (1980). **Engineering Application of Auto Correlation and Spectral Analysis**. John Wiley & Sons, Inc., New York.
- BERTSEKAS, D. P. (1995). Nonlinear Programming. Belmont, MA: **Athenas Scientific**.
- BLESS, D. M. (1991). Measurement of Vocal Function. **Otolaryngologic Clinics of North America**, v. 24, p. 1023–1033.
- BOONE, D. R. E MCFARLANE, S. C. (1994). **A voz e a Terapia Vocal, Artes Médicas**. Porto Alegre.
- BOSI, M.; GOLDBERG, R. **Introduction to Digital Audio Coding and Standards**. 2ed., Kluwer Academic Publishers, Massachusetts, 2003.
- BRUCE, W. SUTER (1998). **Multirate and wavelet signal processing**. San Diego: Academic Press.
- BURGES, C. J. C. (1999). Geometry and invariance in kernel based methods – support vector learning. **MIT Press**, Cambridge, MS, p. 89-116.

CAVALLINI JR., O.; PEREIRA, J. C. (2001). Determinação do nível de soproidade da voz através da transformada wavelet, **Anais do XIII Congresso Argentino de Bioengenharia**, Tafí del Vale, Tucuman, Argentina, Setembro.

CHUI, C. K. (1992). **An Introduction to Wavelets**, Academic Press.

CORTES, C.; VAPNIK, V. (1995). Support Vector Networks. **Machine Learning**, v. 20, p. 273–297.

COHEN, A. (1992). **Ondelettes, analyses multirésolution et traitement numérique du signal**. Tese de Doutorado, Universidade de Paris IX, Daphine.

CRISTIANINI, N.; SHAW-TAYLOR, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. **Cambridge University Press**, Cambridge.

DAVIS, S. B. (1979). Acoustic Characteristics of Normal and Pathological Voices, In Lass, N. J. (Ed.) **Speech and Language: Advances in Basic Research and Practice**, vol. 1, New York Academic Press.

DAUBECHIES, C. I. (1992). **Ten lectures on wavelet**. SIAM.

DAUBECHIES, C. I.; GILBERT C. A. (1997). **Harmonic Analysis, Wavelets and Applications**, New Jersey.

DELLER JR., J. R.; PROAKIS, J. G.; HANSEN, J. H. L. (1993). **Discrete-Time Processing of Speech Signals**. MacMillan Publishing Co., New York.

DENG, L; SHAUGHNESSY, D. (2003). **Speech Processing: A Dynamic and Optimization-Oriented Approach**. New York: Dekker.

EMANUEL, F. W.; LIVELY, M. A. and McCOY, J. F. (1973). Spectral noise levels and roughness ratings for vowels produced by males and females. **Folia Phoniatica**, v. 25, p. 110-120.

ERÇELEBI, E. (2002). Second Generation Wavelet Transform-Based Pitch Period Estimation and Voiced/Unvoiced Decision for Speech Signals. **Elsevier**, Applied Acoustics.

FAHRMEIR, L.; TUTZ, G. (2001). **Multivariate Statistical Modeling Based on Generalized Linear Models**. Springer Series in Statistics, 2nd edition.

FONSECA, E. S.; MACIEL, C. D.; MORI, R. N.; PEREIRA J. C. (2004). A Transformada Wavelet Discreta de Daubechies Aplicada na Identificação de Sinais de Vozes Patológicas. In: CONFERÊNCIA LATINO-AMERICANA DE ENGENHARIA BIOMÉDICA (CLAEB), João Pessoa, PB. **Anais CLAEB 2004**. 1 CD-ROM.

FONSECA, E. S.; GUIDO, R.C.; SILVESTRE, A. C.; MACIEL, C. D.; PEREIRA J. C. (2005). Discrete Wavelet Transform and Support Vector Machine Applied to Pathological Voice Signals Identification. **7th IEEE International Symposium on Multimedia, Special Session on Audio Processing**, Irvine, EUA.

FONSECA, E. S.; GUIDO, R. C.; SCALASSARA, P. R.; MACIEL, C. D.; PEREIRA, J. C. (2007a). Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders, **Elsevier, Computers in Biology and Medicine**, v. 37, p. 571-578.

FONSECA, E. S.; GUIDO, R. C.; PEREIRA J. C.; MACIEL, C. D. (2007b). Combination of LP Inverse Filter and DWT as a New Parameter to Identify Pathology in Voice Signals. **6^o CONGRESSO TEMÁTICO DE DINÂMICA, CONTROLE E APLICAÇÕES (DINCON)**, UNESP, São José do Rio Preto, SP. **Anais DINCON 2007**. 1 CD-ROM.

