# Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas

Silvana Luciene do Nascimento Cunha Costa

Tese de Doutorado submetida à Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências no domínio da Engenharia Elétrica.

Área de Concentração: Processamento da Informação

Benedito Guimarães Aguiar Neto - Dr.-Ing.
Orientador

Campina Grande, Paraíba, Brasil
novembro/2008

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG**

C837a
2008 Costa, Silvana Luciene do Nascimento Cunha.
Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas / Silvana Luciene do Nascimento Cunha Costa.— Campina Grande, 2008.
160 f.

Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.
Referências.
Orientador: Prof. Dr. Benedito Guimarães Aguiar Neto.

1. Processamento Digital de Sinais de Voz 2. Discriminação de Vozes Patológicas 3. Análise Acústica I. Título.

CDU – 621.391 (043)

**ANÁLISE ACÚSTICA BASEADA NO MODELO LINEAR DE PRODUÇÃO DA FALA PARA DISCRIMINAÇÃO DE VOZES PATOLÓGICAS**

**SILVANA LUCIENE DO NASCIMENTO CUNHA COSTA**

Tese Aprovada em 03.11.2008

**BENEDITO GUIMARÃES AGUIAR NETO, Dr.-Ing., UFCG**
**Orientador**

**JOSÉ CARLOS PEREIRA, Dr., USP**
**Componente da Banca**

**ABRAHAM ALCAIM, Ph.D., CETUC-PUC-RJ**
**Componente da Banca**

**JOSEANA MACEDO FECHINE, D.Sc., UFCG**
**Componente da Banca**

**MARCELO SAMPAIO DE ALENCAR, Ph.D., UFCG**
**Componente da Banca**

**CAMPINA GRANDE – PB**
**NOVEMBRO - 2008**

*Dedico este trabalho, em primeiro lugar aos meus pais, Olindina e Euclides (in memorian), ao meu querido esposo Washington e aos meus amados filhos Isabella (15), Eduardo (12), Daniel (8) e Ana Luíza (4 meses).*

"*Quanto mais as pessoas acreditam em uma coisa, quanto mais se dedicam a ela, mais podem influenciar no seu acontecimento*".

Dov Éden

# Agradecimentos

Em primeiro lugar, a Deus, pelo dom da vida, pela luz, pelas oportunidades, dificuldades, conquistas, e por ter sempre colocado pessoas em minha vida que sempre contribuem para o meu desenvolvimento.

Aos meus pais, Euclides (*in memorian*) e Olindina. Em especial, à minha mãe, que sendo pai e mãe de 13 filhos, durante tantos anos, nunca faltou com seu amor e suas palavras de incentivo.

Ao meu amado esposo, Washington, pelo seu amor, sua paciência e colaboração constantes, sempre elevando minha auto-estima e acreditando nessa conquista. Sem o seu apoio, esse trabalho não seria possível.

Aos meus queridos filhos Isabella, Eduardo e Daniel que compreenderam e souberam suportar bem minhas ausências necessárias.

À minha família, em especial à minha irmã Nilza, que tanto contribuiu para a minha formação técnica, moral e espiritual.

Ao meu orientador Benedito Guimarães Aguiar Neto, pela orientação, incentivo e contribuições valorosas, além da compreensão das minhas limitações devido à família, trabalho, etc.

À professora Joseana Macêdo Fechine, amiga, incentivadora constante, co-orientadora não oficial, que tanto contribuiu para este trabalho, sempre pronta a me atender.

Aos meus queridos amigos caravaneiros, guias e protetores, sempre me apoiando e orando pra me fortalecer.

Aos meus amigos de trabalho, principalmente da Coordenação de Telecomunicações do CEFET-PB, especialmente, à minha querida amiga Suzete, pelo carinho, incentivo e contribuições valorosas.

À amiga Daniella, companhia virtual constante, sempre dando força em todos os momentos.

Ao Dr. José Carlos da Silva, otorrinolaringologista, pelo apoio, pelas contribuições e pelo interesse na pesquisa.

A todos da Copele, prof. Benemar, Ângela, Suênia e Pedro, pelo apoio constante.

Aos professores Marcelo Sampaio de Alencar, José Carlos Pereira e Abraham Alcaim pelas valorosas contribuições.

Ao CEFET-PB, à Universidade Federal de Campina Grande e à Capes.

# Resumo

Discriminação de vozes patológicas tem sido realizada por meio de técnicas de processamento digital de sinais, como uma ferramenta auxiliar a exames videolaringoscópicos. Esse método é não-invasivo e mais confortável quando comparado a exames laringoscópicos. Este trabalho trata da análise acústica de sinais de vozes afetadas por patologias na laringe, especificamente, edema nas dobras vocais. O processo de discriminação da voz patológica e o diagnóstico da patologia considerada utilizam basicamente três etapas principais: caracterização acústica, modelagem das características e classificação. A patologia é caracterizada utilizando análise por predição linear, análise cepstral e mel-cepstral. Para a estimação dos coeficientes cepstrais é utilizada uma abordagem paramétrica derivada da análise por predição linear e para os mel-cepstrais uma abordagem não paramétrica, baseada na transformada rápida de Fourier. Cada característica acústica obtida é utilizada para o processo de modelagem paramétrica em um classificador individual, de forma a melhor avaliar sua relevância na detecção da presença da patologia. Para reduzir a quantidade de dados relacionada aos vetores de parâmetros utilizados na análise, é utilizada a técnica de Quantização Vetorial e uma medida de distorção associada para um estágio preliminar do processo de classificação. Para a classificação final e como um refinamento do processo, utiliza-se uma modelagem paramétrica, por meio de Modelos de Markov Escondidos (*Hidden Markov Models* – HMM). Os resultados mostram que os métodos desenvolvidos são eficientes em modelar os efeitos provocados pela patologia em estudo e permitir uma discriminação eficiente da patologia quando comparada a vozes normais.

Palavras-chave: Processamento digital de sinais de voz, Discriminação de vozes patológicas, Análise acústica.

# Abstract

Pathological voice discrimination has been done by means of digital signal processing techniques as a complementary tool to videolaryngoscopic exams. This method is non-invasive to patients and more comfortable when compared to laryngoscopy. This work aims at the acoustic analysis of voice signals affected by laryngeal pathologies, particularly, vocal fold edema. The discrimination process of the pathological voice and consequently the pathology detection consists, basically, in three main stages: acoustic characterization, feature modeling and classification. The pathology is characterized by linear prediction, cepstral and mel-cepstral analysis. To estimate cepstral coefficients, a parametric approach derived from linear prediction analysis is used. The mel-cepstral coefficients estimation uses a nonparametric approach based on Fast Fourier Transform. An individual classifier is applied to each acoustic feature obtained to best evaluate its relevance in detecting the pathology presence. In order to reduce the amount of data related to the parameter vectors used in the analysis, a Vector Quantization technique is applied and a distortion measurement is associated to a preliminary stage of the classification process. For the final classification a parameter modeling is carried out using Hidden Markov Models as a refinement stage of the preliminary classification process. Results show that the developed methods are efficient in modeling the effects caused by the pathology in study and provide an efficient discrimination of the pathology when compared to normal voices.

Keywords: Digital processing of speech signals, Pathological voice discrimination, Acoustic Analysis.

# SUMÁRIO

# Lista de Figuras

# Lista de Tabelas

# Lista de Siglas e Abreviaturas

AMDF - *Average Magnitude Difference Function*
BPL - *Bandpass liftering* ("Lifteragem" ou filtragem linear passa-faixa)
CA - Correta Aceitação
CEP - Coeficientes cepstrais
CEPP - Coeficientes cepstrais ponderados
CR - Correta Rejeição
DCEP - Coeficientes delta-cepstrais
DCEPP - Coeficientes delta-cepstrais ponderados
DET - *Detection-Error Tradeoff*
E - Eficiência
ERG - Excitação do Ruído Glotal
ERN - Energia de Ruído Normalizada
$F_0$ - Freqüência fundamental
FA - Falsa Aceitação
FFT - Transformada Rápida de Fourier (*Fast Fourier Transform*)
FR - Falsa Rejeição
GPS – *Glotal Pathological Shape*
ITV - Índice de turbulência vocal
HMMs - Modelos de Markov Escondidos (*Hidden Markov Models*)
LBG – Algoritmo de Linden, Buzo e Gray
LPC - *Linear Predictive Coding*
MDVP – *Multi-Dimensional Voice Program*
MEL – Coeficientes mel-cepstrais
MFCC - *Mel-frequency Cepstral Coefficients*
MMEEI - *Massachusetts Eye and Ear Infirmary*
NHR - *Noise-to-Harmonic Ratio*
OP – Outras Patologias
QPA - Quociente de Perturbação de Amplitude
QPP – Quociente de Perturbação do *Pitch*
QV - Quantização Vetorial
RHR - Relação Harmônica-ruído
ROC - *Receiver Operating Characteristic*
SE - Sensibilidade
*SP - Especificidade*

# Lista de Símbolos

$c_{mel}(n)$ - n-ésimo coeficiente mel-cepstral

$ex(n)$ - Sinal de excitação do modelo linear de produção de fala

$s_p(n)$ - Amostra pré-enfatizada do sinal $s(n)$

$c_d(n)$ - Cepstrum complexo discreto de $x(n)$

$c_i(n)$ - coeficientes cepstrais obtidos a partir dos coeficientes LPC

$\Delta c_i(n)$ - Coeficientes delta-cepstrais

$\Delta cw_i(n)$ - Coeficientes delta-cepstrais ponderados

$\alpha(k)$ - coeficientes do filtro de predição linear

$\phi$ - Constante de normalização dos coeficientes cepstrais

$\frac{\partial c_i(t)}{\partial t}$ - Derivada dos coeficientes cepstrais

$E_{seg}$ - Energia segmental do sinal de voz

$e(n)$ - erro de predição

$\alpha$ - Fator de pré-ênfase

$H_p(z)$ - Função de transferência do filtro de pré-ênfase

$O_i$ - *i*-ésimo vetor de observação

$\Delta$ - Incremento no índice do estado do HMM

$w_{bpl}(n)$ - Janela de "liftering" ou filtragem linear passa-faixas

$w_{fl}(n)$ - Janela de filtragem linear

$w_h(n)$ - janela de Hamming

$w_h(n)$ - Janela de Hamming

$\delta_t(i)$ - Maior valor de probabilidade ao longo de um único caminho

$\bar{\lambda}_l$ - Modelo HMM re-estimado

$P_l$ - Probabilidade associada ao l-ésimo locutor

$\bar{P}_l$ - Probabilidade associada ao modelo $\bar{\lambda}_l$ re-estimado

$\alpha_T(i)$ - Probabilidade de avanço do HMM

$\beta_t(i)$ - Probabilidade de retrocesso do HMM

$\tilde{s}(n)$ - sinal de estimação do preditor

$X(e^{j\omega})$ - Transformada de Fourier de $x(n)$

$\psi_t(j)$ - Trilha de caminhos ótima

$\mu_s(n)$ - Valor médio de $s(n)$

$\Pi = \pi_i$ - Vetor de probabilidade do estado inicial do HMM

$c(n)$ - Cepstrum de $x(n)$

$w_r(n)$ - Janela retangular

$\delta$ - Limiar de probabilidade

$\lambda$ - Parâmetros do HMM

$\mathbf{Q}s$ - Seqüência de estados ótima

$X(k)$ - Transformada Discreta de Fourier de $x(n)$

$\overline{b_j(k)} =$ Valor re-estimado de $b_j(k)$

$\overline{a_{ij}} =$ Valor re-estimado de $a_{ij}$

$R_{xx}(.)$ – Função de autocorrelação do sinal $x(n)$

$\{S_j\}$ - Conjunto de estados do HMM

$\mathbf{A}$ - vetor coluna de coeficientes LPC

$\boldsymbol{A} = [a_{ij}]$ - Matriz transição de estados do HMM

**B**= [*bj*(*k*)] - Matriz de função de probabilidade das observações do HMM

$c(n, t)$ - o $n$-ésimo coeficiente da predição linear no tempo $t$

$CI$ – Constante de normalização

$cw_i(n)$ – Coeficientes cepstrais ponderados

$d(n)$ – Sinal resultante da diferença entre amostras de s(n)

$d(\boldsymbol{x}, w_i)$ - Distorção entre $\boldsymbol{x}$ e $w_i$

$d(\boldsymbol{x}, w_I)$ – Distorção entre $\boldsymbol{x}$ e $w_I$

$d_p(n)$ - Excitação distorcida pela componente patológica

$E$ – Energia do erro de predição $e(n)$

$Ex(w)$ - Transformadas de Fourier da forma de onda da excitação

$F$ - Freqüência do sinal de voz

$F_0$ – Freqüência fundamental

$F_1$ - Primeiro freqüência formante

$F_2$ - Segunda freqüência formante

$F_3$ - Terceira freqüência formante

$F_a$ – Freqüência de amostragem

$F_{linear}$ - Freqüência linear (em Hz) do sinal de voz

$F_{mel}$ - Freqüência percebida (na escala mel).

$F_n$ – N-ésima freqüência formante

$G$ – fator de ganho

$G(w)$ – Transformada de Fourier do modelo do pulso glotal $g(n)$

$G(z)$ – Transformada z do modelo do pulso glotal $g(n)$

$H(w)$ - Transformada de Fourier da resposta ao impulso $h(n)$

$H(z)$ - Transformada z da resposta ao impulso $h(n)$

$h_{\text{gps}}(n)$ – resposta ao impulso do filtro de conformação glotal

$H_{GPS}(w)$ - Transformada de Fourier de $h_{\text{gps}}(n)$

$H_{TV}(w)$ – Modelo combinado entre $V(w)$ e $R(w)$ em condições saudáveis

$K$ – Dimensão do quantizador

$L$ – Quantidade de amostras de uma janela

$N$ – Número de amostras do sinal de voz

$N_A$ – Número de amostras de um segmento do sinal de voz

$Ne$ – Número de estados do HMM

$Nf$ - Número de filtros digitais

$Nq$ – Número de níveis do quantizador ou número de vetores códigos em $Wqv$

**O** - Vetor de observações

$p$ – Ordem do filtro de predição linear

$P$ – Período do sinal de voz

$P_h(n)$ – Trem de impulsos

$Q$ – Função de mapeamento $Q$ de um espaço Euclidiano $K$-dimensional

$q_i$ – i-ésimo estado inicial do HMM

$R(w)$ – Transformada de Fourier do modelo de radiação $r(n)$

$\mathrm{R(z)}$ – Transformada z do modelo de radiação $r(n)$

$R^K$ - Espaço Euclidiano $K$-dimensional

$\mathbf{R_{xx}}$ – Matriz de autocorrelação de $x(n)$

$\mathbf{r_{xx}}$ – Vetor coluna da matriz de autocorrelação

$s(n)$ – Sinal de voz

$S(z)$ – Transformada de Fourier de $s(n)$

$S_f$ - Estado final do HMM

$Sf(k)$ - Sinal de saída do banco de filtros digitais

$S_i$ – Estado inicial do HMM

$s_s(\mathrm{n})$ – Segmento sonoro do sinal de voz

$Sv_i$ – Partição do espaço Euclidiano (Células de Voronoi)

$T$ - Tamanho da seqüência do vetor de observações

$T_0$ – Período de *pitch*

$T_1$ – Instante de tempo correspondente ao inverso da freqüência fundamental máxima do sinal de voz

$T_2$ – Instante de tempo correspondente ao inverso da freqüência fundamental mínima do sinal de voz

$u(n)$ – sinal de excitação do modelo do trato vocal

$U(\mathrm{z})$ – Transformada z de $u(n)$

$V(\mathrm{w})$ – Transformada de Fourier do modelo do trato vocal $v(n)$

$V(z)$ – Transformada z do modelo do trato vocal $v(n)$

$w(n)$ – Janela de ponderação

$w_{\mathrm{i}}$ - Vetor código

$W_k(j)$ - Janelas de ponderação triangulares associadas às escalas-mel

$Wqv$ – Dicionário ou conjunto de vetores códigos

$\boldsymbol{x}$ – Vetor de entrada de valor real do quantizador

$x(n)$ – sinal de voz após janelamento

# Capítulo 1
# Introdução

A presença de patologias nas dobras vocais causa mudanças significativas em seus padrões vibratórios, afetando a qualidade da produção vocal.

Há uma grande variedade de doenças relacionadas ao trato vocal que causam modificações na voz. Algumas estão relacionadas às patologias do trato vocal, enquanto outras são provocadas por doenças neuro-degenerativas (DAVIS, 1979; QUECK et al, 2002).

Patologias da laringe, como nódulos nas dobras vocais, pólipos, cistos, carcinomas e paralisia dos nervos laríngeos, por exemplo, podem ser corrigidos por meio de: terapia vocal, cirurgia e, em alguns casos, radioterapia (MARTINEZ & RUFINER, 2000).

Atualmente, essas patologias têm aumentado drasticamente, principalmente devido a hábitos sociais não-saudáveis e ao abuso vocal.

Existem diversos procedimentos de rotina para exames da laringe com propósitos clínicos ou de investigação, que incluem vídeolaringoscopia (exame com um instrumento de fibra ótica), vídeoestroboscopia (iluminação estroboscópica da laringe, útil para visualização dos movimentos), eletromiografia (observação indireta do estado funcional da laringe) e vídeofluoroscopia (técnica radiográfica na qual o paciente ingere uma determinada quantidade de uma substância rádio-opaca para avaliar a deglutição) (MARTINEZ & RUFINER, 2000).

Várias técnicas têm sido usadas para avaliar a qualidade vocal do paciente. As técnicas mais usadas são baseadas na escuta da voz do paciente e na inspeção das dobras vocais por exames como os citados. A primeira técnica é subjetiva e, dependendo da experiência do profissional e dos métodos utilizados, pode levar a diferentes resultados. O uso do exame laringoscópico tem a vantagem de ser uma técnica objetiva mais exata, mas é considerada invasiva, causando desconforto ao paciente. Além disso, os instrumentos endoscópicos são caros e sofisticados e requerem equipamentos adicionais, tais como fontes de luz especiais e câmeras de vídeo especializadas. Essa técnica, além de dispendiosa, é ainda considerada de risco, tendo que ser executada em condições controladas por profissional especializado (MANFREDI, 2000; ESPINOSA, 2000).

## 1.1 Motivação da pesquisa

O interesse pela análise acústica na avaliação da qualidade vocal tem crescido muito nos últimos anos. A análise acústica é uma técnica não-invasiva baseada no processamento digital do sinal de voz, podendo ser empregada como uma ferramenta eficiente para o auxílio ao diagnóstico de desordens vocais, classificação de doenças da voz e, particularmente, sua pré-detecção. Além disso, essa técnica pode ser utilizada para a determinação objetiva de alterações da função vocal, avaliações de cirurgias, tratamentos farmacológicos e de reabilitação (GODINO-LLORENTE et al, 2006).

A análise acústica pode ser utilizada como técnica complementar a métodos baseados na inspeção direta das dobras vocais, podendo diminuir a regularidade dos exames invasivos.

Análise acústica não está restrita à área médica, no que se refere à detecção de patologias, podendo ser aplicada também ao controle da qualidade vocal de profissionais que trabalham com a voz, tais como cantores, locutores, professores, etc.

Vários pesquisadores têm dedicado esforços na obtenção de métodos eficientes para discriminar vozes patológicas usando análise acústica. Esses métodos utilizam, em geral, uma abordagem clássica baseada em técnicas de estimação de irregularidades de *pitch* e de amplitude, presença de componentes sub-harmônicas e distorção da envoltória do sinal. Entre as técnicas propostas estão técnicas de estimação do ruído glotal, extração de características dos parâmetros da análise tempo-freqüência, e outras baseadas no modelo de predição linear, análise cepstral e no modelo de percepção auditiva (UMAPATHY et al, 2005; GODINO-LLORENTE et al, 2006; SHAMA et al, 2007; MURPHY & AKANDE, 2007; DIBAZAR et al, 2006; BAHOURA AND PELLETIER, 2004).

Para a discriminação de vozes patológicas, por meio de análise acústica, é fundamental que o processo de caracterização acústica da patologia seja bem elaborado. Para tanto, deve-se buscar uma modelagem acústica para a patologia, utilizando características acústicas que diferenciem uma voz patológica de uma voz normal ou que seja capaz de fazer a distinção entre diversas patologias.

Entretanto, a literatura não é ainda conclusiva com relação às características acústicas ou aos parâmetros mais adequados para modelagem de

uma patologia em particular. A maioria das pesquisas baseia-se na discriminação entre vozes normais e patológicas, sem especificar a patologia. Alguns estudos focalizam uma determinada patologia sem, no entanto, apresentar um modelo acústico correspondente. A pesquisa para uma análise acústica mais representativa e detalhada de sinais de vozes patológicas é ainda um campo promissor.

A abordagem utilizada nesta pesquisa, utilizando Modelos de Markov Escondidos como refinamento do processo de classificação por meio da quantização vetorial, para discriminação de vozes patológicas, não foi utilizada em trabalhos anteriores (AGUIAR NETO, B.G. et AL, 2007a; AGUIAR NETO, B.G. et al, 2007b; COSTA, S. C. et al, 2008a; COSTA, S. C. et al, 2008b; COSTA, AGUIAR NETO and FECHINE, 2008; AGUIAR NETO, COSTA and FECHINE, 2008). Além disso, considerando-se a patologia Edema como a patologia de interesse, não há conhecimento de uma caracterização acústica elaborada pela análise dinâmica linear, utilizando os coeficientes LPC, os coeficientes cepstrais e seus derivados e mel-cepstrais em classificadores individuais. Essa metodologia permite avaliar a relevância de cada parâmetro na caracterização acústica da referida patologia.

## 1.2 Objetivo da pesquisa

No trabalho desenvolvido, técnicas de processamento digital de sinais são usadas para realizar uma análise acústica do sinal de voz patológica. O objetivo principal desta pesquisa é investigar as características acústicas significativas para uma determinada patologia da voz, de forma a desenvolver um método eficiente para discriminação entre vozes normais e vozes patológicas. O foco do trabalho está voltado para o estudo de desordens vocais provocadas por edemas nas dobras vocais. Para tanto, será analisada a importância de características baseadas no modelo de produção da fala. Será levada a efeito uma análise comparativa entre as características da voz normal e da voz patológica conforme as suas irregularidades.

Alguns trabalhos encontrados na literatura, voltados para a detecção de patologias pela análise acústica, estão focados na detecção automática de

alterações vocais por meio de análise de longa duração do sinal de voz. Os parâmetros de longa duração são geralmente calculados a partir da média das perturbações temporais da voz, a fim de avaliar o seu grau de normalidade. Entre esses parâmetros usuais estão: freqüência fundamental (correlato perceptual – *pitch)*, *jitter* (perturbação da freqüência fundamental), *shimmer* (perturbação em amplitude), quociente de perturbação de amplitude (QPA), quociente de perturbação do *pitch* (QPP), relação harmônica-ruído (RHR), energia de ruído normalizada (ERN), índice de turbulência vocal (ITV), excitação do ruído glotal (ERG), entre outros (MANFREDI, 2000; ROSA et al, 2000; ESPINOSA, 2000, ADNENE et al 2003; GARCIA et al, 2005).

Sinais de vozes patológicas apresentam, geralmente, uma característica ruidosa, apresentando dificuldades na obtenção do *pitch*. Dessa forma, medidas acústicas obtidas do *pitch*, como *jitter* e *shimmer*, por exemplo, podem não ser confiáveis para alguns sinais com características bem ruidosas. Com isso, torna-se necessária a busca por outras medidas acústicas que representem bem as mudanças no sinal de voz, provocadas pela patologia em análise.

Além disso, os métodos citados utilizam um único vetor de características obtido a partir da análise de longa duração, com o objetivo de detectar alterações vocais. Entretanto, alguns desses parâmetros são baseados na estimação exata da freqüência fundamental, uma tarefa bastante complexa na presença de certas patologias.

Algumas das propostas mais recentes, baseadas no modelo de produção da fala, não têm apresentado uma preocupação de modificação do modelo tradicional para representar uma patologia específica, limitando-se, em geral à discriminação da voz em normal ou patológica (DIBAZAR, 2002; LI and JO, 2004; UMAPATHY et al, 2005; GODINO-LLORENTE, 2006).

Observou-se que, em vários trabalhos sobre discriminação de vozes patológicas, não há uma descrição detalhada do comportamento de determinadas características dos sinais de voz e sua relação com a patologia em estudo. Neste trabalho, pretende-se observar as mudanças em características e/ou parâmetros do sinal de voz, obtidas do modelo linear de produção de fala. Isto possibilitará o uso mais adequado dessas características no processo de detecção de patologias vocais.

O presente trabalho apresenta um estudo do comportamento do modelo linear de produção de fala diante da presença da patologia edema nas dobras

vocais. Observa-se o comportamento de algumas características acústicas relevantes da fala, a fim de caracterizar acusticamente a patologia Edema. O modelo acústico obtido deve ser tal que possibilite o desenvolvimento de um sistema de reconhecimento de padrões da voz que permita uma detecção automática e eficiente da patologia.

## 1.3 Metodologia da pesquisa

O processo de discriminação da voz patológica e, conseqüentemente, da detecção da patologia considerada, é levado a efeito utilizando três etapas principais: caracterização acústica, modelagem das características e classificação.

Na caracterização acústica do sinal de voz patológico é avaliado, inicialmente, o comportamento de características clássicas tais como: freqüência fundamental (*pitch*), a estrutura dos formantes e energia a curtos intervalos de tempo. Adicionalmente, é realizada uma análise por meio de abordagem paramétrica, baseada no mecanismo humano de produção da fala, considerando o modelo de análise por predição linear, na tentativa de especificar e caracterizar as modificações paramétricas provocadas pela patologia. Além dos coeficientes de predição linear (LPC), o comportamento da patologia é caracterizado por meio de coeficientes obtidos a partir da análise cepstral e mel-cepstral. Para a estimacão desses coeficientes é utilizada uma abordagem paramétrica derivada da análise por meio da Codificação por Predição Linear (análise LPC – *Linear Predictive Coding*) e uma abordagem não paramétrica, pelos coeficientes mel-cepstrais, baseada na transformada rápida de Fourier (*Fast Fourier Transform* - FFT) (RABINER and JUANG, 1993; O'SHAUGHNESSY, 2000, GODINO-LLORENTE, 2006).

As características acústicas destacadas acima são utilizadas para o processo de modelagem paramétrica em um classificador específico, de forma a melhor avaliar e fundamentar a sua importância e o seu comportamento na patologia em estudo.

Para reduzir a dimensionalidade dos parâmetros utilizados na análise, utiliza-se a técnica de quantização vetorial com o algoritmo de Linde, Buzo and Gray - LBG (LINDE et al, 1980).

A modelagem dos parâmetros é feita com Modelos de Markov Escondidos (*Hidden Markov Models* – HMMs). O processo de detecção da presença da patologia é efetuado a partir de uma regra de decisão baseada em uma medida de distorção obtida por comparação entre um vetor de teste, representando uma determinada característica e um vetor da mesma característica pré-armazenado (padrões de referência).

São avaliados os resultados obtidos por processos de classificação, como em Fechine (2000), baseados em: 1) regra do vizinho mais próximo, utilizando a distorção do erro médio quadrático mínimo, obtida a partir da quantização vetorial; 2) a probabilidade obtida do HMM, como um refinamento do processo, quando a medida de distorção empregada não for suficiente para, com segurança, determinar a presença da patologia.

Os dados utilizados neste trabalho foram gravados pelo *Massachusetts Eye and Ear Infirmary (MMEEI) Voice and Speech Lab* (*Kay Elemetrics*, 1994). Este banco de dados de vozes desordenadas (*Disordered Voice database – Model* 4337) inclui mais de 1.400 amostras de voz gravadas em CD-ROM (vogal /a/ sustentada e 12 segundos da *Rainbow Passage*) de aproximadamente 700 locutores. Os dados foram gravados em um ambiente controlado e amostrados a taxas de 25 ou 50 kamostras/s, com resolução de 16 bits/amostra. Neste trabalho, em particular, são utilizados arquivos provenientes da emissão da vogal /a/ sustentada (KAY ELEMETRICS, 1994). Devido ao interesse em patologias nas dobras vocais, a vogal sustentada faz com que elas vibrem durante a produção do som, permitindo a observação de seu comportamento na presença da patologia. Essa base de dados foi desenvolvida com o objetivo de auxiliar a análise perceptual de vozes desordenadas para aplicações clínicas e de pesquisa e tem sido largamente utilizada em diversos trabalhos relacionados ao foco da pesquisa (ESPINOSA, 2000; DIBAZAR & NARAYANAN, 2002; MARINAKI et al, 2004; ZHANG et al, 2005; UMAPATHY et al, 2005; GODINO-LLORENTE, 2006).