FRONTZEK, T.; LAL, T. N.; ECKMILLER, R. (2001). Predicting the non-linear dynamics of biological neurons using support vector machines with different kernels, in **International Joint Conference on Neural Networks**, July 2001, v. 2, p. 1492–1497.

GOMES, J.; VELHO, L.; GOLDENSTEIN, S. (1997). **Wavelets: Teoria, Software e Aplicações**, 21^o Colóquio Brasileiro de Matemática, IMPA.

GRAY, A.H.; MARKEL, J.D. (1976). **Linear prediction of speech**, New York, Springer Verlag.

GRANQVIST S.; HAMMARBERG B. (2003). The Correlogram: a Visual Display of Periodicity, **J. Acoust. Soc. Am.** 114 (5), 2934-2945.

GUIDO, R. C.; PEREIRA J. C. (2004). Trying Different Wavelets on the Search for Voice Disorders Sorting. **37th IEEE Southeastern Symposium on System Theory**, EUA, 2004.

GUIDO, R. C.; SLAETS, J. F. W.; KOBERLE, R; ALMEIDA, L. O. B.; PEREIRA, J. C. (2006). A new technique to construct a wavelet transform matched a specified with applications to digital, real-time spike and overlap pattern recognition, **Digital Signal Process**, v. 16, n. 1, p. 24–44.

HADDJITODOROV, S.; BOYANOV B.; TESTON, B. (2000). Laryngeal Pathology Detection by Mean of Class-specific Neural Maps. **IEEE Trans. Inf. Technol. Biomed**, v. 4, n. 1, p. 89-93.

HAYKIN, S. (1999). **Neural Networks – A Comprehensive Foundation**. Prentice-Hall, New Jersey, 2nd edition.

HECKER M. H. L.; KRUEL, E. J. (1970). Description of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency. **Journal of the Acoustical Society of America**, v. 49, n^o 4, part 2, p. 1275-1282.

HILLEBRAND, J.; CLEVELAND, R. A.; ERICKSON, R. L. (1994). Acoustic Correlates of Breathly Vocal Quality. **Journal of Speech and Hearing Research**, v. 37, p. 769–778.

HORI Y. (1979). Fundamental frequency perturbation observed in sustained phonation. **Journal of Speech and Hearing Research**; v. 22, p. 5-19.

HUGH-MUNIER, C. M.; SHERER, K. R.; LEHMANN, W; SHERER, U. (1997). Coping Strategies, Personality and Voice Quality in Patients with Vocal folds Nodules and Polyps, **Journal of Voice**, v. 11, n. 4, p. 452-461.

ISSHIKI, N.; OKAMURA, H.; TANAVE, M.; MORIMOTO, M. (1969). Differential Diagnosis of Hoarseness, **Folia Phoniatica**, v. 21, p. 9-19.

KADAMBE, S.; BOUDREAUX -Bartels G. F. (1992). Application of the wavelet transform for Pitch detection of speech signals, **IEEE Trans. on Inf. Theory**, v. 38, p. 917-924, Março.

KAISER, G. (1994). **A Friendly Guide to Wavelets**, Boston – Basel – Berlin.

KOIKE, Y. (1973). “Application of some acoustic measures for the evaluation of dysfunction”. *Studia Phonologic*, 7, 17-23.

KOIKE, Y. TAKAHASHI, H. e CANLCATERRA, T. C. (1977). Acoustic measures for detecting laryngeal pathology. **Journal of Phonetics**, v. 14, p. 517-524.

LIEBERMAN, P. (1963). Some acoustics measures of the fundamental periodicity of normal and pathologic larynges. **J. Acoustic Society of America**; v. 35, p. 344-53.

MALLAT, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Pattern Anal. and Machines Intell.**, v. 11, n. 7, p. 674–693.

MALLAT, S. G. (1999). **A wavelet tour of signal processing**, San Diego: Academic Press.

MANFREDI, C. (2000) Adaptive noise energy estimation in pathological speech signals. **IEEE Transactions on Biomedical Engineering**, v. 47, issue 11, p. 1538-1543, Novembro.

MARTINEZ, E.; RUFFINER HUGO, L. (2000). Acoustic Analysis of Speech for Detection of Laryngeal Pathologies, **Proceedings of the 22nd IEEE Annual EMBS International Conference**, p. 2369-2372.

MISITI, M.; MISITI, Y.; OPPENHEIM, G.; POGGI, J. M. (1997). **Wavelet Toolbox Users' Guide for MatLab**, The Mathworks. Disponível em: <<http://www.mathworks.com>>.

MURRY, T.; ABITBOL, J.; HERSAN, R. (1999). Quantitative Assessment of Voice Quality Following Laser Surgery for Reinke's Edema. **Journal of Voice**, v. 13, n. 2, p. 257-264.