A base de dados inclui amostras de pacientes com uma grande variedade de desordens da voz provocadas por causas orgânicas, neurológicas, traumáticas e psicogênicas, entre outras.

Nesta pesquisa, foram utilizados 53 arquivos de vozes normais, disponíveis na base de dados, 44 sinais de vozes afetados por edema nas dobras vocais e 23 sinais com outras patologias da laringe como nódulos, cistos, e paralisia nas dobras vocais.

Para a fase de treinamento do sistema, utilizou-se aproximadamente 50% dos sinais de vozes patológicas com a patologia edema e os demais sinais foram usados na fase de teste.

Alguns dados foram obtidos com o *software Multi-Speech – Signal Analysis Workstation*, Modelo 3700, da Kay Elemetrics, USA (KAY ELEMETRICS, 1994). A obtenção dos parâmetros foi efetuada a partir do uso de rotinas escritas em linguagem C, Borland, como também por meio de implementação em MATLAB 7.0, MathWorks.

## 1.4 Estrutura do trabalho

No presente capítulo é apresentada a contextualização da pesquisa, sua importância na avaliação da qualidade vocal e as respectivas justificativas, bem como os objetivos do trabalho e as linhas gerais da metodologia empregada. É destacado, ainda, o uso da análise acústica como ferramenta auxiliar em procedimentos de diagnóstico de desordens vocais e patologias da laringe.

No Capítulo 2 é feita uma descrição do processo de produção da voz, destacando o papel da laringe, suas patologias e características básicas.

No Capítulo 3 é destacada a importância da análise acústica no processo de detecção de patologias vocais e descreve o comportamento de características e parâmetros da voz afetada por edema nas dobras vocais, tais como freqüência fundamental, energia e formantes. São descritos, ainda, neste capítulo, os métodos utilizados para obtenção dos parâmetros para a análise por predição linear e a análise cepstral, ou seja, a obtenção dos coeficientes LPC, cepstrais, cepstrais ponderados, delta-cepstrais, delta-cepstrais ponderados, além dos coeficientes mel-cepstrais.

No Capítulo 4 é apresentada a descrição da metodologia utilizada no trabalho e uma revisão da Quantização Vetorial (QV) e do modelamento por meio dos Modelos de Markov Escondidos (HMMs).

No Capítulo 5 são mostrados e analisados os resultados obtidos no processo de classificação e no Capítulo 6 são apresentadas as considerações finais e sugestões para trabalhos futuros.

Os nomes dos sinais utilizados da base de dados e algumas características dos locutores, como gênero e faixa-etária são apresentados em quadro anexo (A).

# Capítulo 2
# A Fisiologia da Voz Humana e Patologias da Laringe

## 2.1 Introdução

Neste capítulo é feito um breve estudo da laringe e algumas de suas patologias que atingem as dobras vocais, objeto de estudo do trabalho.

Serão abordadas as seguintes patologias: nódulos vocais, cistos vocais, edema de Reinke, pólipos vocais, paralisia, câncer e laringite crônica.

O objetivo principal deste capítulo é proporcionar uma visão geral da laringe e dos seus aspectos físicos e perceptuais quando esta é afetada pelas patologias citadas.

## 2.2 A Laringe

A função de produção da voz – fonação – depende fundamentalmente da laringe (Figura 2.1). A laringe é um órgão tubular, um arcabouço esquelético membranoso, situada no plano mediano e anterior superficial do pescoço. Comunica-se inferiormente com a traquéia e superiormente com a faringe (DAJER, 2006).

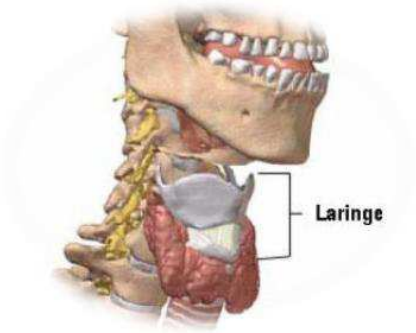



**Figura 2.1 – Representação esquemática da localização da laringe (DAJER, 2006).**

As funções básicas da laringe, em ordem de importância são proteção, respiração e fonação. Na função de proteção a laringe atua como esfíncter evitando a entrada de qualquer coisa, exceto o ar, ao pulmão. Na função de

respiração, as dobras vocais abduzem ativamente durante o movimento respiratório, contribuindo para regular a troca gasosa com o pulmão e a manutenção do equilíbrio ácido-base. Na função de fonação, as mudanças de tensão e longitude das dobras vocais, ampliação da abertura glótica e a intensidade do esforço respiratório provocam variações no tom da voz, tom que resulta da vibração das dobras vocais, modificado pelos movimentos da faringe, língua e lábios (ZITTA, 2005)

O som, que se origina na laringe como um tom fundamental, é modificado por várias câmaras de ressonância acima e abaixo desta, para, finalmente ser convertido em fala por ação da faringe, língua, palato, lábios e estruturas relacionadas (DAJER, 2006).

### 2.2.1 Dobras Vocais

As dobras vocais situam-se entre a parte interna da base das aritenóides e a tireóide (Figura 2.2). A denominação cordas, pregas ou dobras vocais se refere, na verdade, a dois pares de lábios, simetricamente formados por um músculo e um tecido elástico. As extremidades das dobras vocais estão fixadas sobre as aritenóides. Ao espaço, normalmente triangular compreendido entre as dobras vocais, dá-se o nome de glote. O afastamento e a aproximação das dobras vocais, provenientes dos movimentos das aritenóides e dos músculos que as dirigem, são responsáveis pela abertura ou fechamento total ou parcial da glote (PARRAGA, 2002).

As dobras vocais têm papel preponderante na fonação. São estruturas multi-laminadas e cada camada apresenta propriedades mecânicas diferentes (ZITTA, 2005). De um modo geral, as dobras vocais são duas dobras de músculos e mucosas que se estendem horizontalmente na laringe (Figura 2.2 e Figura 2.3). Na Figura 2.3, são ilustrados os processos de abdução (afastamento) e adução (fechamento) das dobras vocais.

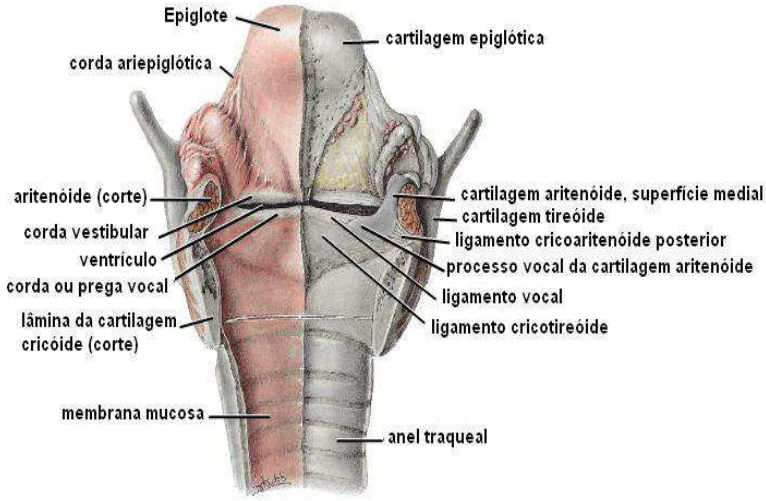
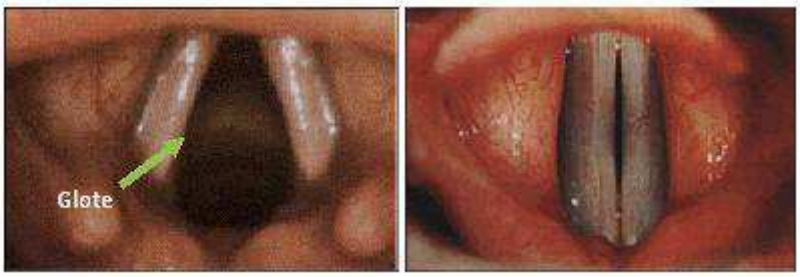



**Figura 2.2 - Vista posterior da laringe.**
Fonte: Dynamic Human Anatomy – Version 2.0 (CD-ROM, Institutional Edition)

O padrão vibratório das dobras vocais pode ser descrito e atribuído à patologia com relação aos diversos traços ou fenômenos principais: freqüência fundamental, periodicidade, movimento horizontal e vertical, onda mucosa e fechamento glótico (HIRANO, 1996; HIRANO e BLESS, 1993).




**(a)** **(b)**

**Figura 2.3 - Dobras vocais normais em: (a) abdução e (b) adução - visão endoscópica**
(Adaptado de Bouchet e Cuilleret, 1998).

A freqüência fundamental ($F_0$) corresponde à freqüência do sinal de excitação proveniente do movimento da glote, ou seja, é o número de vibrações das dobras vocais por segundo. A freqüência fundamental recebe o nome de primeiro harmônico e varia em torno de: 113 Hz para os homens, 220 Hz para as mulheres e de 240 Hz para as crianças, no português falado no Brasil (RUSSO e BEHLAU, 1993).

A freqüência fundamental é determinada por uma interação complexa entre comprimento, massa e tensão das dobras vocais, todos controlados pelos músculos intrínsecos e extrínsecos da laringe.

A freqüência fundamental varia com o padrão da vibração das dobras vocais. É importante considerar que (ZITTA, 2005):

- Quanto mais rígido for o tecido da dobra vocal, maior a freqüência fundamental;
- Quanto mais curta a porção vibrante da prega vocal, maior a freqüência fundamental;
- Quanto maior a massa da prega vocal, menor a freqüência fundamental;
- Quanto maior a pressão subglotal, maior a freqüência fundamental.

Algumas condições podem prejudicar o equilíbrio e resultar em vibrações aperiódicas ou irregulares. São elas: assimetria, interferência na homogeneidade, flacidez, tono oscilante e força inconsistente, que podem ser causadas por diferentes patologias vocais.

Os movimentos horizontais e verticais são descritos por meio da amplitude que é definida como uma extensão da excursão horizontal das dobras vocais durante a vibração. Varias condições afetam a amplitude de vibração, cada uma delas pode ser exemplificada por condições fisiológicas ou patológicas (ZITTA, 2005):

- quanto mais curta a porção vibrante, menor a amplitude;
- quanto mais rígido o tecido da dobra vocal, menor a amplitude;
- quanto maior a massa da prega vocal, menor a amplitude;
- a existência de um obstáculo diminui a amplitude;
- quanto maior a pressão subglotal, maior a amplitude.

A onda mucosa é identificada numa fase específica da vibração, sendo considerada como um traço importante da vibração. Considera-se que quanto mais rígida a mucosa, menos acentuada é a onda; quando a mucosa é parcialmente rígida, a onda interrompe o seu percurso na porção rígida. Quanto maior a pressão subglotal, mais acentuada a onda mucosa. E um fechamento glótico apertado ou frouxo resulta em uma diminuição da onda mucosa. A condição normal resulta de uma onda mucosa claramente observável.

Outro aspecto a considerar é o fechamento glótico, determinado pelo grau de aproximação das dobras vocais durante o fechamento máximo do ciclo vibratório. Pode ser completo, incompleto ou inconsistente. Um fechamento glótico incompleto durante a vibração das dobras vocais pode ser o resultado de uma série de condições, como: comprometimento da adução das dobras vocais; borda não-linear; obstáculo entre as dobras vocais; borda rígida e atividade cricotireóide dominante. O resultado desse tipo de fechamento, bem como do fechamento glótico inconsistente, gera a formação de fendas que podem ser compatíveis com a falta de coordenação funcional ou neurológica do paciente (PARRAGA, 2002).

## 2.3 Fisiologia da Voz Humana

O estudo da fisiologia do processo de produção de voz é essencial para modelamentos físico e matemático que sirvam de base para a construção de sistemas de reconhecimento, síntese e codificação de voz, bem como em sistemas de terapia de voz.

A voz humana é o resultado da ação de um conjunto de estruturas do trato vocal que formam um sistema versátil e intricado para produção de sons, cujas partes mais intimamente associadas à produção são os pulmões, a traquéia, a laringe, a faringe as cavidades nasais e a cavidade oral.

Raymond H. Stentson, pioneiro no estudo da fala, escreveu que a fala é um movimento sonoro audível (STENTSON, 1928). O movimento dos órgãos da fala - estruturas como: língua, lábios, mandíbula, véu palatino e o trato vocal - gera padrões sonoros auditivamente perceptíveis.

A voz, a fala e a audição são elementos fundamentais da linguagem. A voz é a produção de sons que o ser humano faz usando as dobras vocais. É o elemento sonoro da comunicação. Assim, apresenta características acústicas tais como: intensidade, que é a força ou volume com o qual o som é produzido; altura, que determina o quão grave ou agudo é o som, isto é, relaciona-se com a sua freqüência. Quanto maior a freqüência de um som mais agudo ele é percebido. Quanto menor a freqüência, mais grave ele é percebido; e o timbre que, definido pelos componentes harmônicos do espectro do som é o que, por exemplo, caracteriza e diferencia sons de instrumentos distintos, que vibrem na

mesma freqüência. É a detecção e análise de tais características que permitem o reconhecimento da informação contida na voz.

Segundo Read & Kent, pode-se dividir o estudo da fala em três grandes áreas: fisiológica (ou fisiologia fonética), acústica (ou fonética acústica) e perceptiva. A compreensão da fala exige o estudo de cada uma dessas áreas, relacionando-as entre si (READ & KENT, 1992).

### 2.3.1 Teoria Acústica da Produção da Fala

A fala é produzida a partir da liberação de ar dos pulmões para o trato vocal, formado basicamente por cavidades e órgãos articuladores que começa na abertura entre as dobras vocais, ou glote e termina nos lábios (Figura 2.4). O trato vocal é uma estrutura tubular pela qual passa o fluxo de ar vindo dos pulmões, que a seguir é modulado nas dobras vocais. Sua principal função é modular o espectro de freqüência da onda sonora que vem das dobras vocais e promover constrições para a geração de certos tipos de som.

Um esquema simplificado para o sistema vocal é apresentado na Figura 2.5, em que o trato vocal é excitado pelo ar expelido dos pulmões por ação de uma força muscular, e modulado pelo sistema massa-mola correspondente às dobras vocais.

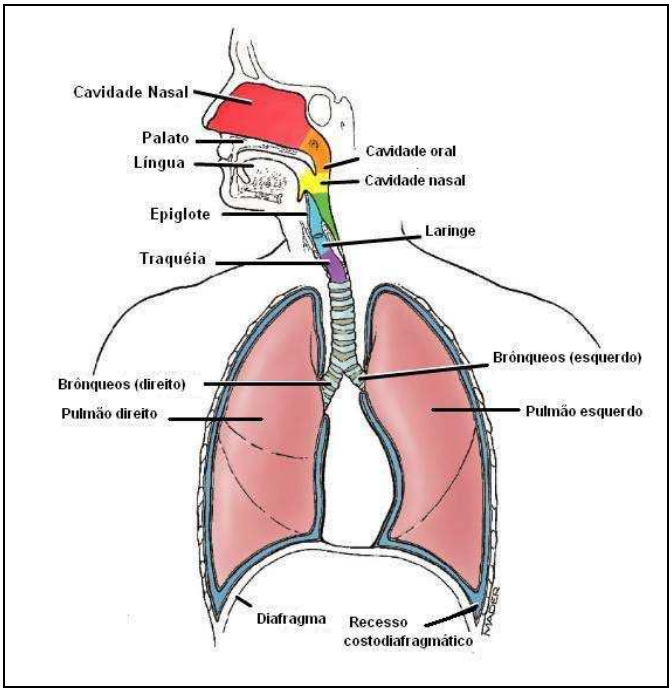



**Figura 2.4- Anatomia do aparelho fonador.**
Fonte: fttp://www.medicalexcel.com (adaptação).

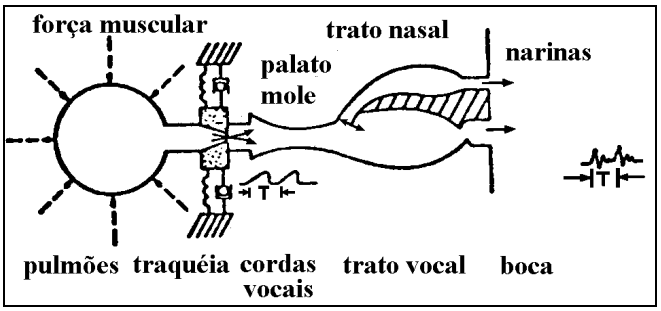



**Figura 2.5 - Modelo do trato vocal (RABINER and SCHAFER, 1978).**

O trato nasal começa na úvula e termina nas narinas. Quando a úvula é abaixada, o trato nasal é acusticamente acoplado ao trato vocal para produzir os sons nasais da voz. Na Figura 2.6, é apresentado um diagrama em blocos da produção de voz humana (DELLER, PROAKIS & HANSEN, 1993), também denominado sistema fonte-filtro, em que as dobras vocais são consideradas a fonte sonora e o trato vocal, o filtro. Nesse modelo, as saídas produzem ondas acústicas que representam a voz humana (RABINER e JUANG, 1993).

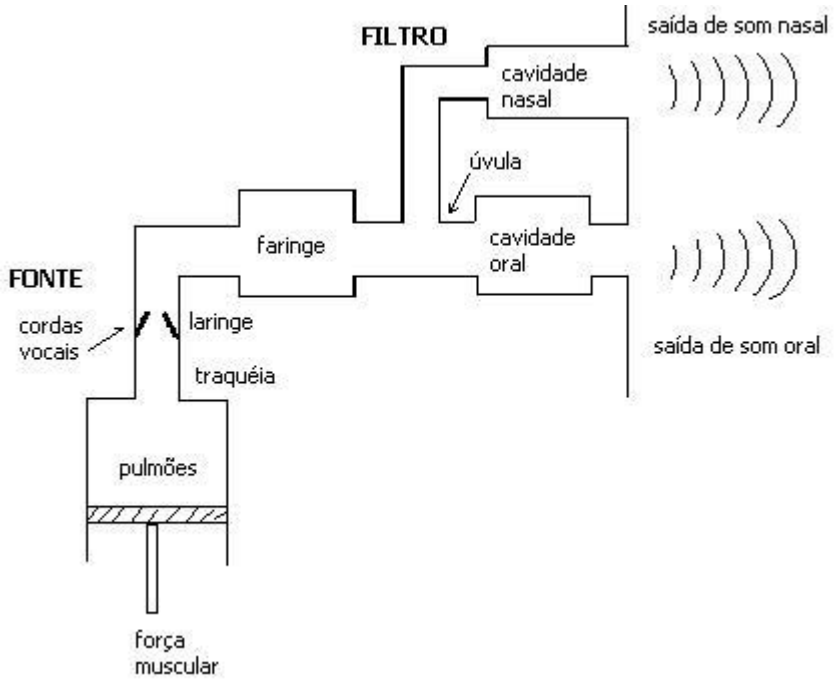



**Figura 2.6 - Um diagrama de blocos da produção de voz humana.**

O ar é conduzido para fora dos pulmões pela traquéia, passando pela laringe, onde estão as dobras vocais. O espaço compreendido entre as dobras vocais é chamado de glote, e sua abertura pode ser controlada movimentando-se as cartilagens aritenóide e tireóide. É lá que o fluxo contínuo de ar dos pulmões é geralmente transformado em vibrações rápidas e audíveis durante a fala. Isso é feito pelo fechamento das dobras vocais, que causa um aumento gradativo da pressão atrás delas, que acaba por fazer com que elas se abram repentinamente, liberando a pressão, para então tornarem a se fechar. Esse processo produz uma seqüência de pulsos cuja freqüência é controlada pela pressão do ar e pela tensão e comprimento das dobras vocais (freqüência fundamental). Os sons assim produzidos são chamados de vozeados, ou sonoros, que normalmente incluem as vogais; caso contrário, são chamados não-vozeados, ou surdos.

Além da vibração das dobras vocais, o fluxo de ar pode tornar-se audível de duas outras maneiras. O fluxo pode ser constrito em algum ponto do trato vocal, por exemplo, elevando-se a língua em direção ao palato, tornando-se turbulento e produzindo um ruído de amplo espectro. Os sons assim formados são chamados de fricativos, normalmente presentes em consoantes como [s] e [f]. Outro método consiste em interromper totalmente o fluxo de ar em algum ponto do trato, e então liberar de uma só vez a pressão formada. Os sons assim produzidos são chamados de plosivos ou explosivos, presentes em consoantes como [p] e [t].

Esses dois últimos tipos são independentes do primeiro, isto é, sons fricativos ou plosivos também podem ser surdos ou sonoros. Ex.: [f] (fricativo surdo) e [v] (fricativo sonoro).

Um som pode ser simultaneamente sonoro e surdo (misto). Por exemplo, considere o som correspondente à letra “z” (símbolo fonético /z/) na frase “três zebras”. Alguns sons da voz são formados por um período de curto de silêncio, seguido por outro de voz sonora, voz surda, ou ambas (DELLER, PROAKIS & HANSEN, 1993).

Os sons fricativos sonoros, como /j/, /v/ e /z/, são produzidos combinando vibração das dobras vocais e excitação turbulenta. Nos períodos em que a região glótica atinge um máximo, o escoamento através da obstrução torna-se turbulento, gerando o caráter fricativo do som; quando a pressão glótica cai abaixo de um determinado valor, termina o escoamento turbulento do ar e as ondas de pressão apresentam comportamento mais suave.

A análise acústica é essencial no processo de compreensão da fisiologia fonética, como área da lingüística que estuda a geração e a estrutura sonora dos fonemas. Assim, o estudo da voz humana requer a definição de conceitos ou propriedades dos sons produzidos que identificam as estruturas sonoras: harmônicos, ressonância e formantes.

Todo som complexo pode ser decomposto em uma combinação de sons mais simples, harmonicamente relacionados, ou seja, em uma série de tons puros, semelhantes ao de um diapasão, e com freqüências que são múltiplos inteiros de uma freqüência fundamental. Quando se decompõe um determinado som em seus diversos componentes, realiza-se uma análise espectral. Cada tom puro corresponde fisicamente a um tipo de oscilação - movimento harmônico simples.

A ressonância é o fenômeno segundo o qual um sistema físico, excitado por outro sistema vibrante, passa a oscilar de forma semelhante a este. No aparelho fonador humano, o trato vocal pode ser visto como uma seqüência de pequenos tubos cilíndricos que formam ressoadores.

O trato vocal supraglótico, acima, portanto, das dobras vocais, se inicia no nível da laringe, prolongando-se até a última fronteira dos lábios e da narina. Essa tubulação de diâmetro variável funciona com uma cadeia de ressoadores, respondendo, seletivamente, às diversas freqüências contidas no som produzido pela fonte sonora.

Assim, se o trato vocal em uma determinada forma responde simpática e naturalmente a determinados sons, como por exemplo, aos de freqüências próximas a 330, 800 e 2200 Hz, pode-se afirmar que estes são os primeiros formantes daquela configuração vocal (FUKS & SUNDBERG, 1999). Modificando os formantes do trato vocal, por meio de alterações em sua forma, pode-se esculpir o som básico gerado pela glote, em uma rica paleta de timbres sonoros, mensuráveis e comparáveis.

## 2.4 Patologias da Laringe

A voz é uma das expressões mais fortes da personalidade humana. Uma voz é considerada normal quando não há esforço na sua produção. Quando a harmonia não é mantida, obtém-se um som de má qualidade para os ouvintes e

emitido com dificuldade e desconforto para o falante. Fatores ambientais e físicos podem interferir no padrão adequado da qualidade vocal (BOONE, 1996).

Desordens vocais são resultantes das alterações morfológicas nas estruturas do trato vocal ou de seu mau funcionamento. Enquanto a audição é essencialmente uma função sensório-neural, a voz depende fundamentalmente da atividade de todos os músculos que participam de sua produção, além da integridade de todos os tecidos do aparelho fonador. Quando essa harmonia é mantida, o som é emitido sem dificuldade pelo falante e apresenta uma boa qualidade para os ouvintes. Quando a voz está fora dos padrões de normalidade, diz-se que está perturbada ou disfônica. A disfonia representa, então, um termo mais geral para qualquer alteração na vocalização normal (DANIEL et al, 1994).

A presença da patologia pode ser percebida por meio de sintomas relatados por pacientes aos seus médicos como queixa de sensações associadas à fonação ou dores na região da garganta. Alguns sintomas podem ser verificados, outros não. Outros sintomas podem referir-se às características perceptuais da voz, como a rouquidão, garganta arranhando ou tremor na voz (COLTON e CASPER, 1996).

Sinais de distúrbios vocais são características da voz que podem ser observadas ou testadas. Os sinais representam um inventário de características vocais embasadas em exames, observações e medições e podem ser: a) perceptuais; b) acústicos e c) fisiológicos.

Os sinais perceptuais de distúrbios vocais são as características da voz de um indivíduo que são percebidas pelo ouvinte/observador. Entre os principais sinais, destacam-se: perturbações ou variações na freqüência, na intensidade, qualidade vocal alterada, rouquidão, soprosidade, tensão, tremor, pigarro, afonia, entre outros (PARRAGA, 2002).

Quanto aos sinais acústicos, é importante lembrar que a voz é produzida por movimentos das dobras vocais interrompendo o fluxo de ar regressivo. Os movimentos das dobras são controlados pelas características biomecânicas das próprias dobras vocais, pela magnitude da pressão de ar abaixo das dobras e por seu controle neural. A patologia pode afetar tais movimentos interferindo em quaisquer dessas variáveis (PARRAGA, 2002).

O movimento das dobras vocais resulta em interrupção periódica do fluxo de ar em velocidade apropriada à percepção do som. A acústica é o estudo do som e a acústica vocal pode fornecer informações importantes referentes ao

movimento das dobras vocais. Há uma forte correspondência entre a fisiologia e a acústica e muito pode ser inferido sobre a fisiologia com base em análise acústica. Além disso, os parâmetros acústicos são provavelmente os mais fáceis de registrar e analisar objetivamente. Por isso, a análise acústica pode ser utilizada como uma ferramenta computacional de baixo custo e de forma não invasiva no monitoramento da qualidade vocal.

Há muitos sinais acústicos que podem ser associados a qualquer patologia, entre os quais estão o *jitter* e o *shimmer*. O *jitter* (perturbação de freqüência) é um dos índices que reflete anomalias das dobras vocais e pode ser de fácil medição (PARRAGA, 2002).

A freqüência fundamental da fala e suas variações, a extensão fonatória, as perturbações na amplitude, ruído espectral e o tempo de fonação, são alguns dos sinais acústicos que podem ser medidos e a partir dos quais é possível avaliar a qualidade vocal e a presença de patologia.

A presença de patologia pode afetar a freqüência fundamental (*pitch)*, fazendo com que homens e mulheres produzam uma freqüência excessivamente elevada ou baixa, respectivamente, ou seja, fora dos valores médios usuais para o sexo.

Os falantes normais apresentam uma pequena perturbação, que pode representar uma variação de massa, tensão, atividade muscular ou atividade neural das dobras vocais. Do mesmo modo que para a freqüência fundamental, é possível que a amplitude do som da prega vocal varie de um ciclo para o seguinte. Essa característica é denominada perturbação de amplitude ou *shimmer* (PARRAGA, 2002).

O tempo de fonação pode estar alterado. Um falante com uma patologia na laringe, pode não conseguir manter, por exemplo, uma vogal sustentada, durante o mesmo período de tempo que uma pessoa com a laringe em condições normais.

Quanto aos sinais fisiológicos que podem ser afetados por patologia, incluem-se as características aerodinâmicas (fluxo de ar e pressão aumentados ou reduzidos), o comportamento vibratório (área de contato, forma de onda) e a atividade muscular.

Alterações na voz podem não estar associadas a uma patologia orgânica ou física, sendo denominadas de disfonia funcional. Algumas das vozes mais ásperas podem não ter relação com uma patologia orgânica. Por outro lado, um

problema orgânico sério, como câncer em fase inicial pode não produzir qualquer alteração na voz. Na disfonia funcional, a qualidade da voz encontra-se alterada podendo apresentar-se rouca, áspera, soprosa ou estridente.