NOLL, A. M. (1966). Cepstrum Pitch Determination. **Journal of Acoustical Society of America**, v. 41, p. 293-309.

PARRAGA, A. (2002). **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. Dissertação de Mestrado, UFRS, Porto Alegre, 2002.

PEAN, V.; OUAYOUN, M.; FUGAIN, C.; MEYER, B.; CHOUARD, C. H. (2000). A Fractal Approach to Normal and Pathological Voices. **Acta Otolaryngol**, v. 120, p. 222-224.

PINTO, N. B.; TITZE, I. R. (1990). Unification of Perturbation Measures in Speech Signals. **Journal of Acoustical Society of America**, 87, 1278-89.

RIS, C.; FONTAINE, V.; LEICH, H. (1995). Speech Analysis based on Malvar wavelet transform. **Proceedings on IEEE International Conf. on Acoustics, Speech and Signal Processing**, v. 1, p. 389-392.

RODRIGUES, S.; BEHLAU, M.; PONTES, P. (1994). Proporção Harmônica-ruído: Valores para indivíduos adultos brasileiros. **Acta Awho**, v. 13, n. 3, p. 112-16.

ROSA, M. O. (1997) **Análise Acústica da Voz para Pré-diagnóstico de Patologias da Laringe**. Dissertação de Mestrado, Depto de Engenharia Elétrica, Universidade de São Paulo, São Carlos, 1997.

ROSA, M.; PEREIRA, J. C.; GRELLET M. (2000). Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis, **IEEE Trans. Biomed. Eng.**, v. 47, p. 96-102.

SCALASSARA, P. R., DAJER, M. E., MACIEL, C. D., PEREIRA, J. C., and GUIDO, R. C. (2007). Relative entropy measures applied to healthy and pathological voice characterization,” **Applied Mathematics and Computation**, Work in press, doi: 10.1016/j.amc.2007.10.068.

SCHOENTGE, J.; GUCHTENEERE, R., DE (1995). Time Series Analysis of Jitter. **Journal of Phonetics**, v. 23, p. 189-201.

SCHUCK Jr., A.; GUIMARÃES, L. V.; WISBECK, J. O. (2003). Dysphonic Voice Classification Using Wavelet Packet Transform and Artificial Neural Network. **Proceedings of the 25th Annual International Conference of the IEEE EMBS**, México.

SHOJI, K.; REGENBOGEN, E.; YU, J. D. and BLAUGRUND, S. M. (1992). High-frequency power ratio of breathy voice, **Laryngoscope**, v. 102, p. 267-271.

STOLNITZ, J. E.; DEROSE, D. T.; SALESIN, H. D. (1996). **Wavelets for computer graphics, theory and applications**. New York.

STRANG, G.; NGUYEN, T. (1996). **Wavelets and filter banks**, Wellesley-Cambridge Press.

SUMMERFIELD, M. (2004). **C++ GUI Programming with Qt 3 - Professional Technical Reference**. New York: Prentice-Hall Int.

SUYKENS, J. A. K.; VANDEWALLE, J. (1999). Least Square Support Vector Machine Classifiers, **Neural Process. Lett.**, v. 9, n. 3, p. 293-300.

SUYKENS, J. A. K.; GESTEL, T. V.; BRABANTER, J.; DE MOOR, B.; VANDEWALLE, J. (2002). **Least Squares Support Vector Machine Classifiers**, World Scientific, Singapore.

SVEC., J.G.; CHUTTE, H. K.; MILLER, D. G. (1996). A Subharmonic Vibratory PATTERN in Normal Vocal Folds. **Journal of Speech and Hearing Research**, v. 39, p.135–143.

UMAPATHY, K.; KRISHNAN, S. (2005). Feature Analysis of Pathological Speech Signals Using Local Discriminant Bases Technique, **IEE Med. Biol. Eng. Comput.**, v. 43, p. 457–464.

VAPINIK, V. (1995). **The Nature of Statistical Learning Theory**, Nova York: Springer-Verlag.

WINHOLTZ, W. S.; RAMIG, L. O. (1992). Vocal Tremor Analysis with The Vocal Demodulator. **Journal of Speech and Hearing Research**, v. 10, p. 534–543.

YATES, A.; DEDO, H. (1984). Carbon dioxide enucleation of polypoid vocal cords. **Laryngoscope**, v. 94, p. 731-735.

ZAR, J. H. (1984). **Biostatistical Analysis**, New Jersey.

ZHANG, L.; ZHOU, W.; JIAO, L. (2004). Wavelet support vector machine. **IEEE Trans. System Man Cybernet.** Part B Cybernet 34.

ZHENG, S.; LIU, L.; TIAN, J.W. (2004). A New Efficient SVM-based Edge Detection Method. **Pattern Recognition Letters**, v. 25, p.1143–1154.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)