O que é uma voz normal? Existe grande variabilidade, dependendo da personalidade do indivíduo, da idade, do sexo, da inteligência, da ocupação e do estado emocional. Pessoas com incapacidade moderada ou leve podem estar satisfeitas com sua voz, enquanto outras podem achar sua disfonia um embaraço ou um impedimento para exercer suas atividades, e procurar ajuda médica.

A natureza da mudança da voz, por vezes, sugere a presença da patologia. A duração da anormalidade da voz pode ser de apenas alguns dias, com a laringite que sucede a uma infecção do trato respiratório superior, ou pode ser de muitos anos, em que a rouquidão é causada pelo edema de Reinke.

A produção de voz variável e irregular, quando as palavras são difíceis de compreender, pode ser a primeira manifestação de um distúrbio neurológico. Alguns sintomas laríngeos podem ser a característica inicial de uma doença sistêmica generalizada ou de condições neurológicas. Como exemplos de doenças por fatores neurológicos, que causam distúrbios na voz, podem ser citadas: doença cerebrovascular, paralisia cerebral, paresia ou paralisia da prega vocal, refluxo gastroesofágico decorrente de incoordenação neuromuscular, parkinsonismo, tremor essencial, disfonia espamódica e esclerose múltipla.

Quanto às doenças orgânicas que causam distúrbios na voz estão: laringite crônica, pólipo de prega vocal, nódulo vocal, edemas de Reinke, granuloma vocal idiopático, paquidermia de contato, hemorragia de prega vocal, trauma laríngeo por intubação prolongada, trauma laríngeo externo, papilomas respiratórios múltiplos, doença laríngea cística, hiperqueratose, displasia, carcinoma ou outra malignidade laríngea, anormalidades congênitas, artrite reumática, entre outras (BENJAMIN, 2000).

Entre as patologias da laringe, que afetam particularmente as dobras vocais, serão aqui destacadas: nódulos vocais, edemas de Reinke, pólipos vocais, cistos, laringite, câncer e paralisia.

### 2.4.1 Nódulos Vocais

Nódulos vocais são protuberâncias bilaterais crônicas na extremidade livre e superfície inferior das dobras vocais membranosas (Figura 2.7). Elas interferem

na produção da voz. Esta definição exclui as protuberâncias agudas na mucosa (KUHL, 1982; BENJAMIN, 2000).

Os nódulos vocais são comuns em crianças em idade escolar. Constituem a causa de distúrbios da voz em 80% das crianças, sendo mais comuns em meninos, especialmente em crianças agitadas, que cometem abusos vocais como gritos e fala excessiva. Crianças com fenda palatina reparada, por vezes, desenvolvem nódulos vocais.

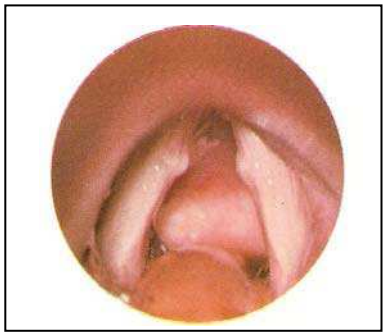


**Figura 2.7 - Nódulos Vocais**
Fonte: (KUHL, 1982).

Os nódulos vocais são as lesões benignas mais comuns das dobras vocais, tanto em crianças como em adultos. Nos adultos, são mais comuns em mulheres do que em homens, incluindo entre estes, cantores, locutores, operadores de telefonia ou telemarketing, professores e outros profissionais que usam a voz em demasia.

Os nódulos interferem na vibração normal e provocam rouquidão, soprosidade e perda do alcance da freqüência, especialmente nas freqüências mais altas.

A avaliação perceptual subjetiva da qualidade da voz é difícil, incerta, não sendo passível de comparação de um examinador para outro e requer experiência extensa. Não obstante, a rouquidão no adulto ou na criança é o indício mais importante isolado da presença de nódulos vocais.

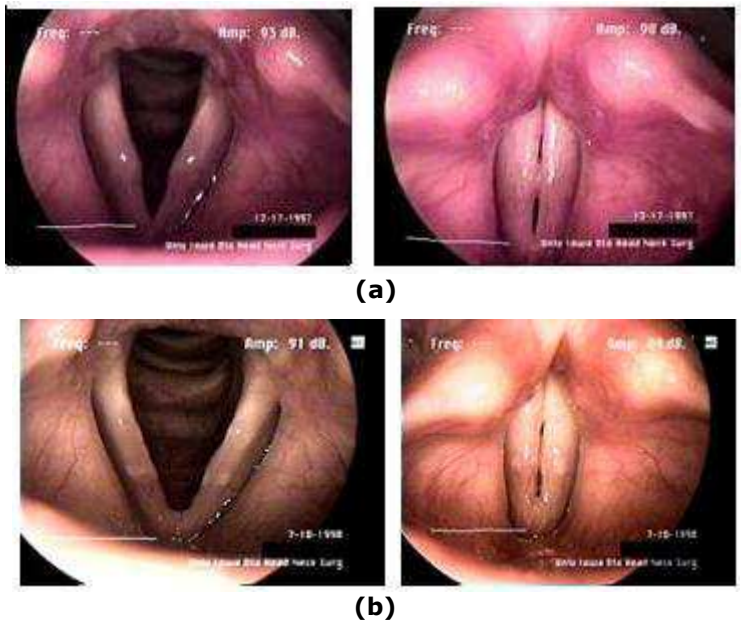
O grau de anormalidade da voz e do aspecto laríngeo na fala e no canto pode ser melhor documentado com o uso de registros audiométricos e em vídeo ou videoestroboscopia.

Os nódulos são sempre bilaterais. Podem variar em tamanho, simetria e coloração. Podem existir nódulos bilaterais, de modo que um deles seja grande, de tal modo que o outro pareça insignificante.

A fonoterapia estimula a redução do nódulo vocal. Quando a fonoterapia e o repouso vocal falham na melhoria do nódulo, a remoção cirúrgica deve ser considerada. Em geral, a cicatrização pós-cirúrgica é rápida, sem complicações e a voz retorna à normalidade algumas semanas após a remoção.

A massa aumentada dos nódulos nas dobras vocais contribui para uma altura de voz mais grave e maior periodicidade (julgada como rouquidão). Isso leva a um tipo de voz soprosa, monótona, que, muitas vezes, parece carecer de ressonância apropriada. As características acústicas do nódulo apresentam *jitter* e *shimmer* elevados. Outro fato são as evidências de ruído no espectro, de cuja intensidade depende a severidade da rouquidão e o tamanho da lesão (PARRAGA, 2002).

Na Figura 2.8, é apresentado o caso de uma paciente de 36 anos, não fumante, mas com histórico de abuso vocal.



**(a)**

**(b)**

**Figura 2.8- Dobras vocais com nódulos: (a) adução e abdução pré-operatório e (b) abdução e adução, pós-operatório.**

Na Figura 2.8(a) e 2.8(b), são apresentadas as dobras vocais com os nódulos (antes da cirurgia) e sem os nódulos (após a cirurgia) [1]. Pode-se observar na Figura 2.8(a) o fechamento incompleto da glote, comparado ao fechamento após a cirurgia.

[1] http://www.medicine.uiowa.edu/otolaryngology/cases/nodules/nodules1.htm

### 2.4.2 Edema de Reinke

Edema de Reinke é o edema bilateral da camada subepitelial no espaço de Reinke, isto é, na região onde o epitélio cilíndrico das dobras vocais se transforma em plano extratificado (KUHL, 1982).

O edema de Reinke caracteriza-se pela expansão, aumento e inchaço das dobras vocais e pelo acúmulo de líquido ou material gelatinoso (ou ainda semi-sólido) na camada superficial da lâmina própria (espaço de Reinke) das dobras vocais (Figura 2.9) (KUHL, 1982). As dobras vocais ficam enormes com um edema claro-hialino, pobre em vasos e que modifica completamente o aspecto anatômico da região glótica (KUHL, 1982; PARRAGA, 2002).

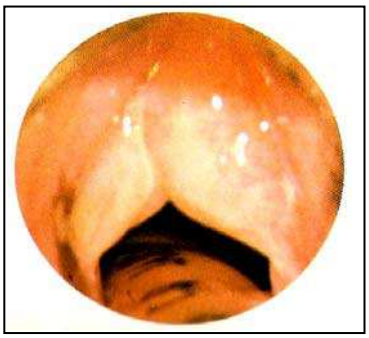


**Figura 2.9– Edema de Reinke.**

É também conhecido como cordite polipóide, degeneração polipóide e polipose difusa bilateral. Reinke foi o primeiro anatomista que descreveu a estrutura fina das dobras vocais, inclusive o espaço – que corresponde à camada superficial da lâmina própria – que leva o seu nome (BENJAMIN, 2002).

Edemas de Reinke são também conceituados como edemas generalizados e bilaterais, ou lesão inflamatória que traduzem a presença de uma laringite crônica (PAPARELLA & SHUMRICK, 1982; BELHAU & PONTES, 1995; BENTO & MINITI, 1997).

O edema é normalmente bilateral e simétrico, mas, quando um lado está mais aumentado do que o outro, uma pesquisa para patologia primária próxima deve ser empreendida. Com o crescimento progressivo do edema, pode haver uma obstrução completa da glote, com conseqüente asfixia, num estágio mais avançado (BENJAMIN, 2002).

A incidência é maior nas mulheres que atingiram a menopausa. Entretanto, encontra-se nos dois sexos e é uma doença de idade adulta. A causa irritativa é o abuso vocal e, especialmente, o fumo, principalmente em mulheres entre 50 e 70 anos de idade. Pode, ocasionalmente, ser encontrada em crianças, se já forem fumantes habituais (BENJAMIN, 2002).

A etiologia do edema de Reinke parece infecciosa, irritativa e hormonal. Alguns doentes sofrem de renites, amidalites e sinusites, que devem agir como causa infecciosa (KUHL, 1982).

Exame ultra-estrutural, em vez da microscopia ótica, contribui para a diferenciação do edema de Reinke do pólipo e dos nódulos de dobras vocais. A membrana basal mostra-se espessada, existem lagos edematosos, sinais de sangramento, espessura aumentada da parede vascular e alguma fibrina, mas há menos lesão de célula endotelial capilar que nos pólipos. O edema de Reinke é limitado às dobras vocais e atinge até a região aritenóidea, onde alcança maior espessura.

Quando o edema de Reinke é severo (Figura 2.10), as grandes bolsas de líquido podem oscilar para dentro e para fora, produzindo uma qualidade rouca, obstrutiva e roncada durante o sono (BENJAMIN, 2002). No edema de Reinke de longa data, alguns pacientes sofrem de disfonia que, diferente de qualquer tumor benigno de dobras vocais, é de um timbre mais baixo, como produzido por dobras vocais mais flácidas.

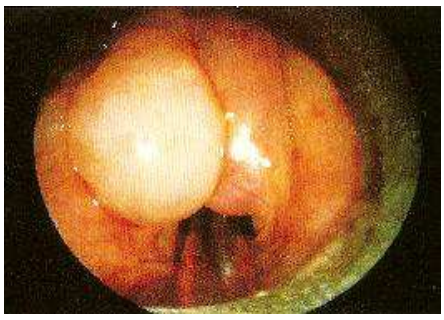
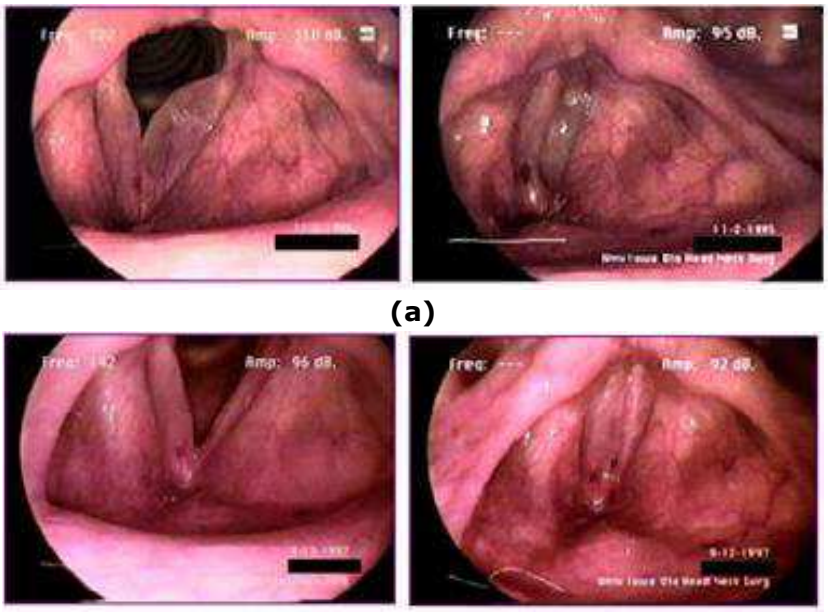


**Figura 2.10 – Edema de Reinke severo.**
(Fonte: BENJAMIN, 2000).

A história típica é de rouquidão consistente, lentamente progressiva em um fumante que fala muito. A voz é rouca e baixa, de modo que a paciente do gênero feminino desenvolve uma voz masculinizada e pode ser confundida por uma pessoa do sexo masculino ao falar ao telefone, por exemplo.

O tratamento do edema de Reinke, a partir de métodos não cirúrgicos, pode ser iniciado pela interrupção do fumo. A fonoterapia pode ajudar na limitação do uso excessivo da voz. Pode-se esperar apenas uma melhora discreta na qualidade da voz. A revisão regular e a laringoscopia indireta normalmente confirmam a lenta progressão da afecção.

Na Figura 2.11, é apresentado o aspecto das dobras vocais em um paciente masculino de 68 anos, fumante de 10-15 cigarros por dia, durante 15 anos, submetido a uma cirurgia. O paciente, além do tabagismo, tinha hábitos não-saudáveis como ingestão demasiada de cafeína e pouca ingestão de água. Na Figura 2.11, são apresentados os processos de abdução e adução das dobras vocais antes da cirurgia (Figura 2.11(a)) e após a cirurgia (Figura 2.11(b)).



**(a)**

**(b)**

**Figura 2.11 – Edema de Reinke: (a) abdução e adução - pré-operatório; (b) abdução e adução - pós-operatório.**
(FONTE: http://www.medicine.uiowa.edu/otolaryngology/cases/reinke/reinke1.htm)

O paciente apresentava uma faixa de *pitch* extremamente baixa[2]. As imagens obtidas por vídeoestroboscopia mostram movimentos muito assimétricos das dobras vocais na fonação. A dobra direita tem uma aparência relativamente normal, com alguma onda mucosa. A prega esquerda é completamente edematosa e tem uma aparência polipóide. O fechamento é incompleto durante o ciclo de fonação.

[2] http://www.medicine.uiowa.edu/otolaryngology/cases/reinke/reinke1.htm

### 2.4.3 Pólipos vocais

Os pólipos ocorrem, normalmente, na parte anterior ou média da prega vocal membranosa e constituem a patologia laríngea que mais comumente exige remoção cirúrgica (Figura 2.12). Em geral, os pólipos são discutidos com o edema de Reinke, pois ambos resultam da permeabilidade aumentada dos vasos, gerando o edema. Porém, enquanto o edema de Reinke afeta a extensão total de ambas as dobras vocais, os pólipos vocais são mais localizados e, usualmente, são unilaterais em 80% dos casos. Acontecem na borda livre ou na superfície inferior da prega vocal no terço anterior ou médio; 20% são bilaterais ou múltiplos (BENJAMIN, 2000).

Os pólipos são duas vezes mais comuns em homens do que em mulheres. São encontrados em adultos de todas as idades, com a maioria dos pacientes apresentando entre 20 e 60 anos de idade.

O uso excessivo da voz e o trabalho em ambiente ruidoso são considerados fatores agravantes. O consumo de tabaco não tem associação com o desenvolvimento de pólipos. A causa é incerta e as teorias relacionadas à patogênese não são convincentes. Um pólipo hemorrágico de prega vocal unilateral pode começar com a ruptura de um capilar no espaço de Reinke, como extravasamento de sangue seguido por organização e formação do pólipo.

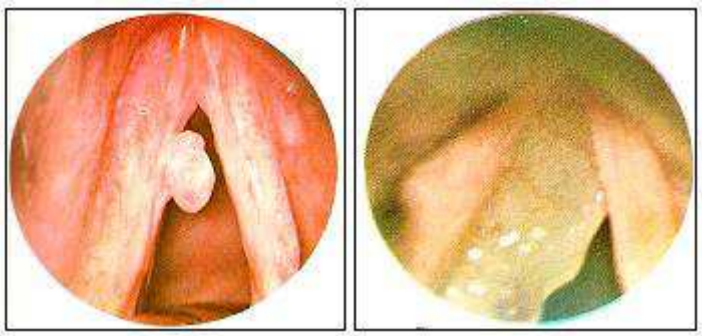


**Figura 2.12 - Pólipos nas dobras vocais: (a) pólipo fibroso e (b) pólipo gelatinoso.**
(Fonte: KUHL, 1982)

Os pólipos e os nódulos são ambos relacionados ao hiperfuncionamento vocal e apresentam algumas semelhanças físicas. Eles têm a mesma etiologia e diferem apenas em grau. Um pólipo é maior e mais vascularizado, edematoso e inflamatório do que um nódulo. O tamanho, o aspecto, a coloração e a

consistência variam consideravelmente. Os pólipos podem ser arredondados, alongados, irregulares ou multilobulados. Usualmente, são pálidos, translúcidos e edematosos e existe, ainda, um tipo hemorrágico, avermelhado e angiomatoso.

As vozes de pacientes com pólipos unilaterais caracterizam-se por disfonia severa. A lesão provocada pelo nódulo altera a vibração das dobras vocais, resultando em rouquidão e soprosidade, requerendo uma limpeza constante da garganta. As características acústicas são semelhantes às dos nódulos (*jittter*, *shimmer* e aumento do ruído espectral) (PARRAGA, 2002).

A rouquidão é, com freqüência, insidiosa, lentamente progressiva, normalmente constante e acompanhada por uma intensidade baixa de voz e uma gama de freqüência reduzida. Algumas vezes, a voz tem uma qualidade ofegante. O grau de incapacidade vocal varia com o tamanho, o local onde está situado o pólipo, a natureza e a pediculação do pólipo, embora, por vezes, a interferência com a fonação possa ser mínima. Ocasionalmente, um grande pólipo provoca uma tosse irritativa e os pólipos muito grandes ou múltiplos podem contribuir para a obstrução parcial da via aérea.

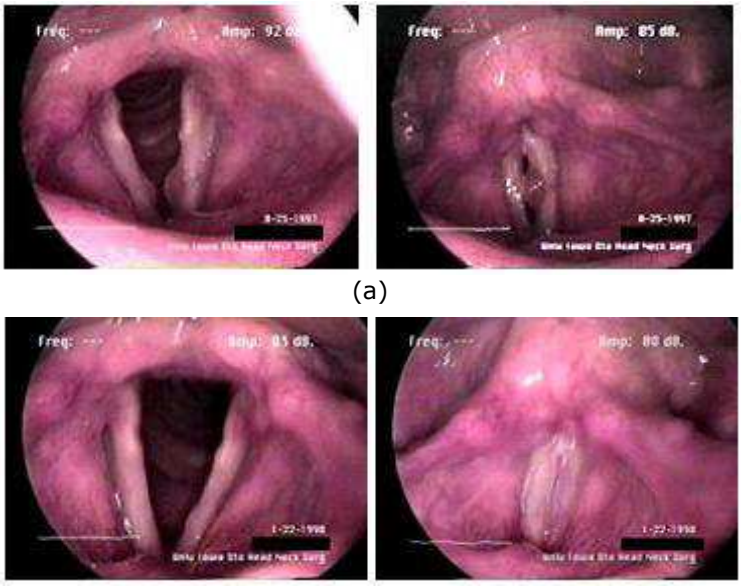
Na Figura 2.13 são mostradas as dobras vocais em adução e abdução, antes e após a retirada dos pólipos com tratamento cirúrgico. Nota-se o fechamento incompleto das dobras vocais antes da cirurgia (Figura 2.13(a)) [3].

Acredita-se que as causas dos cistos sejam, na sua grande maioria, provenientes de processos irritativos e de infecções crônicas da laringe. Os cistos se localizam com mais freqüência na epiglote, depois nas dobras vocais, bandas ventriculares, aritenóides, prega ariepiglótica e seios piriformes (KUHL, 1982).

Os cistos de retenção têm origem glandular. São cistos formados às custas de dilatação dos ácinos e canais glandulares e recoberto sempre por um epitélio glandular. O seu interior apresenta um líquido espesso e de coloração cinza.

Os cistos de tecido linfóide têm a inserção na epiglote e são de diversos tamanhos. São cobertos por um epitélio plano em meio de um folículo linfático com grande quantidade de líquido cinza ou amarelado. Apresentam uma etiologia inflamatória (KUHL, 1982).

[3] http://www.medicine.uiowa.edu/otolaryngology/cases/polyp/polyp1.htm



(a)

(b)

**Figura 2.13 - Dobras vocais com pólipos bilaterais: (a) adução e (b) abdução, pré-operatório; (c) abdução e (d) adução, pós-operatório.**

## 2.4.4 Cistos

Cistos são tumores constituídos por secreções amareladas, recobertas por um epitélio claro e transparente. Os cistos de tamanho muito variável se localizam em várias regiões da laringe (BENJAMIN, 2000).

Os cistos podem ser classificados, de acordo com a sua constituição em: cistos epidermóides, quando revestidos de epitélio plano; cistos de retenção, quando revestidos de epitélio vibrátil do conduto escretor de glândulas mucosas e cistos linfócitos ou de tecido linfóide (Figura 2.14).

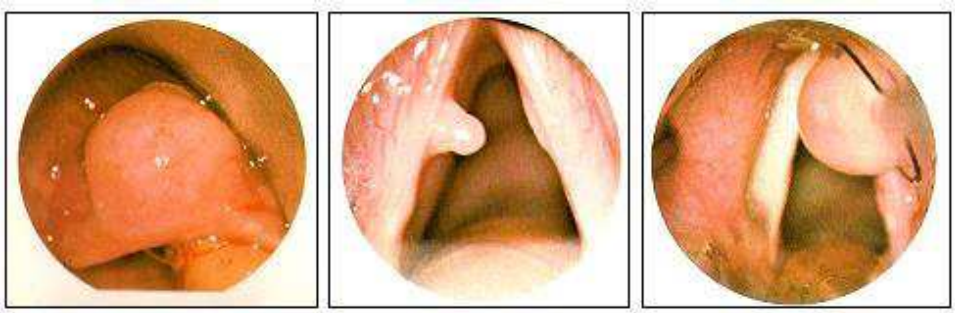
Os cistos intracordais epidermóides provocam disfonia bastante severa. Ressecado o cisto com sua cápsula, a voz volta ao estado normal.

Os cistos linfáticos produzem a sensação da presença de um corpo estranho na garganta e, algumas vezes, disfagias [4].

Do ponto de vista fisiológico, a lesão cística epidermóide provoca enrijecimento da lâmina própria da mucosa. O fechamento glótico se apresenta completo ou com fenda, dependendo das dimensões do cisto. A voz do paciente com cisto apresenta *pitch* rebaixado, dificuldade para regular a intensidade,

[4] Dificuldade de deglutição, de causa otorrinolaringológica, digestiva ou neurológica (http://www.unioeste.br/huop/fonoaudi.htm)

tensão, aspereza, soprosidade e instabilidade vocal mediante demanda vocal (PASSEROTI, em http://www.otorrinousp.org.br).



(a) (b) (c)

**Figura 2.14 - Cistos nas dobras vocais: (a) Cisto de epiglote ou linfócito; (b) Cisto epidermóide e (c) Cisto de retenção na falsa dobra direita, sendo abraçado com alça fria.** (Fonte: KUHL, 1982)

Na Figura 2.15 são apresentadas a adução e abdução das dobras vocais, com a presença de um cisto ventricular[5].

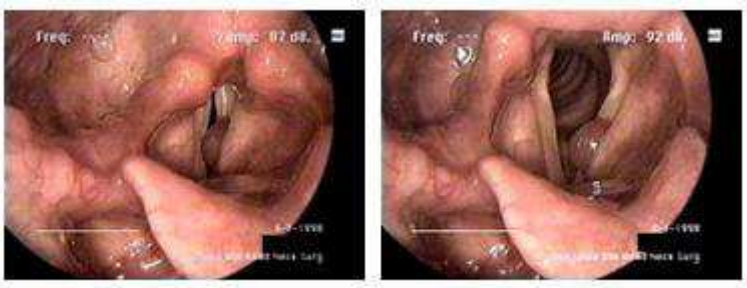


**Figura 2.15 - Adução e abdução em dobras vocais com cistos ventriculares, pré-operatório.**

## 2.4.5 Laringites crônicas

A laringite crônica é uma afecção laríngea com gênese inflamatória e irritativa. A incidência é grande, principalmente em indivíduos de sexo masculino que sejam tabagistas, usem álcool e que abusem do uso da voz. O álcool tem sua ação irritativa mais na região supraglótica, enquanto o cigarro tem a sua ação irritativa em toda a laringe, especialmente sobre o bordo e face superior das dobras vocais. Fatores ambientais (poluição e poeira) também podem

[5] http://www.medicine.uiowa.edu/otolaryngology/cases/ventricu/ventric1.htm

contribuir para a sua incidência, que tem aumentado no sexo feminino entre professoras e fumantes (KUHL, 1982; PARRAGA, 2002).

Na laringite crônica, são esperadas perturbações na freqüência e na amplitude do sinal de voz acima do normal, como também um aumento no ruído espectral (PARRAGA, 2002).

A laringite também pode decorrer de infecções respiratórias superiores que exercem um efeito generalizado sobre a mucosa do trato respiratório, incluindo a laringe. O grande sintoma é a disfonia ou rouquidão. As modificações da voz ocorrem devido à necessidade de maior esforço para movimentar uma dobra vocal de bordo irregular e de maior massa uni ou bilateral. O doente sente algo entre as dobras vocais e procura eliminar o obstáculo com a tosse, não consegue, e a disfonia persiste.

O doente com laringite crônica tem uma voz melhor pela manhã que se agrava durante o dia, pela tosse e uso da voz. Esses doentes normalmente apresentam uma grande disfonia e grande dificuldade de fonação à noite.

As dispnéias são raras, a não ser quando um tumor inflamatório cresce ou se multiplica. A dor na laringe também é rara, a não ser quando o processo inflamatório produz miosites ou artrite em suas articulações.

O pólipo, o nódulo e o edema de Reinke podem ser considerados, por alguns médicos, como lesões crônicas da laringe.

O doente portador de uma laringite crônica deve ser constantemente observado para detectar qualquer lesão que possa levar ao câncer de laringe.

### 2.4.6 Câncer

Câncer de laringe (Figura 2.16) ou carcinoma é um processo mórbido em que determinadas células, que constituem a mucosa da laringe, mostram subitamente um poder de invasão do tecido normal. A qualidade da invasão é que constitui a malignidade que apresenta recorrência quando removida.

Nas Figuras 2.16 e 2.17 (KUHL, 1982), são mostradas imagens referentes às dobras vocais de pacientes com câncer.

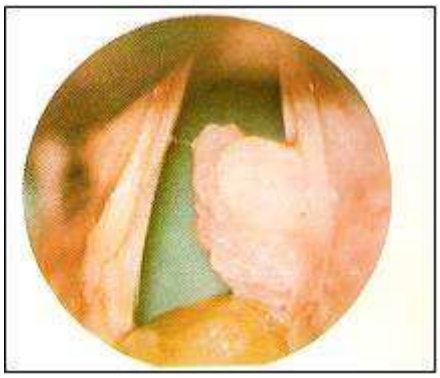


**Figura 2.16 – Tumor da região glótica.**

A etiologia do câncer de laringe é desconhecida, como a do câncer em geral, mas podem ser apresentadas causas que contribuem de uma maneira ou de outra para o desenvolvimento do agente efetivo da doença. Deve ser considerado o seguinte conjunto de fatores: idade, sexo, irritação e hereditariedade.

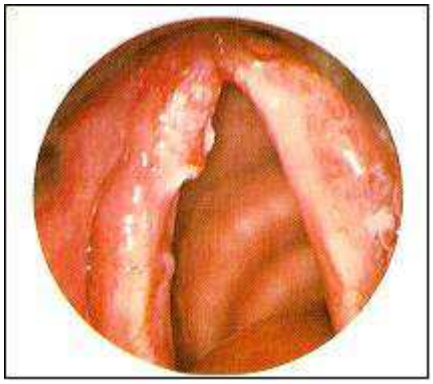


**Figura 2.17 – Câncer da dobra vocal esquerda.**
(Fonte: KUHL, 1982)

Surge normalmente entre 40 e 70 anos, considerando-se assim, as mudanças senis no epitélio como causas prováveis. Mais freqüente em homens, essa doença está intimamente ligada ao tabagismo e ao consumo de álcool.

O câncer laríngeo tem a sua localização mais freqüente na região glótica. Essa localização é explicada pela modificação de epitélio, pelos fatores irritativos e fricção cordal.

O sintoma clássico do câncer laríngeo é a disfonia. Como ocorre mais na região glótica, a disfonia surge logo, fazendo com que o doente procure o médico com certa rapidez.

O diagnóstico precoce é muito importante, sendo um fator decisivo no êxito do tratamento. Muitas vezes o tumor é tratado com métodos inadequados ou confundido com lesões benignas ou doenças comuns da laringe (KUHL, 1982).

### 2.4.7 Paralisia

Segundo Benjamin (2002), a paralisia da dobra vocal pode ser causada por patologia central ou periférica e pode envolver as conexões centrais, o nervo vago, o nervo laríngeo recorrente e/ou o nervo laríngeo superior.

A paralisia da prega vocal pode não ser um distúrbio neurológico isolado, mas pode ser uma manifestação de uma patologia maior ou de uma patologia sistêmica. A avaliação neurológica completa, com estudos intracranianos, deve excluir a doença neurológica central, como a esclerose múltipla em adultos ou hipertensão intracraniana aumentada em lactentes. A paralisia pode ser unilateral ou bilateral, podendo ser encontrada em lactantes, crianças e adultos. Podem ser ainda completas ou incompletas, aproximando ou afastando as dobras vocais numa maior ou menor amplitude.

Entre as causas de paralisia unilateral das dobras vocais em lactantes e crianças estão: lesão por trauma de parto, anomalias congênitas do coração e grandes vasos, cirurgias para corrigir essas anomalias, cirurgias intratorácicas de cistos e tumores, pós-intubação endotraqueal, entre outras ou, ainda, alguma causa não-definida. Acomete mais freqüentemente o lado esquerdo do que o direito.

As principais causas de paralisia unilateral de prega vocal em adultos são: viral ou idiopático, pós-cirúrgico (tireoidectomia, cirurgia cardíaca, entre outras), doença maligna do pescoço ou mediastino, pós-intubação endotraqueal, trauma cervical, entre outras.

A causa principal da paralisia bilateral das dobras vocais em adultos tem sido a tireoidectomia. Entre outras causas, podem ser citadas, ainda, doenças malignas do pescoço ou mediastino, pós-intubação endotraqueal, trauma cervical, doença neurológica ou alguma causa desconhecida.

Os principais tipos de paralisia motora são: paralisia em adução e paralisia da laringe em abdução. Uma lesão neurológica, unilateral ou bilateral, do grupo dos músculos adutores, pode anular a função de fechamento da fenda glótica.

Conseqüentemente, a fenda glótica permanece aberta, apresentando uma paralisia em abdução. Quando a paralisia ocorre no grupo dos músculos abdutores, isto é, dos músculos que abrem a fenda glótica, sendo bilateral, fecha a fenda glótica produzindo uma paralisia em adução (Figura 2.18) (KUHL, 1982).

Os sintomas perceptuais mais comuns da paralisia unilateral são a soprosidade e a rouquidão. Ocasionalmente, a diplofonia pode estar presente. A paralisia bilateral do tipo adutor causará soprosidade severa ou afonia, porém, uma voz quase normal pode estar presente no tipo abdutor. Pode ser observada uma maior aperiodicidade (*jitter* e *shimmer*), uma extensão de freqüência reduzida, níveis de ruídos elevados e uma extensão de intensidade vocal reduzida (PARRAGA, 2002).

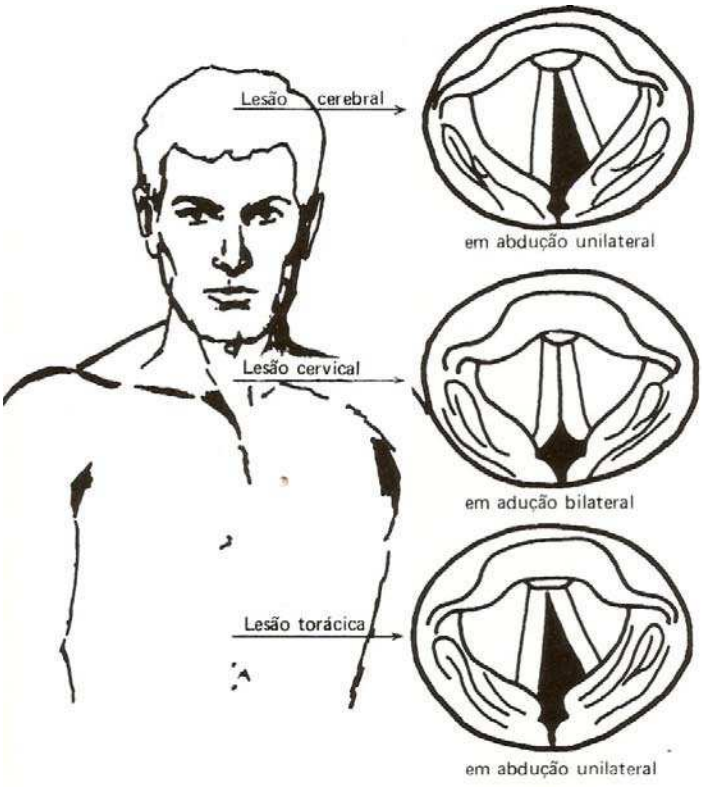



**Figura 2.18 – Paralisia nas dobras vocais.**

## 2.5 Exames e Procedimentos Realizados para Diagnóstico de Doenças da Laringe

As mudanças da voz causadas por patologias na laringe podem ser analisadas e diagnosticadas por um especialista, com o uso de diversos instrumentos e técnicas.

A laringoscopia é um dos procedimentos realizados para diagnóstico das doenças da laringe. Quando o exame é visualizado em monitor é denominado videolaringoscopia[6]. O exame é realizado sob anestesia tópica da faringe e da laringe supraglótica com o paciente sentado.

O exame é realizado pela boca, permitindo o diagnóstico das doenças da cavidade oral, orofaringe, hipofaringe e da laringe, em especial das dobras vocais. Atenção é dada a todas as estruturas da laringe em busca de lesões orgânicas ou funcionais.

Na maioria das vezes, as crianças permitem a realização da laringoscopia com relativa facilidade. Eventualmente, a laringoscopia pode não ser tolerada por náuseas ou resistência do paciente. Nessa eventualidade, pode ser feita a fibronasofaringolaringoscopia.

Fibronasofaringolaringoscopia é um exame realizado utilizando aparelho flexível com diâmetro de 3 *mm*. Permite a visualização das fossas nasais, rinofaringe, orofaringe, hipofaringe e laringe supraglótica, glótica e subglótica, podendo incluir a porção superior da traquéia. É feito sob anestesia tópica com o paciente sentado.

Broncofibroscopia ou videobroncofibroscopia é outro tipo de exame para diagnóstico de doenças da laringe. Mais apropriado seria chamá-lo de Laringotraqueobroncofibroscopia, uma vez que com este exame observa-se a laringe, traquéia, trônquios até brônquios subsegmentares, dependendo do diâmetro do aparelho empregado. O exame é feito com o paciente deitado, sob sedação ou, quando necessário, utilizando a anestesia ministrada por anestesiologista, o que torna o exame mais rápido e confortável para o paciente.

A broncoscopia rígida é feita sob anestesia geral e, na maioria das vezes, com propósitos terapêuticos, sendo então denominada Broncoscopia Terapêutica.

## 2.6 Discussão

Maus hábitos, como o tabagismo e o consumo excessivo de álcool, além do abuso vocal são as causas principais das patologias mais comuns da laringe como pólipos, nódulos, cistos e edemas de Reinke. Essas são as patologias das dobras vocais abordadas com maior aprofundamento neste trabalho, com ênfase para Edemas de Reinke.

[6] http://www.marloscoelho.com.br/conteudo.php?area=endoscopia_respiratoria&idioma=1

Vários exames e procedimentos são realizados por especialistas para detecção dessas patologias, bem como, em alguns casos são realizados procedimentos cirúrgicos, farmacológicos e fonoterápicos.

Geralmente, os exames para diagnóstico dessas patologias são considerados invasivos, podendo haver casos nos quais há a recusa do paciente em se submeter ao exame. Alguns necessitam de procedimentos de anestesia.

O diagnóstico precoce é essencial para que a patologia não se agrave, podendo, em alguns casos, comprometer as dobras vocais e, conseqüentemente, a produção vocal (paralisia, cânceres, etc.).

Neste trabalho, é indicado o uso da análise acústica do sinal de voz para auxílio ao pré-diagnóstico e acompanhamento fonoterápico das doenças da laringe que acometem as dobras vocais. É uma técnica não-invasiva, na qual, a partir da voz gravada do paciente, é possível, utilizando técnicas de processamento digital de sinais, a extração e a análise de características e parâmetros do sinal de voz. Por meio do estudo desses parâmetros e características é possível fazer um modelamento acústico da patologia em estudo.

A observação das mudanças no sinal de voz provocadas pela patologia possibilita a detecção ou o seu pré-diagnóstico, ao se comparar o comportamento do sinal de uma voz normal com o sinal patológico, sob os mesmo métodos de análise.

No capítulo a seguir, é feita uma revisão bibliográfica da análise acústica de sinais de vozes, em que serão apresentados vários parâmetros, usualmente empregados em técnicas de terapia vocal, importantes para avaliar a qualidade da voz.

# Capítulo 3

# Análise Acústica de Sinais de Vozes Normais e Patológicas

## 3.1 Introdução

A análise acústica de sinais de voz tem sido apontada como uma técnica não-invasiva, comparada aos exames laringoscópicos usuais, que pode ser utilizada como uma ferramenta adicional ao diagnóstico de uma dada patologia. A comparação do comportamento de várias medidas acústicas do sinal de voz patológico em relação à voz normal pode contribuir para pré-diagnosticar a presença de patologia.

A análise acústica pode ser utilizada, ainda, para acompanhamento de tratamento farmacológico, diminuindo a quantidade de exames laringoscópicos para o mesmo fim. Não se trata de eliminar o uso dos exames laringoscópicos, como também não se questiona a sua eficiência. No entanto, como muitos pacientes consideram o exame invasivo e, muitas vezes, se recusam a fazê-lo, a análise acústica pode ser utilizada para diminuir o número de exames laringoscópicos.

Técnicas de processamento digital de sinais de voz têm sido utilizadas para o desenvolvimento de ferramentas que proporcionem um suporte objetivo para o diagnóstico de desordens vocais, a determinação objetiva de alterações de funções vocais, avaliações de cirurgias e tratamentos farmacológicos e reabilitação.

A técnica de análise acústica também vem sendo utilizada para terapia vocal em pessoas com problemas na fala e por profissionais da voz. É uma técnica de custo relativamente baixo quando comparada aos exames usuais que precisam de fontes de luz especiais, instrumentos endoscópicos, e equipamentos de vídeo-câmera especializados (GODINO-LLORENTE et al, 2006).

Para a detecção eficiente de uma determinada patologia, com a análise acústica, torna-se necessário estudar e analisar o comportamento do sinal de voz patológico para que se possa levar a efeito um bom modelamento acústico da patologia em estudo. Um modelo acústico adequado possibilita a melhor escolha

acerca de quais características ou parâmetros definem e discriminam uma dada patologia de forma mais eficiente.

A maioria dos métodos descritos na literatura realiza a detecção automática de alterações vocais por meio da análise a longos intervalos de tempo (QI and HILMAN, 1997; YUMOTO et al, 1982; MICHAELIS et al, 1997; KROM, 1993; FEIJOO and HERNÁNDEZ, 1990; WINHOLTZ, 1992). Esses parâmetros a longos intervalos de tempo fornecem estimativas do grau de normalidade da voz, sendo obtidos a partir da média das perturbações locais medidas. Dentre esses parâmetros estão: *pitch*, *jitter*, quociente de perturbação de amplitude (APQ), quociente de perturbação do *pitch* (PPQ), relação harmônica-ruído (HNR), energia de ruído normalizada (NNE), índice de turbulência vocal (VTI), etc. Segundo estudos anteriores, a detecção de alterações na voz pode ser realizada utilizando esses parâmetros em um único vetor. No entanto, alguns desses parâmetros são baseados na estimação exata da freqüência fundamental, uma tarefa considerada complexa na presença de certas patologias (BOYANOV et al, 1993; MANFREDI et al, 1999)

Mais recentemente, métodos mais modernos vêm utilizando análise a curtos intervalos de tempo baseados na análise linear a partir de técnicas de filtragem inversa. Alguns métodos sugerem que a patologia introduz não-linearidades no sinal e, portanto, utilizam técnicas de análise não-linear (ZHANG et al, 2004; ZHANG & JIANG, 2004; DAJER, 2006; JIANG et al, 2006; MERGEL et al, 2000).

Neste capítulo, são introduzidos alguns conceitos referentes às medidas utilizadas na análise acústica de sinais de voz, que são relevantes ao estudo em foco, tais como energia, formantes, freqüência fundamental e análise por codificação preditiva linear (análise LPC – *Linear Predicitive Coding*) e Análise Cepstral. É feita uma revisão bibliográfica sobre algumas características e parâmetros do sinal de voz, a longos e a curtos intervalos de tempo, de forma a fornecer subsídios consistentes para a escolha daqueles que sejam bem representativos para as variações impostas, pela patologia em estudo, ao sinal de voz.

Esse estudo é focado, principalmente, em patologias orgânicas que afetam as dobras vocais e que aparecem como uma modificação da morfologia da excitação, produzindo um padrão de vibração irregular. Esse grupo inclui patologias como pólipos, nódulos, cistos e edemas, entre outras.

A voz patológica é induzida por um aumento de massa, uma falha de fechamento, ou uma mudança na elasticidade das dobras vocais. A conseqüência disto é que o movimento das dobras vocais não é equilibrado e um fechamento incompleto das dobras vocais pode aparecer em alguns ou em todos os ciclos glotais. Por esse motivo, mudanças podem surgir na estrutura harmônica completa (aumentando a energia inter-harmônicas e a perturbação na freqüência fundamental). Além disso, há um aumento de energia nas componentes mais altas devido à turbulência de ar induzida pelo fechamento incompleto da fenda glotal.

As faixas de freqüência mais altas, com menor energia, representam as componentes ruidosas introduzidas pela patologia, sendo a diferença fundamental entre a resposta do modelo para um sinal saudável ou normal e uma voz patológica.

Essas e outras alterações são mostradas por meio de gráficos de características de exemplos de sinais saudáveis e patológicos com a patologia edema nas dobras vocais, para representações de componentes temporais e espectrais desses sinais, no decorrer deste capítulo, a longo e a curto intervalo de tempo.

## 3.2 Medidas acústicas do sinal de voz

A seguir, são descritas algumas das medidas acústicas de sinais de vozes normais e sinais afetados pela patologia edema nas dobras vocais: freqüência fundamental, energia, formantes, além do comportamento do espectro LPC e do cepstro.

### 3.2.1 Freqüência Fundamental $F_0$ – *Pitch*

Freqüência fundamental ($F_0$) corresponde à freqüência do sinal de excitação proveniente da glote (Figura 3.1), ou seja, é o número de vibrações das dobras vocais por segundo. Sabe-se que a freqüência fundamental é determinada por uma interação complexa entre comprimento, massa e tensão

das dobras vocais, todos controlados pelos músculos intrínsecos e extrínsecos da laringe.

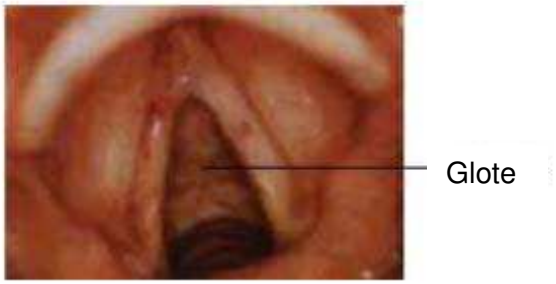



**Figura 3. 1– Visualização das dobras vocais e glote.**

Existem vários métodos para medição da freqüência fundamental (RABINER & SCHAFER, 1978). A freqüência fundamental pode ser mensurada determinando o inverso do intervalo de tempo transcorrido entre dois pulsos glotais sucessivos, ou tomando a freqüência correspondente à primeira harmônica do espectro de freqüências.

Outras formas de medição da freqüência fundamental são realizadas diretamente no domínio do tempo, por exemplo: a) método da Função da Média de Diferenças de Amplitudes (AMDF - *Average Magnitude Difference Function*) (RABINER & SCHAFER, 1978); b) método da função de autocorrelação (SONDHI, 1968; RABINER & SCHAFER, 1978); c) algoritmos que utilizam análise cepstral (NOLL, 1967) e d) medição a partir do resíduo da análise LPC (MARKEL & GRAY, 1976), sendo os dois primeiros os mais freqüentemente utilizados, descritos a seguir.

- **Método da Função da Média de Diferenças de Amplitudes - AMDF**

É considerado um método simples e eficiente. Considera-se o sinal $s(n)$ periódico, de período $P$. Seja

$$d(n) = s(n) - s(n+k)\text{,} \qquad (3.1)$$

em que $d(n)$ é zero para $k = 0, +P, -P, +2P, -2P$....

Para pequenos intervalos do sinal de voz, *d(n)* será mínimo para os valores de $k$, mas dificilmente será zero.

A definição da AMDF é dada pela equação

$$AMDF(k) = \frac{1}{F}\sum_{n=0}^{k_{\max}-1}\left|s(n)-s(n+k)\right|, \qquad k = 0,1,2,...,k_{\max}, \tag{3.2}$$

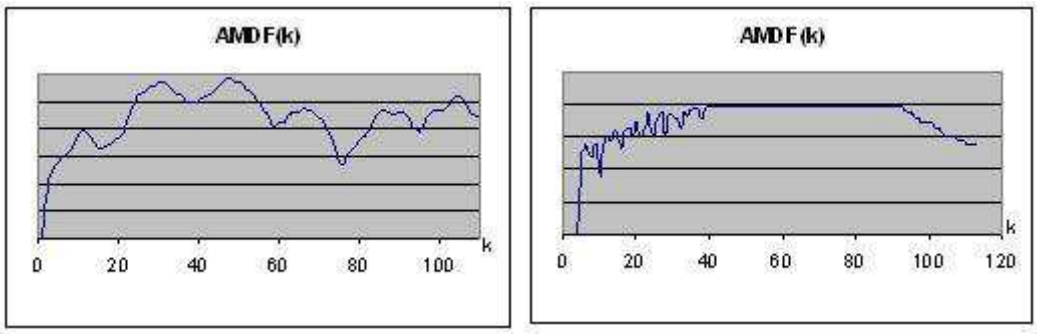
sendo *AMDF*($k$) o valor da AMDF para um atraso $k$ e $F$ é escolhido apropriadamente. Pode-se utilizar $F = k_{máx} = N_A/2$ ($N_A$ é o comprimento do quadro ou segmento da voz em análise) e eliminar a divisão por $F$, por ser desnecessária. Dessa forma, a Equação (3.2) pode ser reescrita como

$$AMDF = \sum_{n=0}^{\frac{N_A}{2}-1}\left|s(n)-s(n+k)\right|, \qquad k = 0,1,2,...,\frac{N_A}{2}. \tag{3.3}$$

A função AMDF de *d(n)* deve ser mínima para os valores de *k*, múltiplos do período. Para sons sonoros, ocorrem vales acentuados nos atrasos correspondentes ao período de *pitch*. Já para os sons surdos, estes vales não são observados (FECHINE, 2000).

Na Figura 3.2 são apresentados dois exemplos típicos da AMDF. Verifica-se que a função AMDF é mínima no período correspondente à freqüência fundamental e que não há mínimos comparáveis nos segmentos sem voz.

Para detectar o período correspondente à freqüência fundamental, é suficiente detectar o primeiro mínimo da função AMDF. É preciso, antes, fazer a detecção surdo/sonoro para os segmentos de voz, já que a função AMDF é aplicada nos segmentos sonoros do sinal de voz (FECHINE, 2000).




**(a)** **(b)**

**Figura 3.2 Exemplo de AMDF: (a) para quadro sonoro; (b) para quadro não sonoro**

- **Método da Autocorrelação**

A freqüência $F_0$ pode ser obtida tomando-se o inverso do tempo em que ocorrem dois picos sucessivos na função de autocorrelação. O menor valor limite ($T_1$) corresponde ao inverso da freqüência fundamental máxima admitida e o maior valor limite ($T_2$) corresponde ao inverso da freqüência fundamental mínima aceita pelo algoritmo. Como exemplo, a forma de onda do sinal de voz de uma criança do sexo masculino, de 8 anos de idade, produzindo a vogal /⊃/ (vovó) e a função de autocorrelação correspondente, são apresentadas na Figura 3.3 (ARAÚJO, 2000).

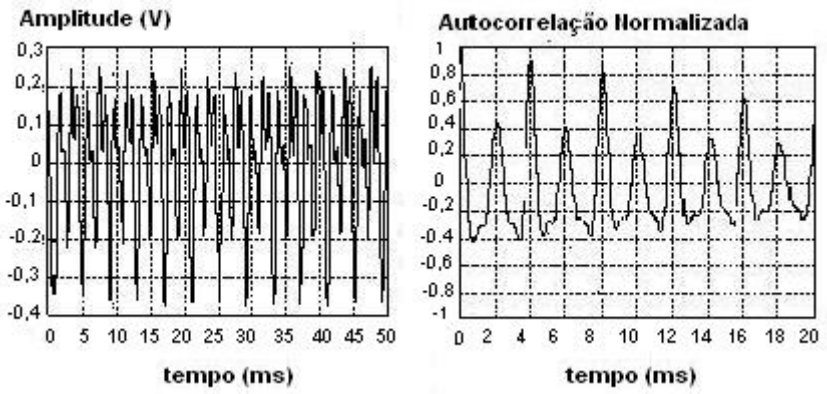



**Figura 3.3 - Função de autocorrelação obtida para vogal /⊃/ de uma criança do sexo o masculino, de 8 anos de idade.**

O pico máximo que ocorre em 4 ms corresponde à freqüência fundamental de 250 Hz. O pico em 2 ms, na função de autocorrelação, resulta das oscilações amortecidas intra-ciclos, introduzidas pela ação do trato vocal. O valor de pico da função de autocorrelação normalizada é comumente utilizado para definir se o quadro de sinal sob análise é ou não sonoro.

Markel & Gray (1976) propõem realizar a medida de $F_0$ a partir do resíduo da análise LPC. Rabiner & Schafer (1978) afirmam que esse procedimento é ineficiente para vozes com valores elevados de $F_0$, sendo, portanto, inadequado para vozes de crianças.

A partir da Figura 3.3 é possível perceber que a função de autocorrelação apresenta vários picos, cuja maioria pode ser atribuída às oscilações amortecidas entre pulsos glotais.

Em 1968, Sondhi propôs uma operação não linear sobre o sinal, buscando reduzir o efeito do trato vocal sobre a função de autocorrelação (SONDHI, 1968). Um dos efeitos da ação do trato sobre os pulsos glotais é a introdução de oscilações amortecidas entre estes pulsos. Sinais de voz apresentam diferenças de intensidade a cada ciclo, mesmo durante fonação sustentada.

Na transformação proposta por Sondhi, as amostras são subtraídas por uma constante (Cl). Diferenças absolutas de amplitude existentes entre os picos dos ciclos sucessivos do sinal, após o processamento não linear, são mantidas fixas sobre um sinal resultante com menor intensidade, acentuando, portanto, a diferença relativa entre ciclos sucessivos do sinal. Esse procedimento, entretanto, reduz o valor máximo da função de autocorrelação normalizada entre $T_1$ e $T_2$.

Um algoritmo para encontrar o valor do *pitch*, baseado no método da função de autocorrelação, pode ser resumido da seguinte forma (RABINER & SCHAFER, 1978):

1. O sinal de voz é filtrado com um filtro passa-baixas de 900 Hz e amostrado a uma taxa de 10 k amostras/s.
2. Segmentos de 30 ms de comprimento são selecionados em intervalos de 10 ms, havendo uma sobreposição nos segmentos de 20 ms.
3. A magnitude média é calculada com uma janela retangular de 100 amostras. O nível de pico do sinal é comparado a um limiar determinado pela medição do pico do sinal para 50 ms de ruído de fundo. Se o nível de pico do sinal está acima do limiar, significa que o segmento é voz e não ruído, então o algoritmo procede como a seguir; caso contrário, o segmento é classificado como silêncio e nenhuma ação a mais é executada.
4. O nível de truncamento é determinado como uma porcentagem fixa (*e.g.*, 68%) do mínimo dos valores máximos absolutos nas primeiras e últimas 100 amostras do segmento de voz.
5. Usando esse nível de truncamento, o sinal de voz é processado por um truncador central de três níveis e a função de autocorrelação é calculada sobre um intervalo dentro da faixa esperada de períodos de *pitch*.

6. O maior pico da função de autocorrelação é localizado e o valor de pico é comparado com um limiar fixo (isto é, 30% de R(0)). Se o pico cai abaixo do limiar, o segmento é classificado como surdo e se estiver acima, o período de *pitch* é definido como a localização do maior pico.

No caso do método da AMDF, tem–se certa facilidade de cálculo e custo computacional reduzido. Já no método da autocorrelação, observa-se uma maior complexidade computacional. No entanto, não há a preocupação em se fazer um algoritmo de detecção surdo/sonoro. Valores altos na função de autocorrelação já são uma indicação de que o segmento em análise é sonoro.

Para o caso de vozes patológicas, o ruído introduzido pela patologia (ver Figuras 3.4, 3.5 e 3.6) pode dificultar e até mesmo impedir o cálculo da freqüência fundamental pelo método da AMDF. Assim, pode haver grande dificuldade na detecção dos segmentos sonoros do sinal de voz em análise. Nesse caso, o método da autocorrelação é preferível.

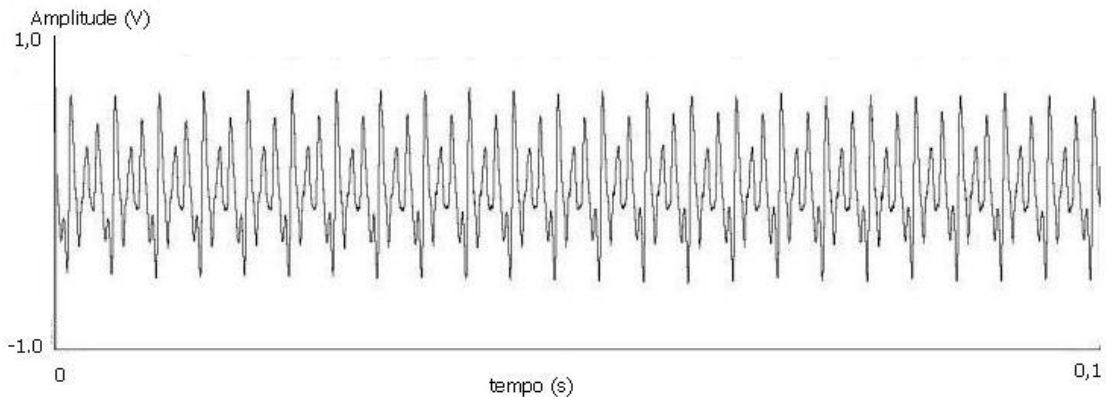



**Figura 3.4 – Trecho de 100 ms da vogal sustentada /a/, para voz normal.**

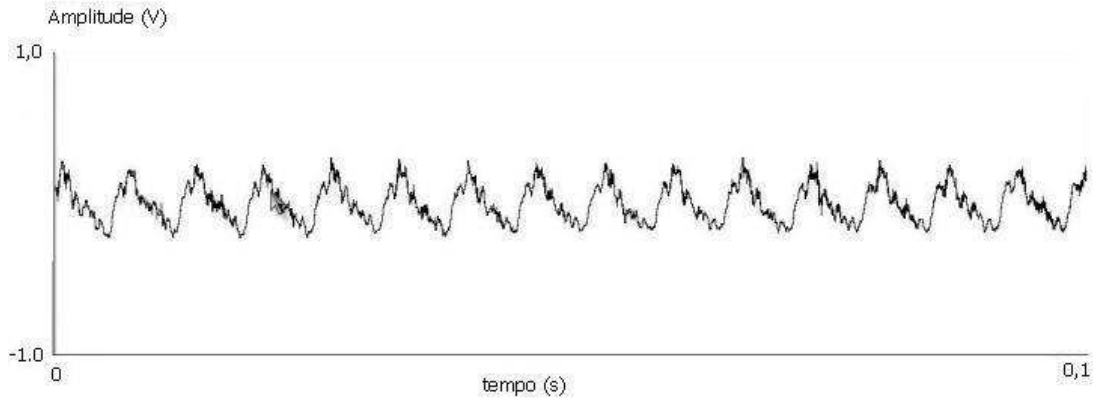



**Figura 3.5 – Trecho de 100 ms da vogal sustentada /a/, para voz afetada por edema unilateral nas dobras vocais.**

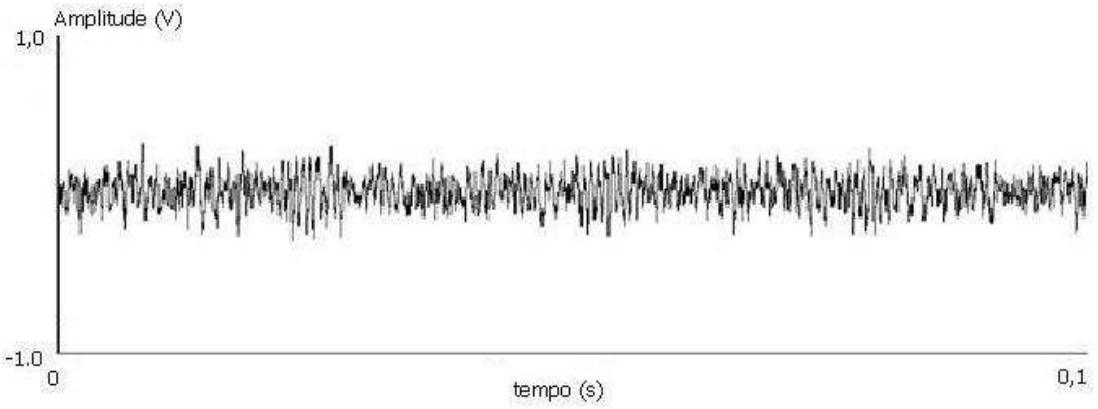



**Figura 3.6 – Trecho de 100 ms da vogal sustentada /a/, para voz afetada por edema bilateral nas dobras vocais.**

As medidas apresentadas neste capítulo foram obtidas com os *softwares Multi-Speech*, *Model* 3700, versão 2.5.2 e do *Multi-Dimensional Voice Program* (MDVP), *Model* 5105 da *Kay Elemetrics*, versão 2.6.2.

Observando as formas de onda apresentadas nas Figuras 3.4 a 3.6, as alterações na amplitude e na freqüência fundamental são evidentes. Entretanto, essas alterações observadas não são suficientes para determinar a existência de uma patologia e/ou classificá-la. Observando as formas de onda apresentadas, é possível sugerir uma anormalidade ou desordem na voz, sendo necessário um maior aprofundamento no estudo das medidas acústicas que representem bem a patologia que está causando a desordem.

Na Figura 3.7 e na Figura 3.8 é mostrado o comportamento da freqüência fundamental para os sinais das Figuras 3.4 e 3.5, durante 3 *s* para a voz normal e 1 segundo para a voz patológica (ambos do sexo feminino). Os valores médios da freqüência fundamental $F_0$ encontrados foram 234,97 Hz e 152 Hz, respectivamente. Já para o exemplo da Figura 3.6, locutor do sexo masculino, com edema bilateral severo, o valor médio de $F_0$ encontrado foi igual a zero.

É evidente a alteração na freqüência fundamental para os sinais de vozes afetadas por edema nas dobras vocais.

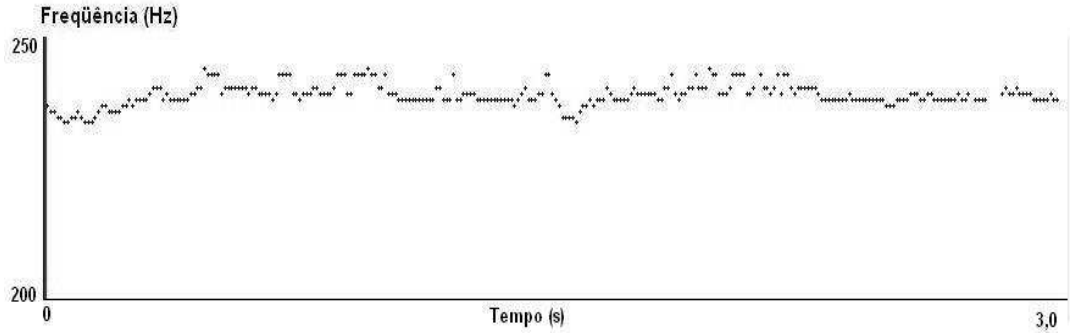



**Figura 3.7 – Comportamento do *Pitch* para um sinal de voz normal (vogal /a/ sustentada).**

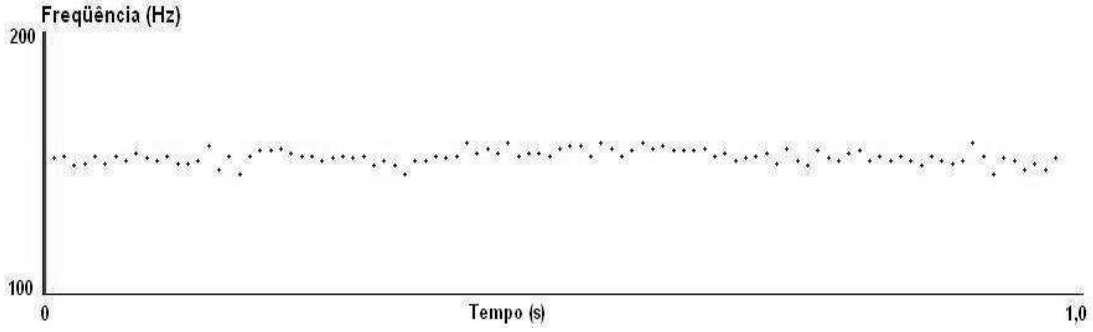



**Figura 3.8 – Comportamento do *Pitch* para um sinal de voz (vogal sustentada /a/)afetado por edema unilateral nas dobras vocais.**

Na Figura 3.9 é apresentado o comportamento do *pitch* (freqüência fundamental) para 53 sinais de vozes normais e 43 de vozes patológicas. Um dos sinais da base de dados, representado parcialmente na Figura 3.6 (Sinal DXC22AN.NSP da base de dados), é um caso de edema bilateral severo, no qual não foi possível encontrar o valor da freqüência fundamental. Observa-se que os valores encontrados nos casos patológicos são bem inferiores aos casos de voz normal. Ao escutar esses sinais, observam-se vozes roucas e muito roucas, nos casos mais severos da patologia. Há ainda a dificuldade em manter a vogal sustentada. Para os casos de vozes normais, a base de dados conta com três segundos de gravação, enquanto que nos casos patológicos, em média a duração é de um segundo.

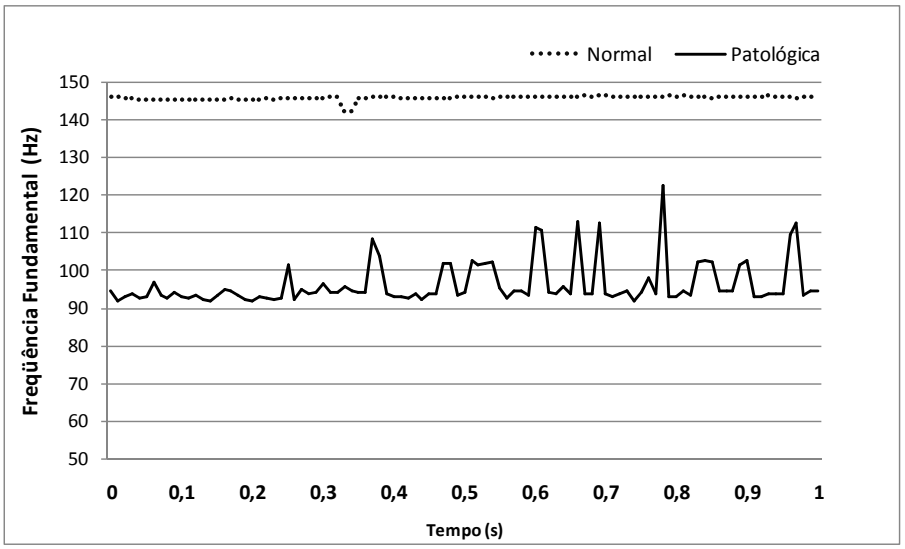



**Figura 3.9 – Comportamento da freqüência fundamental (*pitch*) em um segundo de voz masculina, normal e patológica.**

Na Figura 3.10 está ilustrada a evolução ou contorno de *pitch* para as vozes femininas. Novamente, se mantém o padrão dos valores mais baixos para os casos patológicos. Pode-se, inclusive, confundir-se uma voz feminina patológica com uma voz masculina, se for observado apenas o valor do *pitch*.

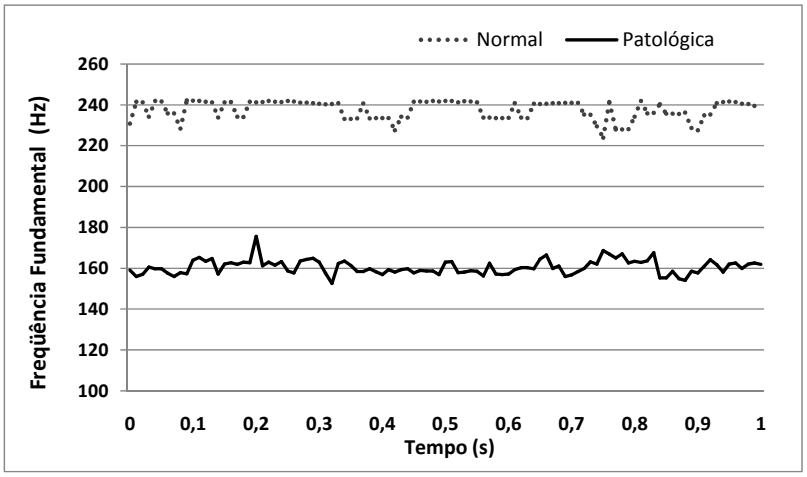



**Figura 3.10 – Comportamento da freqüência fundamental (*pitch*) em 1 segundo de voz feminina, normal e patológica.**

### 3.2.2 Energia

A medição da intensidade sonora, em processamento digital de sinais de voz, pode ser realizada por meio do cálculo da energia do sinal. Para tanto, são usadas técnicas no domínio do tempo (análise temporal) e/ou no domínio da freqüência (análise espectral) (RABINER & SCHAFER, 1978; DELLER, PROAKIS and HANSEN, 1987).

As propriedades estatísticas dos sinais de voz podem ser consideradas invariantes no tempo, para curtos intervalos, até 32 ms, sendo um valor típico, 16 ms. Utilizando dessa característica, procura-se obter os parâmetros temporais do sinal a partir de segmentos que se situem nesse intervalo de interesse mantendo-se, assim, a estacionaridade do sinal de voz (RABINER & SCHAFER, 1978; FECHINE, 2000).

Assim, a energia segmental, $E_{seg}$, é definida como

$$E_{seg} = N_A \cdot E\{[s(n) - \mu_s(n)]^2\}\,, \tag{3.4}$$

em que $s(n)$ é o sinal de voz, $\mu_s(n)$ a média de $s(n)$ e $N_A$ é o número de amostras do segmento em análise.

Considerando, ainda, ergodicidade, estacionaridade no sentido amplo e média nula para o sinal de voz no intervalo citado, a $E_{seg}$ é definida por

$$E_{seg} = N_A \cdot E\{[s(n)]^2\} = \sum_{n=0}^{N_A-1} [s(n)]^2 \tag{3.5}$$

e

$$E_{seg}(dB) = 10\log[E_{seg}]\,. \tag{3.6}$$

Na Figura 3.11 está representado o comportamento da energia média segmental para os 53 sinais de vozes normais e 44 de vozes patológicas afetados por edema. A intensidade média para os sinais de vozes com a citada patologia é cerca de 3 dB mais baixa.

A energia é um parâmetro útil também na diferenciação entre segmentos surdos e sonoros do sinal de voz, já que a amplitude nos segmentos surdos é bem mais baixa do que nos segmentos sonoros.

Quando se pretende distinguir entre surdos e fricativos ou até mesmo os períodos de silêncio, pode haver certa confusão se for usada unicamente a energia. Isto ocorre porque, nesses casos, os valores de energia são bem próximos (valores baixos), havendo a necessidade de outros parâmetros para análise e decisão corretas (RABINER &SCHAFER, 1978; FECHINE, 2000).

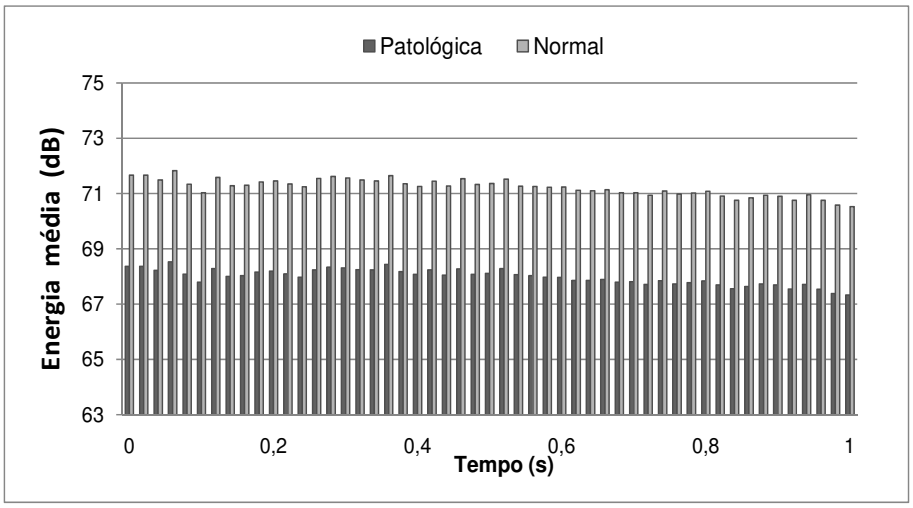



**Figura 3.11 - Energia segmental média para sinais de vozes normais e patológicas com edema nas dobras vocais.**

### 3.2.3 Formantes

O sinal de excitação é modulado em sua passagem pelo trato vocal, por uma envoltória correspondente à função de transferência do trato vocal (Figura 3.12). Os picos dessa envoltória correspondem às freqüências de ressonância do trato vocal, que por sua vez, dependem da posição dos articuladores. Desse modo, sua posição, amplitude e largura de banda, podem dar uma idéia da configuração do trato vocal no momento da articulação. As freqüências correspondentes a esses picos são chamadas de formantes, geralmente designados por $F_1$, $F_2$, ..., $F_n$ (primeiro formante, segundo formante, ..., $n$-ésimo formante) (TUJAL, 1998).

As vogais podem ser classificadas segundo sua posição num plano $F_1xF_2$, visto na Figura 3.13, na qual são apresentados os valores médios dos primeiros formantes para homens, mulheres e crianças, para o português brasileiro.

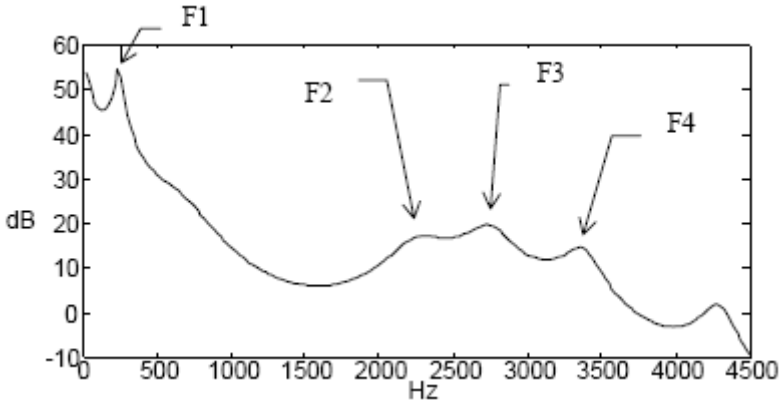



**Figura 3.12 – Espectro da vogal */i/.***
Fonte: (TEIXEIRA, 1995).

Assim como cada vogal apresenta suas freqüências formantes características, devido à configuração geométrica do trato vocal, cada indivíduo apresenta seus formantes particulares, para uma determinada vogal, devido às dimensões das estruturas do trato vocal, além do padrão articulatório pessoal. Sendo os valores absolutos das freqüências dos formantes variáveis de indivíduo para indivíduo, é a relação entre as freqüências $F_1$ e $F_2$, que determina a qualidade de uma vogal, em termos acústicos.

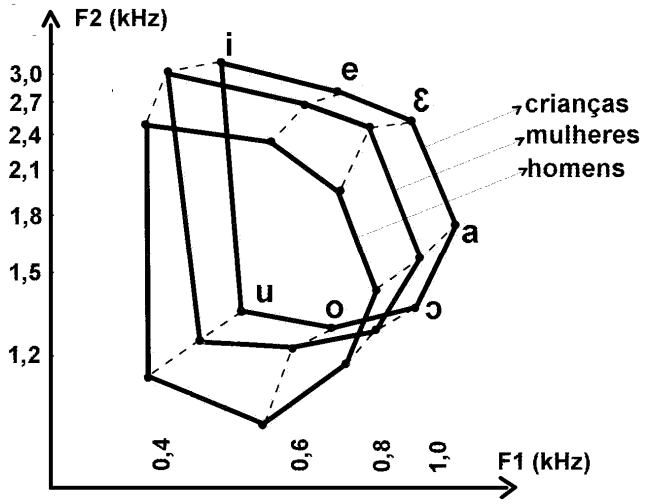



**Figura 3.13 - Classificação das vogais pela sua localização no espaço formado pelo primeiro e segundo formantes, $F_1$ e $F_2$ (RUSSO e BEHLAU, 1993).**

O tamanho das cavidades ressonantes do trato vocal varia de acordo com o sexo e com a idade da pessoa, o que implica em uma variação na freqüência dos formantes, mais graves para adultos do sexo masculino e mais agudas para crianças (RUSSO e BEHLAU, 1993).

De acordo com as freqüências dos formantes e tendo em vista o grau de anteriorização das vogais classificadas no plano horizontal, pode-se considerar duas diferentes situações (ver gráfico na Figura 3.13):

- Das vogais anteriores /i,e, ɛ/ em direção à vogal central /a/ - neste caso, observa-se um incremento na freqüência $F_1$ e, por sua vez, as freqüências $F_2$ e $F_3$ relativamente próximas entre si, decrescem em direção à vogal central /a/;
- Da vogal central /a/ em direção às sucessivas vogais posteriores /ɔ,o,u/ a situação não é tão clara como no caso das vogais anteriores, embora seja evidente o decréscimo nas freqüências $F_1$, não há aumento significativo de $F_2$ e $F_3$, que têm valores praticamente iguais, em algumas vogais, como no caso de /o/ e /u/.

Portanto, em relação às formantes $F_2$ e $F_3$, não há uma definição clara para as vogais posteriores, o que é também observado na língua inglesa (RUSSO e BEHLAU, 1993), o que pode explicar a maior quantidade de erros em discriminação gerada por essas vogais.

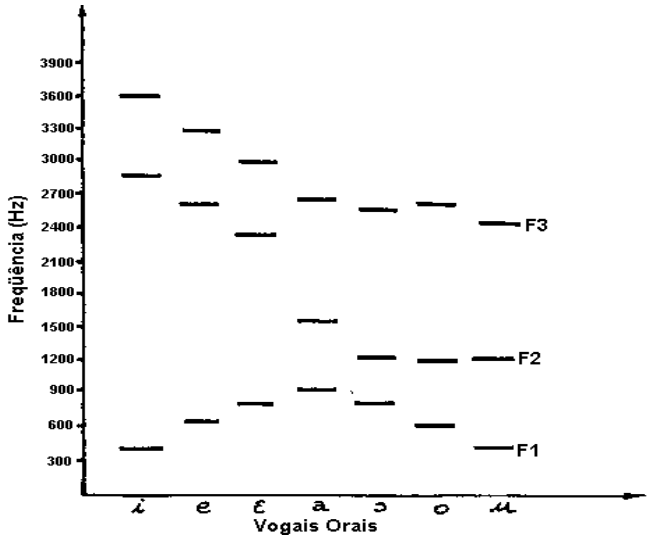
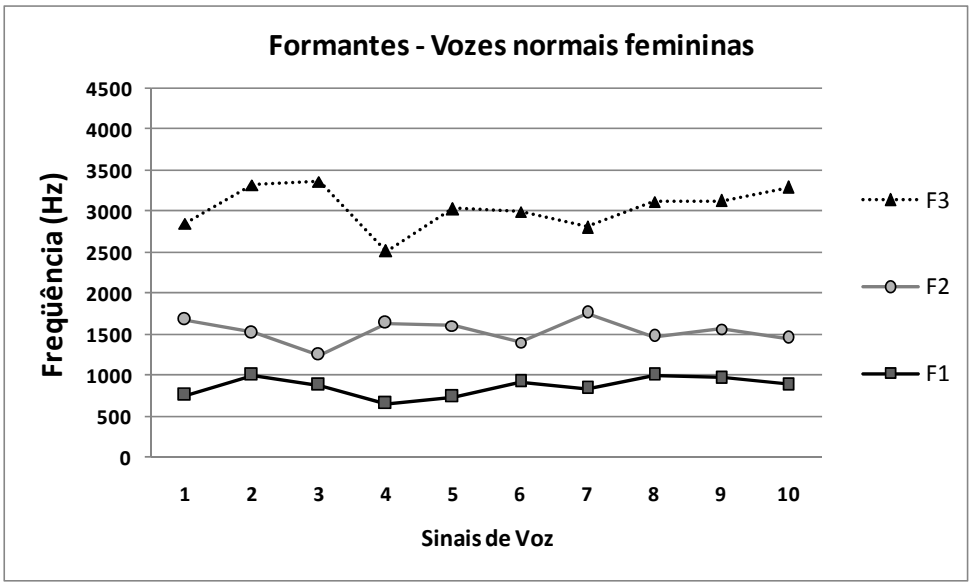
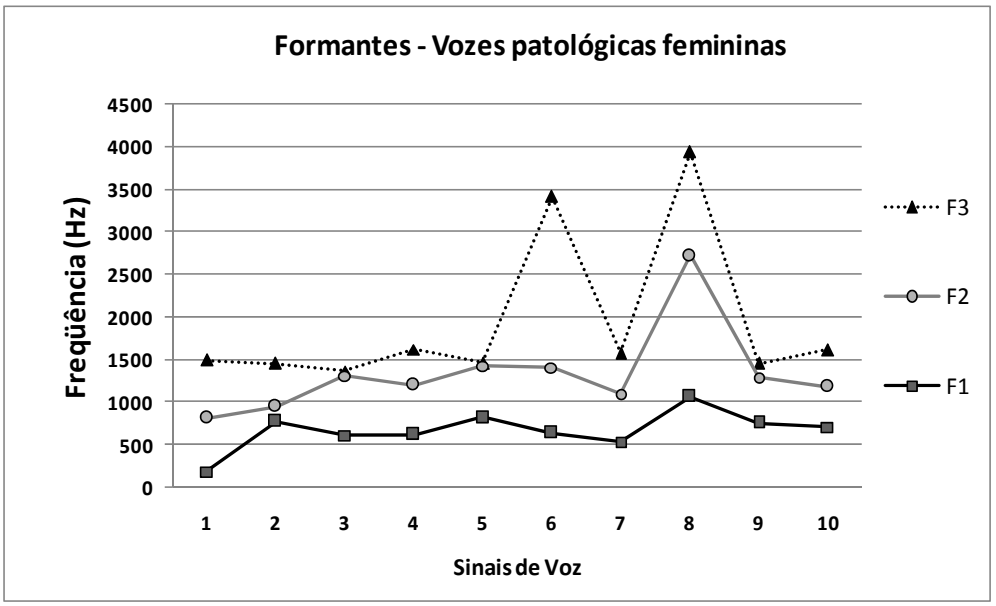



**Figura 3.14 - Freqüências dos três primeiros formantes das vogais do português brasileiro (RUSSO e BEHLAU, 1993).**

Na Figura 3.15 é apresentado o comportamento dos formantes para 10 sinais de vozes femininas normais (Fig. 3.15(a)) e patológicas (Fig. 3.15(b)). Os formantes foram extraídos usando o *software Multi-Speech,* da *Kay Elemetrics* (KAY ELEMETRICS, 1994).

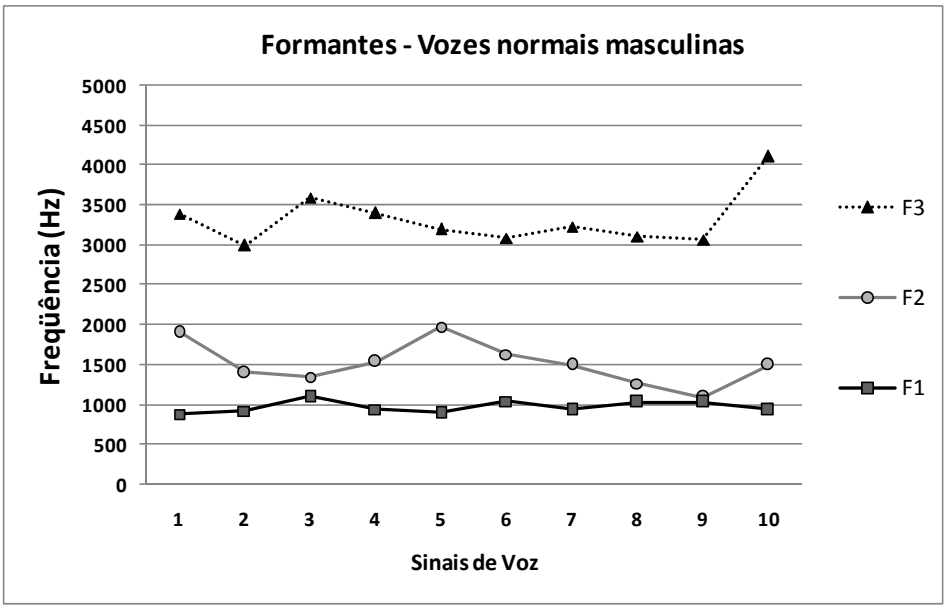



(a)

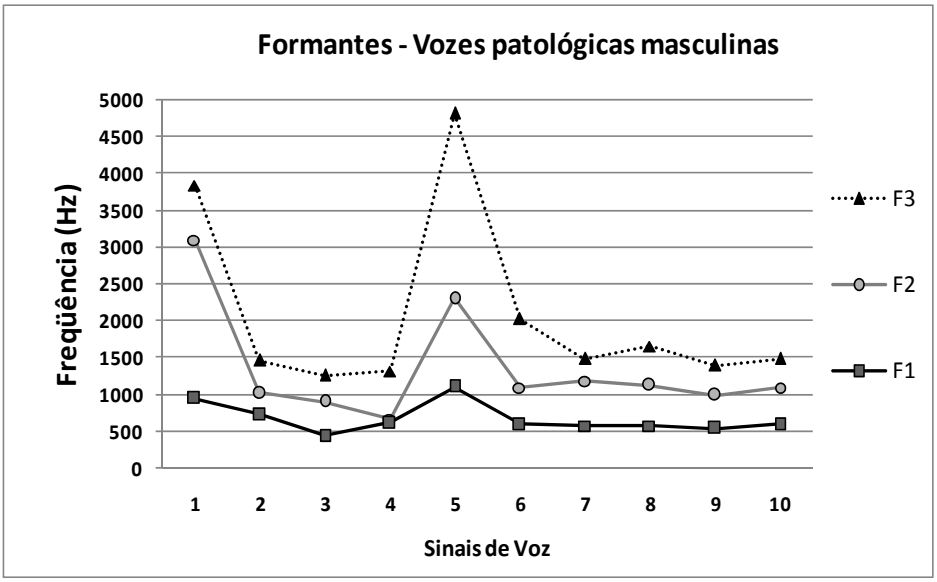



(b)

**Figura 3.15 – Comportamento dos formantes para sinais de vozes femininas: (a) Vozes normais e (b) Vozes patológicas (edemas).**




(a)




(b)

**Figura 3.16 – Comportamento dos formantes para sinais de vozes masculinas: (a) Vozes normais e (b) Vozes patológicas.**

Ao comparar o caso normal com o patológico, observa-se um comportamento mais uniforme e mais de acordo com os valores indicados na literatura, como aqueles indicados na Figura 3.13. Os valores dos formantes para o caso patológico encontram-se, em geral, mais baixos, apresentando, no entanto, alguns valores acima da faixa para voz normal. Esses casos ocorrem para sinais nos quais a patologia encontra-se mais severa.

Algumas técnicas usuais de análise acústica utilizam medidas de perturbação da voz, entre as quais Quociente de Perturbação de Amplitude, Índice de Tremor de Amplitude, *Shimmer* (índice de perturbação em amplitude do *pitch*), *jitter* (índice de perturbação em freqüência do *pitch*), Quociente de Perturbação do Período de *pitch* (PPQ). Muitas dessas medidas dependem da obtenção correta do *pitch* para que a análise seja eficiente. Em sinais de vozes afetados por patologias nas dobras vocais, principalmente em casos em que a patologia é de moderada a severa, o *pitch* é bastante afetado. Em alguns casos, torna-se muito difícil sua obtenção, dependendo do tipo de algoritmo empregado.

Nesse tipo de análise, necessita-se, geralmente, de uma grande quantidade de parâmetros para se chegar a um diagnóstico conclusivo.

Neste trabalho, em particular, sugere-se o uso de técnicas de análise acústica baseada em medidas acústicas que não dependam diretamente da obtenção do *pitch* e que utilize uma quantidade reduzida de parâmetros, facilitando a análise. Para tanto, são utilizados os coeficientes obtidos por meio da análise LPC e cepstral. Como base para discussão dos resultados obtidos na abordagem proposta, é apresentada uma revisão da literatura sobre Codificação por Predição Linear do sinal de voz (análise LPC – *Linear Predictive Coding*), sobre Análise Cepstral e Mel-cepstral. Os resultados obtidos no trabalho, a partir da análise LPC e cepstral, e sua discussão são apresentados no Capítulo 5.

## 3.3 Análise por Predição Linear e Análise Cepstral do Sinal de Voz

A tarefa de avaliação acústica de vozes patológicas está relacionada à extração de características. Parâmetros estatísticos específicos, baseados no modelo de produção de fala, podem ser usados como características acústicas significativas.

Sabe-se que o sinal de voz é produzido como o resultado de pulsos glotais ou como um sinal que varia aleatoriamente, como excitação ruidosa filtrada pelo trato vocal (RABINER AND SCHAFER, 1978).

Patologias como edemas afetam as dobras vocais ou outros componentes do sistema vibratório, produzindo uma vibração mais irregular. De fato, é amplamente conhecido que patologias nas dobras vocais podem apresentar

variação do movimento vibratório por causa das mudanças na elasticidade das dobras vocais. Isso ocorre devido ao fechamento incompleto das dobras vocais em todos os ciclos glotais. Portanto, as mudanças na morfologia das dobras vocais podem provocar modificações significativas no sinal de voz (GODINO LLORENTE et al, 2006).

Embora a patologia esteja localizada no sistema vibratório pode, ainda, afetar o movimento articulatório regular durante a produção da fala. Além disso, componentes do sistema de ressonância podem ser afetadas, resultando em mudanças no trato vocal, produzindo irregularidades nas propriedades espectrais. Portanto, alterações nas vozes desordenadas podem ser observadas tanto nas modificações da freqüência fundamental quanto da envoltória espectral do sinal (GODINO-LLORENT et al, 2006).

A chave para a modelagem acústica de vozes desordenadas é o entendimento das mudanças, referentes às medidas acústicas, produzidas pelos efeitos da fonte de excitação e do trato vocal (COSTA et al, 2008a).

Um dos principais desafios da modelagem acústica é capturar a variabilidade presente no sinal de voz. A variabilidade provém da natureza dinâmica do trato vocal. Assim, a voz é dinâmica ou variante no tempo e a modelagem precisa considerar dois aspectos: 1) as dependências temporais explícitas do sinal de voz, e 2) a estimação das características. Esses aspectos têm que estar baseados na análise estatística a curto intervalo de tempo. O modelo deve representar o comportamento das irregularidades introduzidas pela própria patologia.

Essencialmente, dois métodos paramétricos baseados no modelo linear para o mecanismo de produção da fala humana têm sido considerados até então. O primeiro é obtido da análise preditiva linear (*Linear Predictive Coding* – LPC) e o segundo é baseado na análise cepstral (AGUIAR NETO et al, 2007a; AGUIAR NETO et al, 2007b; MARINAKI et al, 2004; PARSA & JAMIESON, 2001; ROSA et al, 2000; GAVIDIA-CEBALLOS & HANSEN, 1996).

### 3.3.1 Codificação por predição linear do sinal de voz - Análise LPC

A teoria acústica da produção da fala é constituída de representações matemáticas do processo de produção da fala e tem sido usada como base para

toda a análise e síntese realizada com os sinais da fala (RABINER and SCHAFER, 1978).

O modelo básico para produção da fala é constituído por um gerador de excitação e um sistema linear variante no tempo (Figura 3.17). O gerador de excitação deve fornecer dois tipos de saída: um trem de pulsos (glotais) para sinais sonoros e ruído aleatório para sinais não-sonoros. Os efeitos de radiação dos lábios e do trato vocal são produzidos pelo sistema linear. Os parâmetros da fonte e do sistema devem ser escolhidos de forma tal que a saída resultante tem as propriedades semelhantes à voz desejada. Se isto puder ser feito, o modelo serve como uma base útil para o processamento de sinais de voz (RABINER & SCHAFER, 1978).

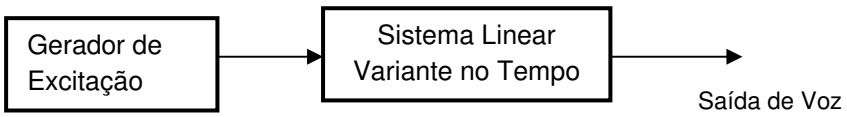



**Figura 3.17 - Modelo simplificado de produção de fala.**

Um modelo detalhado para geração propagação e irradiação do som pode, em princípio ser solucionado com valores adequados dos parâmetros da excitação e do trato vocal para calcular uma forma de onda da voz na saída. A teoria acústica fornece uma técnica simplificada, bastante utilizada, para modelar sinais de voz, que apresenta a excitação separada do trato vocal e da radiação. Os efeitos da radiação e do trato vocal são representados por um sistema linear variante no tempo (RABINER & SCHAFER, 1978).

O modelo completo é mostrado na Figura 3.18. Para a produção dos sinais sonoros é gerado um trem de impulsos unitários cuja periodicidade é determinada pelo período de *pitch* $T_0$ ($T_0$=1/$F_0$), com $F_0$ representando a freqüência fundamental do sinal da fala. Esse trem de impulsos é aplicado a um filtro digital $G(z)$ que simula o efeito dos pulsos glotais, que são devidamente selecionados e aplicados ao trato vocal, após um controle de ganho $A_V$. Para a produção dos sinais não-sonoros é utilizado um gerador de ruído aleatório com espectro plano e um controle de ganho $A_N$ (RABINER & SCHAFER, 1978).

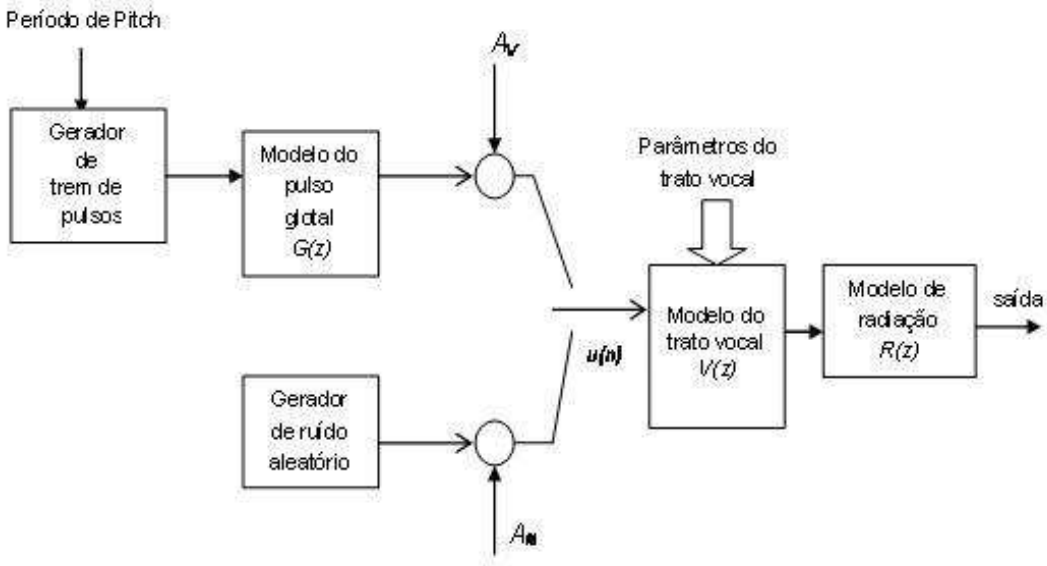



**Figura 3.18 - Modelo geral discreto no tempo para produção de fala.**

Na Figura 3.18, $u(n)$ é o sinal de excitação e $A_V$ e $A_N$, agem controlando a intensidade da excitação do sinal de voz e do ruído, respectivamente.

Chaveando entre geradores de excitação sonora e não-sonora alterna-se o modo de excitação. O trato vocal pode ser modelado de várias formas. Em alguns casos, é conveniente combinar o pulso glotal e modelos de radiação em um sistema simples. No caso de análise por predição linear, as funções do pulso glotal, radiação e componentes do trato vocal, podem ser combinadas em uma única função $H(z)$, representando o processo de produção da fala, como descrito na Equação (3.7).

$$H(z) = G(z) \cdot V(z) \cdot R(z), \tag{3.7}$$

em que $G(z)$, $V(z)$ e $R(z)$, representam a transformada-*z* dos modelos do pulso glotal, do trato vocal e da radiação, respectivamente.

Um segmento de voz sonoro $s_s(n)$ pode ser modelado como a saída gerada por um trem de impulsos, $P_h(n)$, convoluído com as respostas do filtro $g(n)$ (resposta glotal), $v(n)$ e $r(n)$ (radiação nos lábios), como mostrado na Equação (3.8) (GAVIDIA-CEBALLOS AND HANSEN, 1996; DELLER, PROAKIS e HANSEN, 1993).

$$s_s(n) = p_h(n) * g(n) * v(n) * r(n) \quad (3.8)$$

O termo $p_h(n)$ representa o modelamento da taxa periódica do movimento das dobras vocais em condições saudáveis, $g(n)$ modela a resposta ao impulso da excitação de um único período de *pitch* na glote, $v(n)$ modela a estrutura ressonante do trato vocal e $r(n)$ representa o modelo de radiação nos lábios. O termo $H_{TV}(n)$, na Figura 3.19, representa o modelo combinado para $V(w)$ e $R(w)$. Isso representa o modelo geral usado para representação da produção de fala em condições saudáveis.

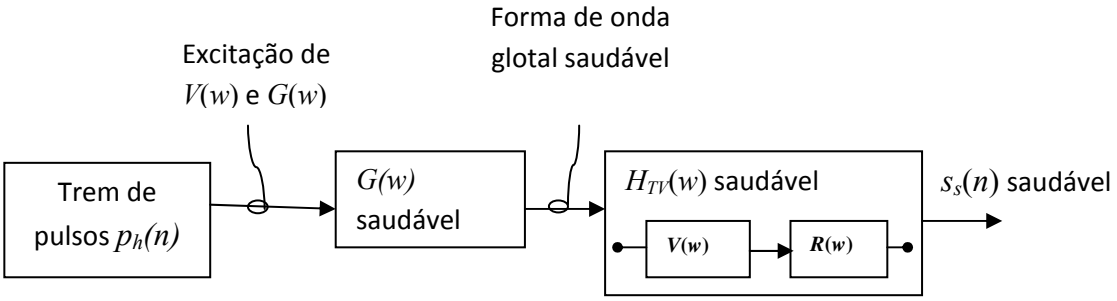



**Figura 3.19 – Modelo de produção de fala sob condições saudáveis.**
(FONTE: GAVIDIA-CEBALLOS AND HANSEN, 1996).

Considerando um modelo de produção de fala sob condições patológicas (Figura 3.20), supõe-se que: o fator que causa a patologia é estacionário durante a produção da fala; que a análise é limitada a curtos intervalos de tempo de vogal sustentada e que a detecção da patologia pode ser realizada sem exercitar a faixa de movimento dos articuladores do trato vocal, ou seja, a função ou comportamento do trato vocal é o mesmo que no caso patológico. A diferença fundamental, para o foco deste trabalho, que trata de doenças nas dobras vocais está na resposta glotal. Embora o foco principal seja o de vozes afetadas por edemas nas dobras vocais, o estudo pode também ser estendido para outras patologias nas dobras vocais como nódulos, pólipos, cistos, câncer nas dobras vocais, granulomas, papilomas e paralisia nas dobras vocais.

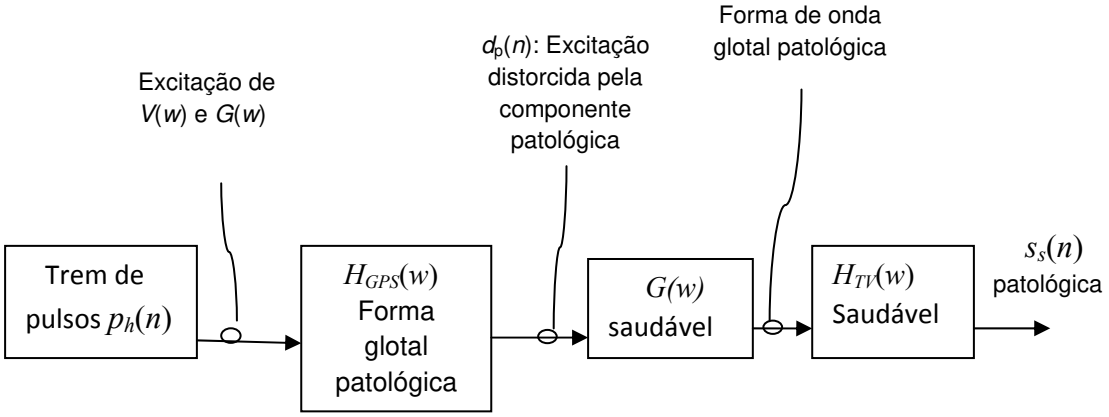



**Figura 3.20 – Modelo de produção de fala para o caso de voz patológica.**
(FONTE: GAVIDIA-CEBALLOS AND HANSEN, 1996).

Para o presente estudo, no caso ideal, as características saudáveis $G(w)$ e $H_{TV}(w)$ são exatamente as mesmas que na condição patológica. Um filtro de conformação patológica $H_{GPS}$ (GPS – *Glotal Pathological Shape*) é incorporado para levar em consideração a distorção introduzida pela patologia na glote na característica glotal e de excitação saudáveis, ou seja,

$$d_p(n) = p_h(n) * h_{gps}(n) \,. \tag{3.9}$$

O modelo linear de produção da fala (Figura 3.18) incorpora os efeitos dos pulsos glotais, trato vocal e da radiação dos lábios como um filtro linear (RABINER and SCHAFER, 1978). A fonte é uma seqüência de impulsos quase-periódicos utilizados para gerar sons sonoros ou a adição de uma seqüência de ruído aleatório para sons surdos. Um fator de ganho, $G$, é ajustado para controlar a intensidade da excitação. Combinando os efeitos dos pulsos glotais do trato vocal e da radiação o modelo pode ser representado por uma função de transferência de apenas pólos, $H(z)$, como

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} \alpha(k) z^{-k}}, \tag{3.10}$$

em que $S(z)$ e $U(z)$ representam as transformadas Z do sinal de voz, $s(n)$, e do sinal de excitação, $u(n)$, respectivamente. Os termos $\alpha(k)$ representam os

coeficientes de predição linear (coeficientes LPC) e $p$ a ordem do filtro de predição.

A principal vantagem do modelo é que o ganho, G, e os coeficientes do filtro podem ser estimados, de uma forma eficiente computacionalmente pelo método da predição linear (AGUIAR NETO, 1987; MAMMONE et al, 1996; RABINER & SCHAFER, 1978). Desde que a voz é invariante no tempo e que a configuração do trato vocal muda com o tempo, um conjunto preciso de coeficientes do preditor é determinado, adaptativamente, sobre intervalos curtos de tempo (tipicamente 16 a 32 ms), assumindo invariância no tempo (MAMMONE et al, 1996).

O método LPC estima cada amostra de voz baseado numa combinação linear de $p$ amostras anteriores. Um valor de $p$ maior representa um modelo mais preciso. A análise LPC fornece um conjunto de parâmetros da fala que representa o trato vocal. Espera-se que, qualquer mudança na estrutura anatômica do trato vocal, devido à patologia, afete os coeficientes LPC. Um preditor linear com coeficientes de predição, $\alpha(k)$, é definido como um sistema cuja saída é (RABINER and SCHAFER, 1978)

$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha(k) s(n-k). \tag{3.11}$$

em que $p$ é a ordem do preditor.

Existem várias formulações diferentes para a predição linear, sendo que algumas delas são equivalentes entre si. O método da autocorrelação e o método da covariância são dois métodos padrões de solução para cálculo dos coeficientes do preditor (RABINER AND SCHAFER, 1978; O'SHAUGHNESSY, 2000). Ambos os métodos são baseados na minimização do valor médio quadrático do erro de estimação $e(n)$, ou sinal residual, como dado por

$$e(n) = s(n) - \sum_{k=1}^{p} \alpha(k) s(n-k), \tag{3.12}$$

em que $p$ é a ordem do preditor.

O método utilizado neste trabalho foi o método da autocorrelação, descrito a seguir.

- **Determinação dos coeficientes LPC pelo método da autocorrelação**

O sinal $s(n)$ é janelado (janela de *Hamming*, por exemplo) para limitar a extensão do sinal de voz em análise, assegurando a sua estacionaridade, no intervalo considerado usualmente de 16 a 32 ms (RABINER & SCHAFER, 1978)

$$x(n) = w_h(n) \cdot s(n). \tag{3.13}$$

Os coeficientes LPC descrevem, portanto, uma média suavizada do sinal.

Seja $E$ a energia do erro (O'SHAUGHNESSY, 2000)

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} [x(n) - \sum_{k=1}^{p} \alpha(k) \cdot x(n-k)]^2, \tag{3.14}$$

em que $e(n)$ é o sinal residual correspondente ao sinal janelado $x(n)$.

Os valores de $\alpha(k)$ que minimizam $E$ são encontrados fazendo $\partial E / \partial \alpha_k = 0$, para $k$=1, 2, 3, ....$p$. Isso produz $p$ equações lineares

$$\sum_{n=-\infty}^{\infty} x(n-i)x(n) = \sum_{k=1}^{p} \alpha(k) \sum_{n=-\infty}^{\infty} x(n-i)x(n-k), \quad \text{para } i = 1, 2, 3, ..., p. \tag{3.15}$$

Tem-se que

$$R_{xx}(i) = \sum_{n=1}^{N-1} x(n) \cdot x(n-i), \qquad \text{para } i = 1, 2, 3, ..., p, \tag{3.16}$$

em que $N$ representa o número de amostras do quadro em análise e $p$ a ordem do preditor.

Portanto, a Equação (3.14) se reduz a

$$\sum_{k=1}^{p} \alpha(k) R_{xx}(i-k) = R_{xx}(i), \qquad \text{para } i = 1, 2, 3, ..., p. \tag{3.17}$$

A Eq. (3.17) pode colocada no formato $\mathbf{R}_{xx} A = \mathbf{r}_{xx}$, em que $\mathbf{R}$ é uma matriz $p$ $x$ $p$ de elementos $R_{xx}(i,k)=R_{xx}(|i-k|),\ (1 \leq i,\ k \leq p)$, $\mathbf{r_{xx}}$ é um vetor coluna $(R_{xx}(1),\ R_{xx}(2),\ ...,\ R_{xx}(p))^{\mathrm{T}}$, e $\mathbf{A}$ é um vetor coluna de coeficientes LPC $(a_1,\ a_2,\ ...,\ a_p)^{\mathrm{T}}$. A matriz quadrada de autocorrelação do sinal $\mathbf{R_{xx}}$ tem forma *Toeplitz* simétrica (simetria diagonal e todos os elementos iguais em qualquer linha paralela à diagonal), o que facilita a resolução do sistema, cujos coeficientes pode ser determinados numericamente pelo método de Levinson-Durbin (MAKHOUL, 1975; RABINER and SCHAFER, 1978).

O *software* Multi-Speech 3700 oferece a opção da obtenção do espectro LPC de um sinal de voz, a partir da gravação pelo usuário, ou dos sinais da base de dados , modelo 4337 (KAY ELEMETRICS, 1994).

Para efeito visual comparativo, na Figura 3.21 é apresentado o espectro LPC, obtido pelo método da autocorrelação, com um filtro de predição de 12 coeficientes (podendo variar até 36), janela de *Hamming*, pré-ênfase (0,95) e tamanho do quadro de 20 ms, com sobreposição de 50%.

Na Figura 3.22 é apresentado o espectro LPC para um sinal de voz com edema unilateral nas dobras vocais e na Figura 3.23, para um sinal de voz com edema bilateral. Os três espectros referem-se aos sinais de voz mostrados nas Figuras 3.4, 3.5 e 3.6

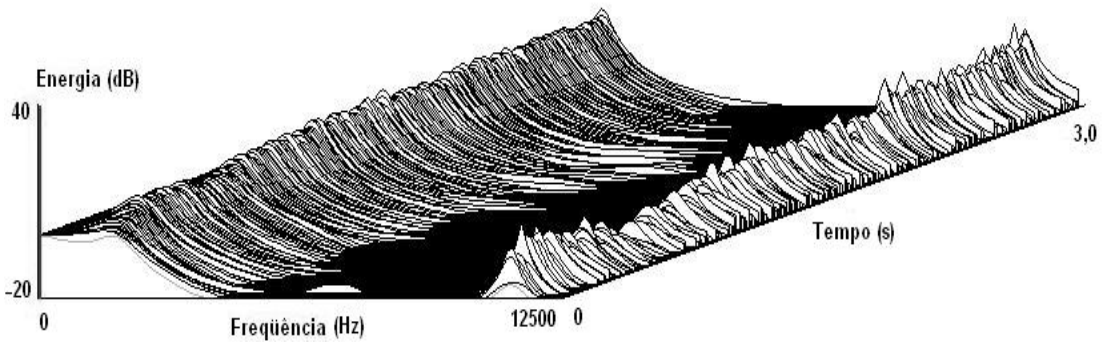



**Figura 3.21 – Espectro LPC para um sinal de voz normal.**

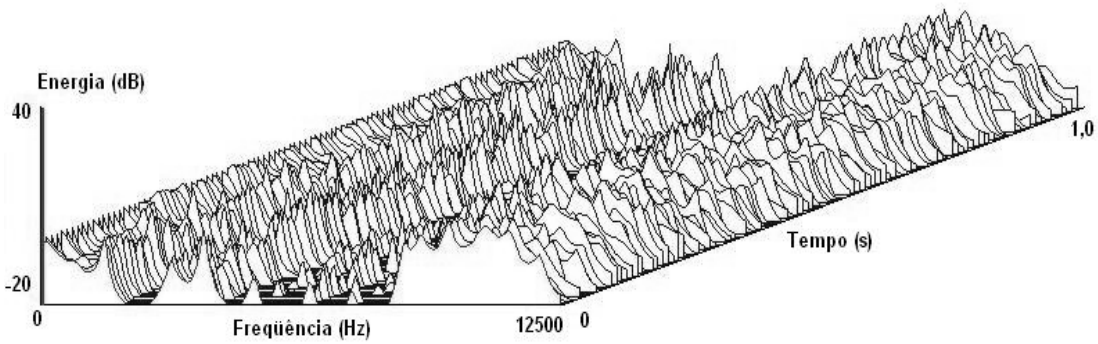



**Figura 3.22 – Espectro LPC para um sinal de voz com edema unilateral.**

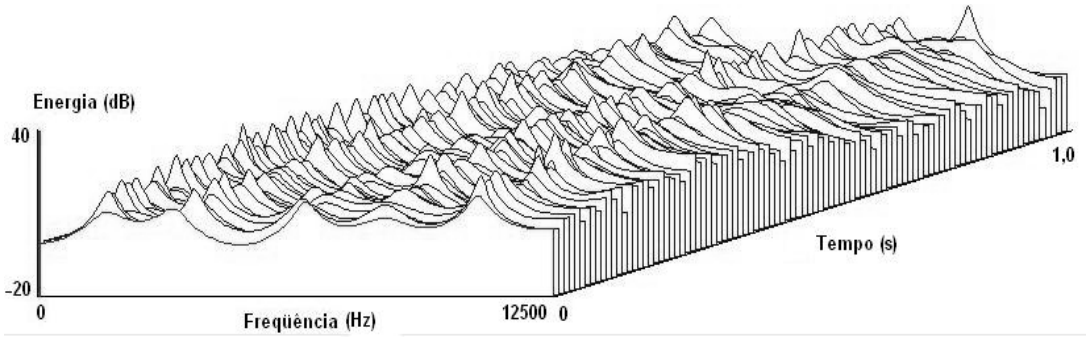



**Figura 3.23 – Espectro LPC para um sinal de voz com edema bilateral.**

Pode-se observar pelos espectros, a presença de componentes de alta energia nas regiões de freqüências mais altas do espectro para os sinais de vozes afetadas por edemas unilaterais nas dobras vocais. Há uma maior concentração de energia nas baixas freqüências no sinal normal, enquanto que no caso em que a patologia é mais severa (Figura 3.23), o espectro é mais denso, com componentes de alta energia por todo o espectro do sinal.

Dessa forma, torna-se interessante analisar os aspectos de um sinal patológico pelo comportamento também da análise LPC, dado que a variação de seu espectro em relação ao espectro da voz normal é evidente. Essas alterações sugerem uma melhor investigação do comportamento de sinais de vozes patológicas pela análise por predição linear.

Rabiner & Juang (1993) justificam, ainda, o uso da análise LPC pelas seguintes razões: 1) LPC fornece um bom modelo do sinal de voz, especialmente para o sinal de voz no estado quase estacionário das regiões sonoras da voz nas quais o modelo LPC de apenas pólos fornece uma boa aproximação da envoltória

espectral do trato vocal. Durante as regiões surdas ou transientes do sinal de voz, o modelo LPC é menos eficaz do que em regiões sonoras, mas ainda fornece um modelo aceitável para propósitos de reconhecimento; 2) Fornece uma separação razoável fonte-trato vocal; 3) O modelo LPC é analiticamente tratável, matematicamente simples e direto para implementar em *software* ou em *hardware*. A computação envolvida em processamento LPC é consideravelmente menor que para uma implementação toda digital do modelo de banco de filtros, por exemplo.

### 3.3.2 Análise Cepstral

A análise cepstral do sinal de voz para o estudo das alterações laríngeas pode ser muito útil, uma vez que permite se trabalhar com o sinal da glote (excitação) separadamente das repercussões ressonantais do trato vocal, facilitando o entendimento das modificações que ocorrem nas dobras vocais. A aplicação dessa técnica no estudo do sinal acústico de vozes alteradas poderia detectar modificações no sinal de voz que se relacionem com as alterações laríngeas e, conseqüentemente, identificar modelos para uma classificação, permitindo a obtenção de uma ferramenta de diagnóstico não-invasiva (ZWETSCH, 2006).

A qualidade da voz depende do modo de fechamento e abertura da glote e da vibração das dobras vocais. Certas alterações laríngeas impedem que dobras vocais tenham uma vibração glotal harmônica. Os principais fatores que determinam a vibração vocal são:

1. Posição da prega vocal, ou a extensão em que as dobras vocais são aduzidas ou abduzidas;
2. Mioelasticidade, ou o grau de elasticidade das dobras vocais (determinado pela posição e grau de tensão decorrente da contração do músculo vocal);
3. Nível de pressão do ar através das dobras vocais.

As alterações das dobras vocais também podem determinar que estas não vibrem em concordância, resultando em uma área vocal onde o trato vocal é excitado em duas freqüências fundamentais diferentes (ZWETSCH, 2006).

O uso da análise cepstral é direcionado para problemas centrados em voz sonora. No presente estudo, a voz obtida a partir da produção da vogal sustentada /a/ (som sonoro) permite avaliar o comportamento das dobras

vocais, com vibração obtida na produção desse tipo de som. De acordo com o modelo de produção de fala mais comumente empregado (Figura 3.18), a voz é considerada a saída de um sistema linear, variante no tempo (o trato vocal) excitado ou por um trem de pulsos quase-periódico (em sons sonoros) ou por ruído aleatório (em sons não-sonoros). Visto que o sinal de voz (Eq. 3.18) é o resultado da convolução da excitação ($ex(n)$) com a resposta do trato vocal ($\theta(n)$), seria útil separar ou "deconvoluir" as duas componentes.

$$s(n) = ex(n) * v(n). \tag{3.18}$$

Como na Eq. (3.17) os termos $ex(n)$ e $v(n)$ não são combinados linearmente, as técnicas lineares usualmente empregadas não podem ser utilizadas (DELLER, PROAKIS and HANSEN,1993; O'SHAUGHNESSY, 2000).

A deconvolução cepstral converte um produto de dois espectros na soma de dois sinais, que podem ser separados por um processo de filtragem linear, a "lifteragem" ou *liftering,* facilitando o estudo individualizado das modificações ocorridas na excitação e da parte ressonantal. Dentre as propriedades matemáticas envolvidas no processo, destacam-se, principalmente, as transformadas de Fourier e funções logarítmicas que resultam em uma função chamada cepstral ou cepstro, responsável pela dissociação do sinal de voz. A transformação desejada é logarítmica, na qual $\log(Ex(w).V(w)) = \log(Ex(w)) + \log(V(w))$, sendo $Ex(w)$ *e* $V(w)$ as transformadas de Fourier da forma de onda da excitação e da resposta do trato vocal, respectivamente (DELLER, PROAKIS E HANSEN, 1987et al, 1993; ZWETSCH, 2006; O'SHAUGHNESSY, 2000).

Assim, a análise cepstral do sinal de voz permite trabalhar com as componentes do sinal correspondentes à excitação e as respectivas modificações introduzidas pelo efeito de renossância no trato vocal, separadamente. O estudo das alterações na voz produzidas pelas dobras vocais é, então, facilitada devido às suas propriedades homomórficas, que permitem a separação das características do filtro do trato vocal da seqüência de excitação.

A análise cepstral pode ser aplicada no projeto de codificadores de voz do tipo *vocoders*, análise de formantes e detecção de freqüência fundamental (O'SHAUGHNESSY, 2000).

Na prática, o cepstrum complexo não é necessário, sendo suficiente o cepstrum real, definido como a transformada inversa do logaritmo do espectro de magnitude do sinal de voz (O'SHAUGHNESSY, 2000):

$$c(n) = \frac{1}{2\pi} \int_0^{2\pi} \log |X(e^{j\omega})| e^{j\omega n} d\omega. \tag{3.19}$$

Para sinais reais $x(n)$, $c(n)$ é a parte par de $\hat{x}(n)$, porque

$$\hat{X}(e^{j\omega}) = \log(X(e^{j\omega})) = \log |X(e^{j\omega})| + j \arg[X(e^{j\omega})]. \tag{3.20}$$

A magnitude da transformada de Fourier é real e par, enquanto a fase é imaginária e ímpar. A fase pode ser descartada, sem risco de degradação da qualidade do sinal de voz de saída (O'SHAUGHNESSY, 2000).

Para algoritmos de processamento digital de sinais, aplica-se a DFT (*Discrete Fourier Transform*), obtendo

$$c_d(n) = \frac{1}{N} \sum_{k=0}^{N-1} \log[X(k)] e^{j2\pi kn/N} \qquad n = 0, 1, ...., N\text{-}1. \tag{3.21}$$

A troca de $X(e^{j\omega})$ por $X(k)$ é equivalente à amostragem da transformada de Fourier (multiplicação por um trem de impulsos) em $N$ freqüências igualmente espaçadas de $\omega = 0$ a $2\pi$. O efeito é convoluir o sinal original $c(n)$ com um trem de amostras uniformes de período $N$, em que $c_d(n) = \sum_{i=-\infty}^{\infty} c(n + iN)$.

A concentração das componentes aparece no cepstro como picos. O eixo horizontal da função cepstral tem dimensões temporais e o nome de quefrências. Com isso, no cepstro da voz se obtém uma clara distinção entre a componente de excitação e a contribuição do trato vocal (ver exemplo na Figura 3.24).

Na Figura 3.24, vê-se o cepstro de um segmento de voz, em que o pico correspondente ao período fundamental (excitação) está próximo da quefrência de 10 ms, separado das componentes do trato vocal, que são as de baixas quefrências.


**Figura 3.24 – Cepstro de um segmento de fala (ZWETSCH, 2006).**

Na Figura 3.25 é apresentada a representação do cepstro para um sinal de voz normal e, na Figura 3.26, o cepstro para um sinal de voz patológico com edemas nas dobras vocais, ambos obtidos pelo *software Multi-Speech* 3700 (AGUIAR NETO et al, 2007b).

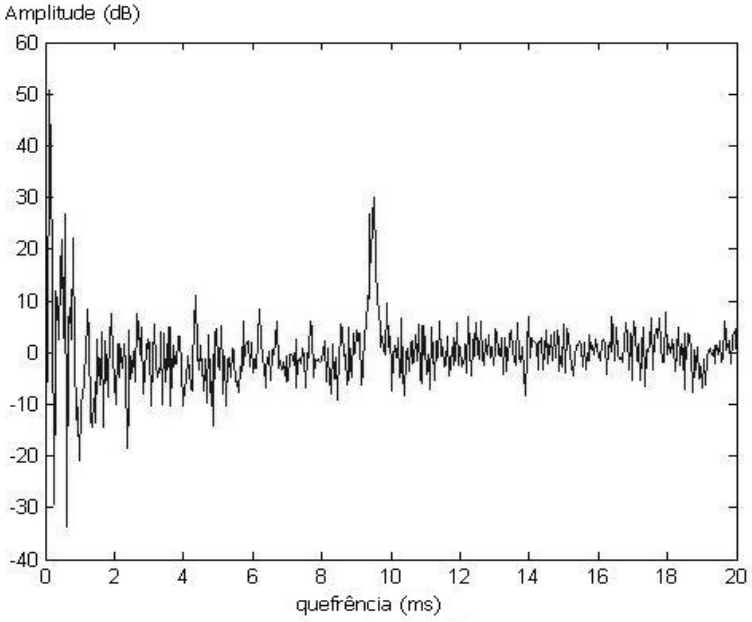



**Figura 3.25 – Cepstro para uma voz normal.**

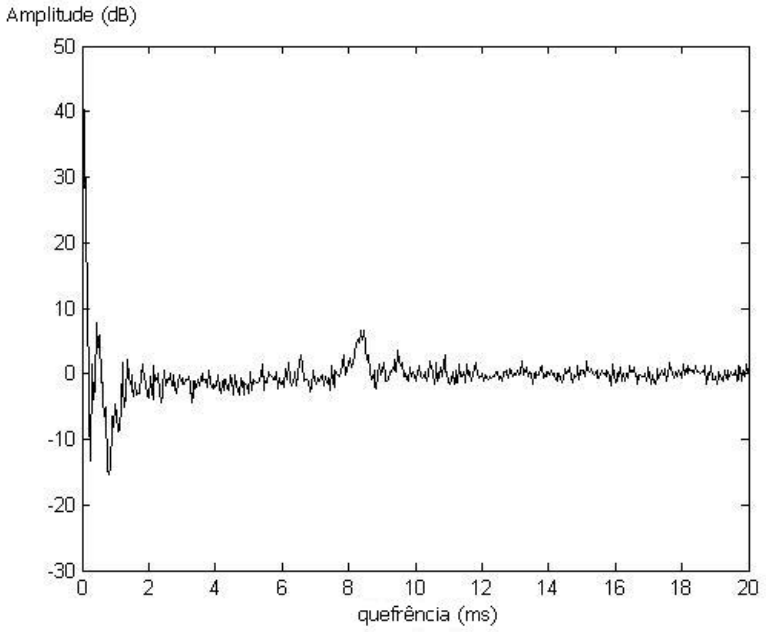



**Figura 3.26 – Cepstro para uma voz patológica.**

Os coeficientes cepstrais podem também ser obtidos a partir dos coeficientes LPC, mantendo a validade para análise dos efeitos das mudanças provocadas pelas dobras vocais no sinal de voz. Mantendo-se o trato vocal inalterado, ou seja, supondo que o trato vocal é saudável, as mudanças ocorridas no parâmetro, pelas alterações vocais, serão provenientes da excitação.

### 3.3.2.1 Coeficientes Cepstrais

Na análise cepstral LPC, a transformada z é aplicada ao sinal de voz modelado pela análise LPC. Os coeficientes cepstrais podem ser calculados recursivamente a partir dos coeficientes de predição linear, $\alpha(k)$, por meio de (MAMMONE et al, 1996; FECHINE, 2000):

$$\begin{cases} c(1) = -\alpha(1) \\ c_i(n) = -\alpha(n) - \sum_{j=1}^{n-1} (1 - \frac{j}{n}) \alpha(j) c(n-j) \quad 1 < n \leq p \end{cases} . \tag{3.22}$$

O uso dessa recursão permite um cálculo eficiente dos coeficientes cepstrais e evita fatoração polinomial. Os coeficientes cepstrais obtidos (Eq.

3.21) fornecem uma boa medida das diferenças na envoltória espectral dos segmentos de voz em análise (MAMMONE et al, 1996).

### 3.3.2.2 Coeficientes Delta Cepstrais

A representação cepstral do espectro da voz fornece uma boa representação das propriedades locais do sinal para um dado quadro em análise. Uma representação melhorada pode ser obtida estendendo a análise para incluir informação acerca da derivada cepstral temporal (tanto a primeira como a segunda derivada tem sido investigadas para melhorar o desempenho de sistemas de reconhecimento de voz). Para introduzir ordem temporal na representação cepstral, denota-se o i-ésimo coeficiente cepstral no tempo discreto $t$, por $c_i(t)$. Na prática, o tempo $t$ se refere ao quadro em análise, em vez de um instante de tempo arbitrário. A maneira pela qual a derivada cepstral no tempo é obtida é descrita a seguir: A derivada no tempo do espectro da log magnitude tem uma representação da série de Fourier da forma (RABINER & JUANG, 1993)

$$\frac{\partial}{\partial t}[\log | S(e^{jw}), t |] = \phi \sum_{k=-\infty}^{\infty} \frac{\partial c_i(t)}{\partial t} e^{jw_i}, \tag{3.23}$$

Em que $S(e^{jw})$ representa a densidade espectral do sinal de voz e $\phi$ uma constante de normalização.

Conseqüentemente, a derivada temporal cepstral deve ser determinada de uma maneira adequada. Sabe-se que se $c_i(t)$ é uma representação discreta no tempo (em que t é o índice do quadro), simplesmente usar uma diferença de primeira ou segunda ordem para aproximar a derivada é inadequado. Um compromisso razoável é aproximar $\frac{\partial c_i(t)}{\partial t}$ por um polinômio ortogonal apropriado (uma estimativa dos mínimos quadrados da derivada) sobre uma janela de comprimento finito, isto é,

$$\frac{\partial c(n,t)}{\partial t} = \Delta c_i(n) \approx \phi \sum_{k=-K}^{K} kc(n,\ t+k), \tag{3.24}$$

em que $c(n,\ t)$ é o $n$-ésimo coeficiente da predição linear no tempo *t*, *Φ* é a constante de normalização e $2K+1$ é o número de quadros sobre os quais o cálculo é realizado. Neste trabalho, os coeficientes delta cepstrais são obtidos como uma versão simplificada da Eq. (3.24), a partir de

$$\Delta c_i(n) = [\sum_{q=-K}^{K} k c_{i-q}(n)]G, \quad 1 \le n \le p, \tag{3.25}$$

em que $G$ é um termo de ganho (por exemplo: 0.375), $p$ é o número dos coeficientes delta cepstrais, $\phi$, K é o número de coeficientes, $n$ representa o índice do coeficiente e $i$ o quadro em análise (MAMMONE et al,1996; FECHINE, 2000).

### 3.3.2.3 Coeficientes Cepstrais Ponderados

Coeficientes ponderados são o resultado da ponderação dos coeficientes cepstrais com o objetivo de minimizar a sensibilidade dos coeficientes cepstrais de baixa ordem em relação à envoltória espectral e à sensibilidade dos coeficientes cepstrais de alta ordem em relação ao ruído. A ponderação é realizada multiplicando os coeficientes cepstrais, $c_{\mathrm{i}}(n),$ por uma janela $w(n)$, usando cepstrum ponderado, $cw_{\mathrm{i}}(n)$, como um vetor de características (Eq. 3.26). A escolha adequada da janela melhora a robustez dos coeficientes. A operação de ponderação é também conhecida como *liftering* ou suavização (MAMMONE et al, 1996). Os coeficientes cepstrais ponderados são obtidos por

$$cw_i(n) = c_i(n) \cdot w(n). \tag{3.26}$$

Há várias formas de ponderação que diferem entre si pelo tipo de janela cepstral, $w(n),$ utilizada. A mais simples é a janela retangular, dada por

$$w_r(n) = \begin{cases} 1, & n = 1,2,...L \\ 0, & \text{caso contrário.} \end{cases} \tag{3.27}$$

em que *L* é o tamanho da janela. As primeiras *L* amostras, que são as mais significativas devido à propriedade do decaimento, são mantidas (MAMMONE et al, 1996).

Outra forma de ponderação é baseada na técnica de filtragem linear ou *quefrency liftering*, que pondera cada componente individual pelo índice *n*, suavizando as componentes de ordem inferior, dada por

$$w_{fl}(n) = \begin{cases} n, & n = 1,2,...L \\ 0, & \text{caso contrário.} \end{cases} \tag{3.28}$$

E a filtragem linear *(liftering*) passa-faixa (*Bandpass liftering – BPL*), método utilizado neste trabalho (MAMMONE et al, 1996):

$$w_{bpl}(n) = \begin{cases} 1 + \frac{L}{2}\text{sen}(\frac{n\pi}{L}), & n=1,2,...,L \\ 0, & \text{caso contrário.} \end{cases}, \tag{3.29}$$

em que *L* é o tamanho da janela. A técnica baseada em BPL pondera uma seqüência cepstral pela janela da Eq. (3.29) tal que as componentes de ordem mais alta e de ordem mais baixa são de-enfatizadas, sendo o tipo de janela utilizada neste trabalho.

### 3.3.2.4 Coeficientes Delta-Cepstrais Ponderados

Os coeficientes delta-cepstrais são obtidos a partir das Eqs. (3.25) e (3.26), associando as características dos coeficientes cepstrais ponderados com os delta-cepstrais, resultando em

$$\Delta cw_i(n) = \Delta c_i(n) \cdot w_{bpl}(n). \tag{3.30}$$

#### 3.3.2.5 Coeficientes Mel-cepstrais

Nos anos 80, o cepstro tornou-se uma característica importante na modelagem de sinais de voz em sistemas de reconhecimento de voz como uma forma de melhorar as taxas de reconhecimento. Os processos de filtragem linear e ponderação que suavizam o espectro baseado em LP (predição linear), removendo a variabilidade inerente, devido à excitação, melhora, aparentemente, o desempenho do reconhecimento de fala (DELLER, PROAKIS e HANSEN , 1987).

Uma segunda forma de melhoramento no desempenho do reconhecimento pode ser obtida pelo cepstro baseado na escala mel, ou simplesmente mel-cepstro.

Os coeficientes mel-cepstrais (*Mel-frequency Cepstral Coefficients* – MFCC) surgiram devido aos estudos na área de psicoacústica (ciência que estuda a percepção auditiva humana), que mostraram que a percepção humana das freqüências de tons puros ou de sinais de voz não segue uma escala linear. Isso estimulou a idéia de serem definidas freqüências subjetivas de tons puros, da seguinte forma: para cada tom com freqüência *f*, medida em *Hz*, define-se um tom subjetivo medido em uma escala que se chama escala mel. O mel, então, é uma unidade de medida da freqüência percebida de um tom (DELLER, PROAKIS e HANSEN, 1987).

A diferença entre o cálculo dos coeficientes cepstrais e dos coeficientes mel-cepstrais está na aplicação de um banco de filtros digitais ao espectro real do sinal, antes da aplicação da função logarítmica. Tais filtros não estão linearmente espaçados no domínio da freqüência. Esses filtros têm por objetivo aproximar a resposta humana a sinais sonoros.

É possível traçar uma comparação entre a freqüência real (medida em Hz) e a freqüência percebida (medida em mels). O mapeamento entre a escala de freqüência real, em Hz, e a escala de freqüências percebida, em mel, é aproximadamente linear abaixo de 1000 Hz e, logarítmica, acima. Logo, o espaçamento dos filtros digitais deve respeitar a escala de freqüências percebidas (escala Mel). Pode-se definir uma função para mapeamento da freqüência acústica *f* (em $\mathrm{Hz}$) para uma escala de freqüências percebidas Mel (em mels) como

$$F_{mel} = 2595 \cdot \log_{10}(1 + \frac{F_{linear}(Hz)}{700}), \quad (3.31)$$

em que $F_{linear}$ é a freqüência linear (em Hz) e $F_{mel}$ é a freqüência percebida (em mel).

Para obtenção dos coeficientes mel-cepstrais (MFCC), a partir dos coeficientes cepstrais, deve-se a seguir aplicar filtros digitais espaçados segundo uma escala acusticamente definida (escala mel). Após o mapeamento das freqüências acústicas para a escala de freqüências percebidas, pela Equação (3.31), é aplicado um banco de filtros espaçados linearmente no domínio mel. Isso corresponde à aplicação de filtros digitais espaçados segundo a escala mel, no domínio da freqüência (ver exemplo na Figura 3.27).

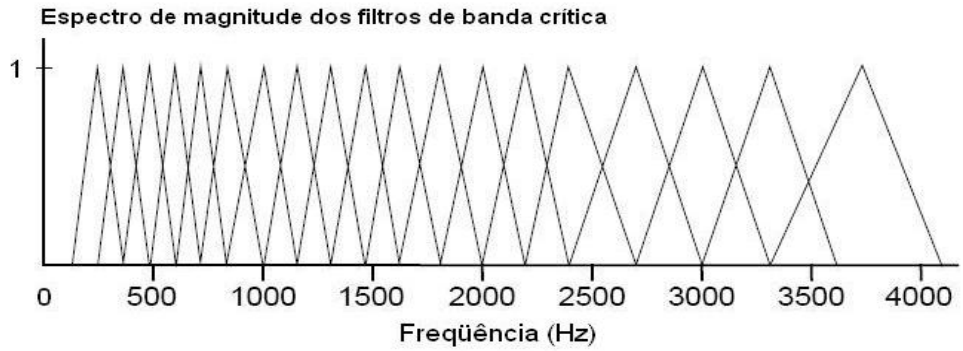



**Figura 3.27 – Banco de filtros digitais na escala mel.**
FONTE: (O'SHAUGHNESSY, 2000)

Na Figura 3.28 é ilustrado o processo de obtenção dos coeficientes mel-cepstrais (DIAS, 2000).

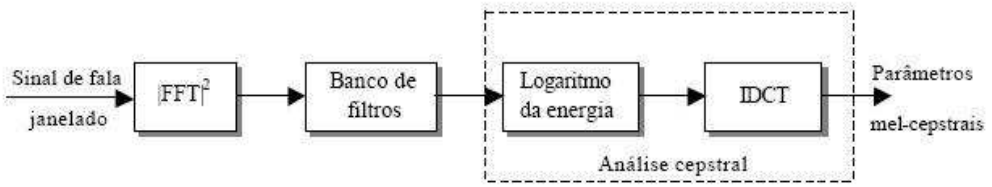



**Figura 3.28 – Processo de obtenção dos coeficientes mel-cepstrais.**
FONTE: (DIAS, 2000)

O sinal de voz é pré-processado (segmentação, janelamento e pré-ênfase). Os coeficientes mel cepstrais são obtidos a partir de cada janela do sinal, depois de realizados os seguintes processamentos (ANDREÃO, 2001):

- Cálculo do espectro de magnitude do sinal ($|FFT|^2$);

- Aplicação do banco de filtros triangulares em escala mel;
- Cálculo do logaritmo da energia de saída de cada filtro. A aplicação do logaritmo é necessária para a obtenção do cepstro. São utilizados geralmente 20 filtros de formato triangular. No entanto, a quantidade de filtros é baseada na freqüência de amostragem ($F_a$) (3.ln($F_a$)).
- Finalmente, o processo de obtenção dos coeficientes MFCC pode ser matematicamente descrito por (O'SHAUGNESSY, 2000; GODINO-LLORENTE, 2006; ANDREÃO, 2001)

$$c_{mel}(n) = \sum_{k=1}^{Nf} \log(Sf(k)).\cos[n(k-\frac{1}{2})].\frac{\pi}{Nf} \qquad n = 0,1,....,Nf \tag{3.32}$$

em que $Nf$ é o número de filtros digitais utilizados, $c_{mel}(n)$ é o $n$-ésimo coeficiente mel-cepstral e $Sf(k)$ é o sinal de saída do banco de filtros digitais, dado por

$$Sf(k) = \sum_{j=1}^{NFFT} W_k(j).X(j) \quad k = 1,..., Nf, \tag{3.33}$$

em que $W_k(j)$ são as janelas de ponderação triangulares associadas às escalas-mel e $X(j)$ é o espectro de magnitude da FFT de $N$ pontos (O'SHAUGNESSY, 2000; GODINO-LLORENTE, 2006).

## 3.4 Discussão

A partir das informações deste capítulo, observa-se que a análise acústica dos parâmetros e características apresentados é importante para um acompanhamento terapêutico de desordens vocais, que pode também ser utilizada como uma ferramenta de auxílio a pré-diagnósticos de patologias nas dobras vocais.

Patologias nas dobras vocais afetam o *pitch* de tal forma que sua determinação fica difícil e, em alguns casos, impossível, o que torna comprometida a análise acústica utilizando medidas que dependam da obtenção do *pitch*. Além disso, o grande número de parâmetros necessário nesse tipo de análise a torna difícil e dispendiosa.

Neste trabalho, propõe-se o uso da análise cepstral como uma ferramenta opcional para analisar desordens vocais provocadas por patologias nas dobras vocais.

As técnicas de análise da voz por predição linear e análise cepstral foram descritas neste capítulo, como embasamento teórico para a descrição do processo de caracterização e modelagem acústica dos sinais de vozes patológicas.

O uso das referidas técnicas justifica-se pelas características que cada uma oferece no sentido de representarem parametricamente o sinal de voz. Assim, em uma análise comparativa entre os parâmetros obtidos para voz normal e para vozes patológicas, é possível perceber as alterações provocadas por uma patologia no sinal de voz.

A observação visual dos gráficos representativos dos sinais no domínio do tempo, assim como o espectro LPC, formantes, energia e *pitch*, sugerem anormalidades no sinal de voz patológico. No entanto, há que considerar que se a análise for feita visualmente torna-se, ainda, uma avaliação subjetiva, dependendo da experiência e qualificação do profissional responsável pela avaliação da qualidade vocal ou do diagnóstico da patologia que causa a desordem vocal.

Portanto, é interessante que, uma boa avaliação subjetiva possa ser acompanhada de uma avaliação objetiva, que proporcione um diagnóstico eficiente e confiável de uma patologia. O processo de discriminação entre uma voz normal e uma voz patológica será tanto mais eficiente quanto mais exato for o método empregado para análise acústica do sinal de voz.

Dessa forma, torna-se possível a discriminação entre uma voz normal e uma voz patológica a partir do emprego de classificadores que explorem as alterações provocadas nos parâmetros representativos da fala.

No caso dos coeficientes cepstrais, por exemplo, propõe-se representar as variações da excitação imposta por uma patologia laríngea, foco deste estudo.

Para uma melhor fundamentação da análise acústica do sinal por meio da análise LPC e cepstral, leva-se a efeito um processo de classificação/discriminação de uma dada patologia em relação a sinais de vozes normais. As técnicas de classificação utilizadas neste trabalho são descritas no Capítulo 4 e os resultados obtidos no Capítulo 5.

# Capítulo 4

# Técnicas para Classificação de Sinais de Vozes Patológicas

## 4.1 Introdução

Técnicas de processamento digital de sinais de voz têm sido cada vez mais propostas para o desenvolvimento de ferramentas aplicadas à terapia vocal (GAVIDIA-CEBALLOS & HANSEN, 1996; YIN & CHIU, 2004; GARCIA et al, 2005; GODINO-LLORENTE et al, 2006; UMAPATHY et al, 2005;FONSECA et al, 2005; AGUIAR et al, 2007a). Algumas dessas técnicas são usuais em sistemas de reconhecimento de fala e/ou de identidade vocal (FREDOUILLE et al, 2005).

Em discriminação de vozes patológicas, ou deteção de patologias que provocam alterações na fala, busca-se uma ferramenta não-invasiva que possa auxiliar o profissional da área médica, não só na terapia vocal para melhorar a qualidade da voz, como também para pré-diagnósticos, acompanhamentos pós-cirúrgicos e farmacológicos. Usualmente, são feitos exames laringoscópicos que utilizam equipamentos endoscópicos com microcâmeras, que muitos pacientes consideram desconfortáveis.

Neste capítulo, serão abordadas a técnica de classificação utilizada para discriminar vozes patológicas com edema nas dobras vocais, a saber, o uso de Quantização Vetorial (QV) e Modelos de Markov Escondidos (*Hidden Markov Models* - HMMs). Antes da etapa de classificação, foi feito um pré-processamento no sinal e a extração de parâmetros representativos do modelo. A partir da quantização vetorial foi levada a efeito uma redução da dimensionalidade dos vetores de parâmetros. Em seguida, foi realizada uma classificação dos sinais baseada numa medida de distorção (denominada aqui de pré-classificação). Um classificador baseado em HMMs é usado para refinar o processo de pré-classificação, a fim de melhorar os resultados obtidos. O embasamento teórico necessário sobre HMM e Quantização Vetorial e o procedimento utilizado neste trabalho, são descritos neste capítulo e os resultados obtidos são abordados no Capítulo 5.

## 4.2 O Processo de Classificação de Vozes Patológicas

A tarefa de discriminação de voz patológica é uma questão de reconhecimento de padrões. O diagrama em blocos da Figura 4.1 mostra a representação do procedimento geral utilizado para o processo de classificação do sinal de voz em análise, baseado no modelo utilizado para reconhecimento de padrões (DIAS, 2000; adaptado de RABINER & SCHAFER, 1978).

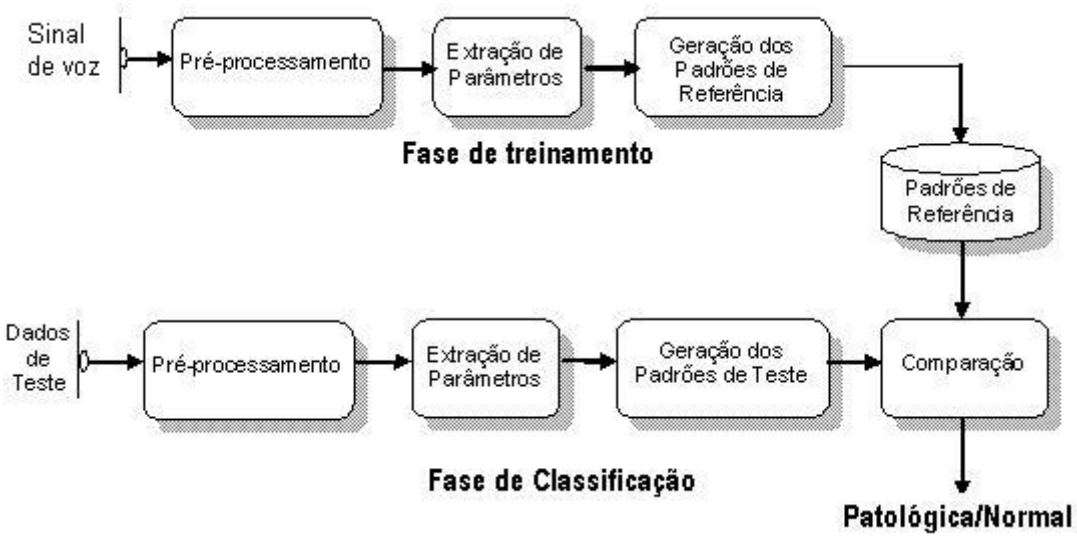



**Figura 4.1 – Diagrama em blocos para o procedimento de discriminação do sinal de voz (normal/patológica).**

### 4.2.1 Pré-processamento

A etapa de pré-processamento consiste em (RABINER and SCHAFER, 1978):

- ***segmentação do sinal em quadros e janelamento***

  - Divisão do sinal em quadros correspondentes a intervalos de tempo de 16 a 32 ms de forma a assegurar a sua estacionaridade;
  - Ponderação do quadro do sinal por função que proporciona a manutenção das características espectrais do centro do quadro e a eliminação das

transições abruptas das extremidades. Para um janelamento de *Hamming* a função é

$$w_h(n) = \begin{cases} 0,54 - 0,46\cos[2\pi n(N_A - 1)], & 0 \le n \le N_A - 1 \\ 0, & \text{caso contrário} \end{cases} \quad , \quad (4.1)$$

em que $n$ é a amostra e $N_A$ é *o* número de amostras da janela em análise.

Considerando-se janelas *de* 20ms com superposição de 50%, os parâmetros do sinal de fala são atualizados a cada 10 ms. A superposição proporciona a suavização da amplitude do sinal amostrado, nos extremos do segmento de análise, dando maior ênfase às amostras localizadas no centro da janela (Figura 4.2) (DIAS, 2000).

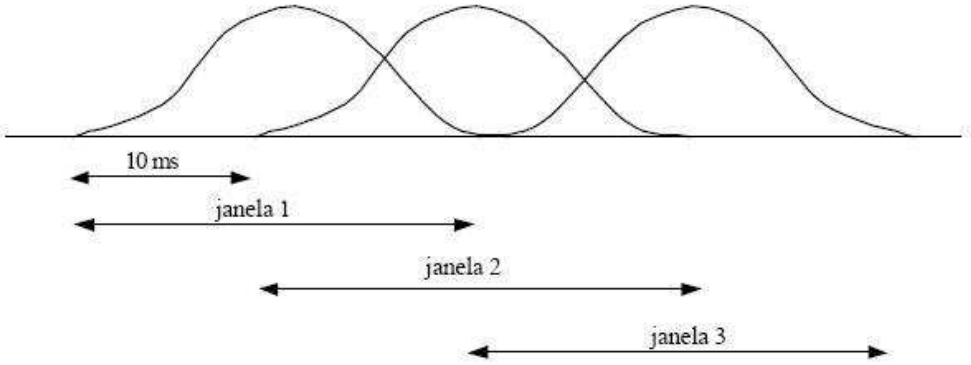



**Figura 4.2 – Processo de janelamento do sinal, com superposição de quadros.**

- **pré-ênfase**

  – proporciona compensação das perdas durante a passagem do sinal pelo trato vocal e pela radiação nos lábios. Para solucionar esse problema é aplicado um filtro, de resposta de aproximadamente +6dB/oitava. A função de transferência da pré-ênfase consiste de um sistema de primeira ordem fixo, cuja função é

$$H_p(z) = 1 - \alpha \cdot z^{-1}, \quad 0 \le \alpha \le 1 , \quad (4.2)$$

em que $\alpha$ é o fator de pré-ênfase (valor típico usado: $\alpha = 0,95$).

O sinal resultante da $s_{\mathrm{p}}(n)$ está relacionado à entrada $s(n)$ pela equação diferença

$$s_p(n) = s(n) - \alpha \cdot s(n-1), \tag{4.3}$$

sendo $s_p(n)$ a amostra pré-enfatizada, $s(n)$ a amostra original.

### 4.2.2 Extração de Parâmetros

Na etapa de extração dos parâmetros, é feita a aquisição dos parâmetros para análise acústica. No presente estudo, foram extraídos os coeficientes LPC e os coeficientes cepstrais e mel-cepstrais (descritos no Capítulo 3).

### 4.2.3 Geração de Padrões

Para efeito da classificação, a geração dos padrões pode ser feita por meio de uma abordagem apenas paramétrica (exemplo: Quantização Vetorial) ou estatística (exemplo: Modelos de Markov Escondidos) (RABINER and SCHAFER, 1978; FECHINE, 2000). Nos métodos paramétricos, após a detecção de fim de palavra é levada a efeito uma redução de dados explícita, após a qual é obtido um padrão de referência que continua ainda na forma paramétrica. A regra de decisão no processo de comparação de padrões baseia-se em medidas de distância.

Nos métodos estatísticos, a construção dos padrões é obtida por meio de modelos estatísticos, tais como Modelos de Markov Escondidos (HMMs). Os parâmetros extraídos são, portanto, com o auxílio da teoria das probabilidades, representados por modelos estocásticos. Nesses métodos não é feita uma comparação direta de padrões e a decisão é tomada usando o cálculo de probabilidades associadas aos modelos (FECHINE, 2000).

O processo se repete para o sinal de teste, para o qual é feito o pré-processamento, a extração de parâmetros e a geração de padrões de teste que serão comparados com os padrões de referência pré-armazenados na fase de treinamento.

Finalmente, os padrões são comparados e será dado o resultado final: a voz de teste é normal ou patológica, segundo algum critério de decisão.

Reconhecedores de fala baseados em HMMs têm sido de grande interesse devido ao seu baixo custo computacional, durante a fase de reconhecimento (visto que se baseia apenas no cálculo de uma medida de probabilidade) e por basear-se em modelos estocásticos do sinal de voz sendo capaz de modelar vários eventos, tais como fonemas, sílabas, etc. (RABINER et al, 1985), o que o torna bastante flexível.

O sistema para discriminação de vozes patológicas, neste trabalho, se constitui em um sistema híbrido, que utiliza tanto o método paramétrico quanto o estatístico, para as fases de treinamento e classificação, visando a otimização do processo.

Para a tarefa de treinamento, após a extração e escolha dos parâmetros que melhor representam o sinal patológico, é realizada a quantização vetorial destes parâmetros e obtido o dicionário representativo do sinal de entrada correspondente, sendo um para cada sinal. Para o projeto do dicionário do quantizador vetorial foi utilizado o algoritmo LBG (LINDE et al, 1980). Em seguida, são construídos os Modelos de Markov Escondidos (HMMs) de Densidades Discretas, sendo associado um HMM para cada tipo de sinal (normal ou patológico). Na tarefa de classificação, são utilizados dois parâmetros para discriminação do sinal de entrada: a medida de distorção obtida a partir da quantização vetorial, seguida da probabilidade obtida do HMM. Esse último é utilizado como parâmetro de refinamento do processo de discriminação.

A seguir, são descritas a técnica paramétrica e a técnica estatística utilizadas: Quantização Vetorial e Modelos de Markov Escondidos.

## 4.3 Quantização Vetorial

A técnica da quantização vetorial foi utilizada, na pesquisa, tanto para efeito de redução da dimensão dos dados quanto para uma etapa de classificação dos sinais em normais ou patológicos. Para tanto, a partir da quantização vetorial, foi gerado um dicionário (*codebook*) para cada sinal da base de dados

para determinação da similaridade entre as elocuções dos sinais a serem analisados.

Neste trabalho, a Quantização Vetorial (MAKHOUL et al, 1985) é realizada para cada um dos parâmetros (LPC, cepstral, delta-cepstral, mel-cepstral, etc), usando vozes com edemas nas dobras vocais, ou seja, o treinamento é feito com sinais de vozes patológicas. A base de dados é descrita em detalhes no Capítulo 5.

Dessa forma, classificadores diferentes treinados a partir de uma medida de distância aplicada na Quantização Vetorial, são obtidos para o processo de discriminação.

Após a extração de características, o dicionário é gerado, consistindo da geração de $N$ níveis discretos que cada vetor de parâmetros representativos do sinal de entrada pode assumir.

Assim, um quantizador de $N$ níveis pode ser definido como um mapeamento $Q$ de um espaço Euclidiano $K$-dimensional $R^K$ (Figura 4.3) dentro de um subconjunto $Wqv$ de $R^K$. Assim,

$$Q : R^K \rightarrow Wqv. \tag{4.4}$$

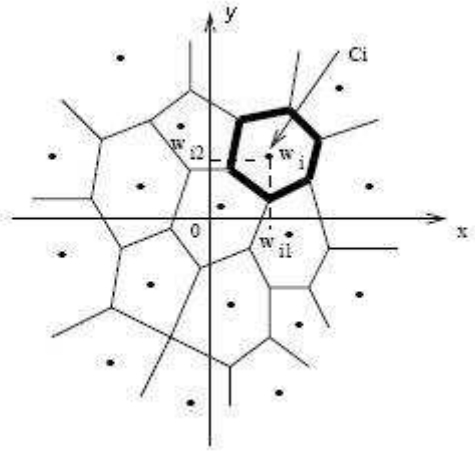



**Figura 4.3 - Partição do espaço bi-dimensional (K = 2).**

O dicionário $Wqv = \{w_i ; i=1, 2, \ldots N\}$ é o conjunto de vetores códigos, $K$ é a dimensão do quantizador e $N$ é o número de vetores códigos em $Wqv$.

O mapeamento $Q$ assume para um vetor de entrada $\mathbf{x}$ de valor real $K$-dimensional um vetor código $K$-dimensional $w_i=Q(x)$.

A Quantização Vetorial define um particionamento do espaço Euclidiano $K$-dimensional dentro de células sem interceptação, com $Nq$ níveis, tal que

$$Sv_i = \{\boldsymbol{x}: Q(x) = w_i\}, \qquad i = 1, 2, \ldots, Nq. \tag{4.5}$$

Como as células de Voronoi, $Sv_i$, coletam juntamente todos os vetores de entrada mapeando-os para o $i$-ésimo vetor código $\boldsymbol{w_i}$, este pode ser visto como o padrão de entrada pertencente a $Sv_i$.

O mapeamento do vetor de entrada $\boldsymbol{x}$ para um vetor código $w_i$ ocorre se

$$d(\boldsymbol{x}, w_I) < d(\boldsymbol{x}, w_i), \ \forall i \neq I, \tag{4.6}$$

em que $d(.)$ é uma função de distorção. É utilizada a regra do vizinho mais próximo para encontrar aquele que apresenta maior similaridade a $\boldsymbol{x}$. Neste trabalho, foi utilizado o algoritmo LBG e a distância do erro médio quadrático mínimo (LINDE et al, 1980; AGUIAR NETO et al, 2007a; COSTA et al, 2007b).

## 4.4 Modelos de Markov Escondidos (*Hidden Markov Models* – HMMs)

Um problema de fundamental importância no estudo dos sinais é encontrar um modelo adequado que caracterize bem o seu comportamento. Por meio de modelos estatísticos, tenta-se caracterizar as propriedades estatísticas do sinal, como é o caso dos processos gaussianos, processos de Poisson, processos de Markov ou Modelos de Markov Escondidos, dentre outros. A hipótese levantada pelos modelos estatísticos é que o sinal pode ser bem caracterizado como um processo aleatório paramétrico, cujos parâmetros podem ser estimados de forma coerente (RABINER, 1989).

Nas tarefas de reconhecimento de padrões, os modelos de sinais estocásticos e, em especial, os Modelos de Markov Escondidos (HMMs) têm apresentado excelente resultados. Como neste trabalho o problema da discriminação de vozes patológicas recai na área de reconhecimento de padrões, os Modelos de Markov Escondidos se mostram adequados.

Na Figura 4.4 é apresentado o diagrama em blocos do procedimento geral a ser utilizado neste trabalho. Após a aquisição do sinal e a etapa de pré-processamento, os parâmetros de interesse (LPC e cepstrais) são extraídos e, em seguida, é aplicada a quantização vetorial. Os parâmetros quantizados são utilizados para o treinamento dos modelos de Markov e a seguir é feita a etapa de classificação. Lembrando que (Figura 4.1) na etapa de treinamento são extraídos os parâmetros que servirão como base de dados para comparação com os padrões de teste.

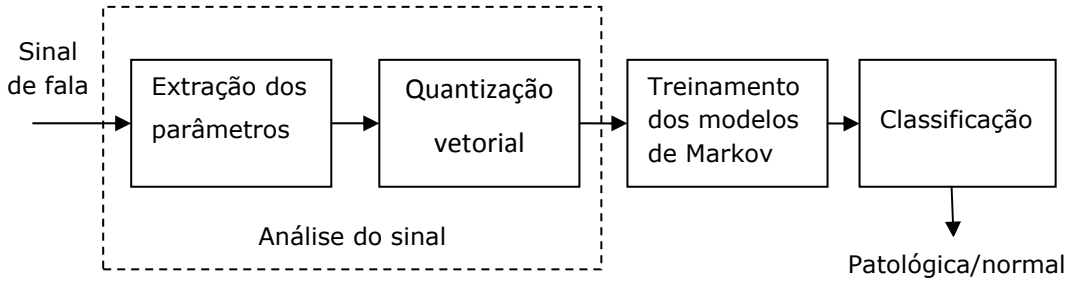



**Figura 4.4 – Diagrama em blocos do processo de classificação de vozes patológicas**

Os Modelos de Markov Escondidos (HMMs) apresentam um baixo custo computacional por ser necessário, na fase de reconhecimento dos padrões, apenas o cálculo de uma medida de probabilidade. Diferentemente dos métodos paramétricos que envolvem, durante a fase de reconhecimento, o cálculo de medidas de distância, o que acarreta um maior tempo de processamento. São bastante flexíveis por basear-se em modelos estocásticos do sinal de voz, sendo capazes de modelar métodos para o reconhecimento automático de padrões.

Dentre as vantagens do uso de HMMs podem ser citadas: A habilidade para treinar vários exemplos; os parâmetros do modelo são automaticamente agrupados para representar as entradas; as características temporais do sinal de entrada são modeladas inerentemente; as variações estatísticas do sinal de entrada são consideradas por estarem implícitas na própria formulação probabilística; não é necessária, a priori, uma distribuição estatística das entradas para estimação dos parâmetros, o que não é o caso, usualmente, em outras técnicas estatísticas (SATISH & GURURAJ, 1993; 2000; FECHINE, 2000).

Sabe-se que o som produzido por um dado locutor é uma função das características fisiológicas do trato vocal, tais como tamanho da garganta e

posição e forma de elementos do sistema articulatório dentre outros fatores. Esses fatores interagem entre si para produzir o som de uma elocução de acordo com as características inerentes de quem fala. Um sistema de reconhecimento de locutor tem como objetivo identificar essas características inerentes de modo a associar uma identidade vocal ao locutor. Por outro lado, no caso da identificação de vozes patológicas não importam as características inerentes do locutor. O interesse é verificar, a partir de um conjunto de elocuções de diferentes locutores, se há alterações intrínsecas à patologia nas características da voz, em relação às de uma voz normal. Assim, espera-se que os Modelos de Markov Escondidos, semelhantemente ao que ocorre em Sistemas de Reconhecimento de Locutor e Reconhecimento de Voz, possam modelar as características intrínsecas do sinal patológico por observação das modificações introduzidas neste com relação ao sinal de voz normal.

Considerando a produção interna da voz como sendo uma seqüência de estados escondidos, e o som resultante uma seqüência de estados observáveis gerados por uma voz processada que mais se aproxima do estado verdadeiro (escondido) (CAMPBEL, 1997), os Modelos de Markov Escondidos podem ser aplicados ao processo de classificação de vozes patológicas em relação a vozes normais. É preciso escolher qual o tipo de HMM, ou seja, a estrutura do modelo mais adequado para a aplicação pretendida.

### 4.4.1 Tipos de HMMs e Descrição do Modelo

É comum considerar três tipos de estruturas para os modelos de HMMs: o modelo sem restrição (o ergódico), o modelo serial restrito e o paralelo restrito (RABINER, LEVINSON and SONDHI, 1983).

No modelo sem restrição pode ocorrer mudança de um estado para outro qualquer. Neste trabalho será utilizado um modelo simplificado de HMM conhecido como modelo *left-right* (serial restrito), ou modelo de Bakis (LEVINSON, RABINER and SONDHI, 1983). Nesse modelo, exemplificado na Figura 4.5 (DIAS, 2000), são permitidos apenas transições para o mesmo estado, ou transições de um estado $i$ para um estado $j$, mais à direita (j ≥i), em que $a_{ij} = 0$ se $j > i + 2$.

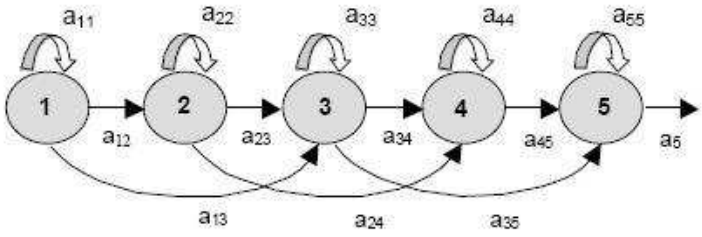



**Figura 4.5 - Exemplo de um HMM tipo *left-right* de cinco estados.**
FONTE: (DIAS, 2000)

Como a estrutura da fala é inerentemente seqüencial, a liberdade adicional de transição de estados presente nos modelos sem restrição não refletem as variações dos parâmetros da fala caracterizados por vetores de padrões. Conseqüentemente, o uso de modelos com restrição torna-se preferível (RABINER, LEVINSON and SONDHI, 1983).

Os modelos esquerda-direita (*left-right*), para os Modelos de Markov Escondidos ou para uma cadeia de Markov, apresentam as seguintes propriedades (LEVINSON, RABINER and SONDHI, 1983):

1. A primeira observação é produzida quando a cadeia de Markov encontra-se em um estado determinado, chamado estado inicial, designado por $q_1$.
2. A última observação é gerada enquanto a cadeia de Markov está em outro estado, chamado estado final ou estado de absorção, designado por $q_N$.
3. Uma vez que em uma cadeia de Markov se deixa um estado, aquele estado não pode ser visitado num tempo posterior.

Um HMM pode ser definido como sendo (DIAS, 2000):

- Um conjunto de estados $\{S_j\}$, incluindo um estado inicial $S_i$ e um estado final $S_f$ ;
- Uma matriz de transições $A=\{a_{ij}\}$, em que $a_{ij}$ representa a probabilidade de se efetuar uma transição do estado $i$ para o estado $j$;
- Uma matriz de probabilidades de saída $B=\{b_j(k)\}$, em que $bj(k)$ define a probabilidade de emissão do símbolo $k$, ao se chegar ao estado $j$

(modelo de Moore). O símbolo $k$ pertence a um conjunto finito ou infinito de símbolos de saída.

Desde que $a_{ij}$ e $bj(k)$ sejam probabilísticos, as seguintes propriedades devem ser satisfeitas

$$\sum_j a_{ij} = 1 \quad \forall\ i, j,\ \text{sendo} \quad a_{ij} \geq 0, \tag{4.7}$$

$$\sum_k b_j(k) = 1 \qquad \forall\ k,\ \text{sendo}\ b_j(k) \geq 0. \tag{4.8}$$

A propriedade fundamental de todos os HMMs do tipo *left-right* é que os coeficientes $a_{ij}$ da matriz de transição de estados, **A**, com $N$ estados, obedecem à seguinte propriedade:

$$a_{ij} = 0, \qquad \text{para}\, j < i \qquad 1 \leq i, j \leq Ne, \tag{4.9}$$

em que $Ne$ representa o número de estados do HMM.

Para modelos *left-right,* restrições adicionais são freqüentemente colocadas sobre os coeficientes da matriz de transição de estados, para assegurar que não ocorram mudanças muito grandes nos índices dos estados, tais como

$$a_{ij} = 0, \qquad \text{para}\, j > i + \Delta \qquad \Delta = 1,2,\ldots \ , \tag{4.10}$$

em que $\Delta$ é um incremento no índice do estado, freqüentemente utilizado (RABINER, 1989).

Além disso, as probabilidades de estado inicial apresentam a propriedade

$$\pi_i = \begin{cases} 0, & \text{para}\ i \neq 1 \\ 1, & \text{para}\ i = 1 \end{cases} \tag{4.11}$$

### 4.4.2 Parâmetros do Modelo

Os parâmetros que caracterizam o HMM, $\lambda = (\mathrm{A}, \mathrm{B}, \Pi)$, da Figura 4.4, são (RABINER et al, 1985):

1. $N$, número de estados do modelo. Estados individuais são denotados por $(q_1; q_2; ...; q_N)$.
2. $\mathbf{A} = [a_{ij}]$, $1 \leq i, j \leq Ne$, a matriz transição de estados. Cada $a_{ij}$ corresponde à probabilidade de ocorrer uma transição do estado $qi$, no instante de tempo $t$, para o estado $q_j$, no instante $t+1$. A transição pode ser de tal forma que o processo permaneça no estado $q_i$ em $t+1$, ou se mova para o estado $q_j$.

   $a_{ij}$ = prob($q_j$ em $t+1$/$q_i$ em $t$). Para modelos *left-right* usa-se a restrição $a_{ij} = 0,\ j < i,\ j > i+2$.
3. $\mathbf{B} = [b_j(k)],\ 1 \leq j \leq N$ e $1 \leq k \leq M$, a distribuição de probabilidades dos símbolos da observação no estado $j$ em que $b_j(k) = P[S_k \text{ em t}|\text{q}_t = S_j],\ 1 \leq j \leq Ne \quad 1 \leq k \leq M.$
4. $\Pi = \pi_i = P\{q_i \mid t = 1\},\ 1 \leq i \leq Ne$, vetor de probabilidade do estado inicial. Esse vetor indica a probabilidade de iniciar o processo no estado $q_i$ para $t$=1.

O sinal a ser representado pelo HMM consiste de uma seqüência de *T* vetores de observações, $\mathbf{O}^l = \{O_1, O_2, ..., O_T\}$ em que cada *i*-ésimo vetor $O_i$ caracteriza o sinal no tempo $t=i$.

No caso discreto, ou de densidade discreta, considerado como uma forma alternativa para o uso de HMMs, faz-se uma combinação com a quantização vetorial, em que os parâmetros de interesse são transformados em um conjunto de observações discretas. Cada vetor $O_i$ é trocado por um dos $M$ símbolos possíveis $w_k \in Wqv,\ 1 \leq k \leq M$, em que $Wqv$ representa um alfabeto discreto obtido por meio da quantização vetorial, tal que a distorção na quantização de $O_i$ seja mínima. Seja $q_j$ o estado no tempo *t*, então $\mathbf{B} = [b_j(k)],\ 1 \leq j \leq N$, é a probabilidade de observação do *k*-ésimo símbolo $w_k$ no estado $q_j$ (RABINER et al, 1983).

Neste trabalho, em particular, são utilizados os HMMs de densidades discretas.

Algumas suposições são tomadas, para os Modelos de Markov, para fins de tratamento matemático e computacional:

- Suposição de Markov: assume-se que o próximo estado depende apenas do estado corrente (HMM de primeira ordem), ou seja,

$$a_{ij} = p\{q_{t+1} = j \mid q_t = i\} \tag{4.12}$$

Quando a mudança de estado depende de $n$ outros estados anteriores, o modelo de HMM é de ordem $n$, resultando em maior complexidade computacional.

- Estacionaridade: as probabilidades de transição de estados são independentes do tempo atual, no qual as transições foram realizadas, ou seja,

$$P\{q_{t+1}\} = P\{q_{t1+1} = j \mid q_{t1} = i\} = P\{q_{t2+1} = j \mid q_{t2} = i\} \tag{4.13}$$

- Independência entre as saídas: Assume-se que a saída ou observação atual é estatisticamente independente das saídas ou observações anteriores. Matematicamente, considerando uma seqüência de observações, $\mathbf{O}^l = \{O_1, O_2, ..., O_T\}$ e um modelo $\lambda$, obtêm-se (FECHINE, 2000)

$$P\{\mathbf{O} = \mid q_1, q_2, ..., q_T, \lambda\} = \prod_{t=1}^{T} P(O_t \mid q_t, \lambda) \tag{4.14}$$

Resumidamente, tem-se que

1. O modelo para cada $l$-ésimo sinal de voz de entrada é denotado por $\lambda_l$= (**A**, **B**, $\Pi$).
2. Inicia-se no estado particular $q_i$ ($t$=1), que depende da distribuição do estado inicial e produz um símbolo de saída $O_t = w_k$, de acordo com $b_i(k)$.
3. Do estado atual, $q_i$, move-se para o estado $q_j$ ou se permanece em $q_i$, de acordo com $a_{ij}$.

4. Esse processo (passos de 1 a 3) se repete até que o objetivo seja atingido (e.g., quando o número de iterações estabelecido é alcançado).

5. Em modelos *left-right*, o processo se inicia no estado $q_1$ ($t$ = 1) e termina quando são atingidos $T$ passos ($t = T$).

6. Assim, a partir da seqüência de observações $\mathbf{O}^l = \{O_1, O_2, ..., O_T\}$ e dos parâmetros necessários, obtém-se o HMM referente a cada $l$-ésimo sinal de voz.

Cada vetor de observação é obtido, neste trabalho, a partir da análise por predição linear (coeficientes LPC) e análise cepstral (coeficientes cepstrais, delta-cepstrais, cepstrais ponderados, delta-cepstrais ponderados e mel-cepstrais), como descrito no Capítulo 3.

Seja $Wq = \{w_1, w_2, ..., w_M\}$, o alfabeto discreto usado para representar a seqüência de observações $O^l$, define-se o cálculo da probabilidade de ocorrência de uma dada seqüência por

$$P\{O_1, O_2, ..., O_T\} = \pi_i \cdot \mathbf{A} \cdot \mathbf{B}. \tag{4.15}$$

Um conjunto de dados de treinamento é assumido, a partir do qual se constroem modelos para cada sinal de voz. Sendo os dados de treinamento compostos de sinais de vozes patológicos, para identificar se os sinais de teste têm ou não a patologia, calcula-se a medida de probabilidade associada aos HMMs de referência pré-armazenados. Determina-se um limiar de probabilidade que, se ultrapassado, indica a presença da patologia. Caso a probabilidade esteja abaixo do limiar pré-estabelecido, o sinal é tido como não-patológico.

Dado um HMM, há três questões básicas (problemas fundamentais) que devem ser resolvidas: o treinamento, a estimação ou reconhecimento e a decodificação (RABINER, 1989).

No treinamento deseja-se determinar os parâmetros do modelo que maximizem a probabilidade de geração da observação. Nesse procedimento, pode-se utilizar como solução o algoritmo *Forward-Backward*, também conhecido como algoritmo de reestimação de Baum-Welch (RABINER, 1989; FAGUNDES & ALENS, 1993; DIAS, 2000).

Na estimação ou reconhecimento, deseja-se determinar qual o modelo, dentre os vários modelos, que mais provavelmente gerou uma dada seqüência de observações. Nesse procedimento utiliza-se como solução o algoritmo *Forward* ou o algoritmo de Viterbi (DIAS, 2000).

Na decodificação utiliza-se como solução o algoritmo de Viterbi que, a partir de uma seqüência de observação, tem como função determinar a seqüência de estados que mais provavelmente produziu as observações. Ou seja, determinar a seqüência ótima de estados do modelo.

Esses problemas podem ser colocados e solucionados, conforme descrito a seguir (RABINER, 1989; COSTA, 1994):

- **Problema 1** – Treinamento

Na fase de treinamento é feita a estimação dos parâmetros dos modelos $\lambda_l = (A, B, \pi)$, um modelo para cada $l$-ésimo locutor $(1 \leq l \leq L)$ (SATISH & GURURAJ, 1993). Desde que exista um procedimento de reestimação convergente para o modelo de densidades discretas, teoricamente é possível escolher aleatoriamente valores iniciais para cada um dos parâmetros do modelo (sujeitos às restrições iniciais) e deixar a reestimação determinar os valores ótimos (máxima verossimilhança), que correspondem aos HMMs de referência, um para cada uma das $L$ elocuções (RABINER, 1989; FECHINE, 2000).

No caso de Modelos de Markov, a estimação pode ser realizada usando o processo iterativo de Baum-Welch, descrito por meio dos seguintes passos (RABINER, 1989; LEVINSON et al, 1983):

1. Atribuição inicial dos valores para os parâmetros do modelo $\lambda_l = (A, B, \pi)$ e para a probabilidade $P_l$;
2. Reestimação dos parâmetros do modelo pelo algoritmo de reestimação de Baum-Welch, obtendo-se $\overline{\lambda}_l$;
3. Cálculo da probabilidade $\overline{P}_l$ associada ao modelo $\overline{\lambda}_l$ reestimado e comparação com a probabilidade anteriormente calculada $P_l$;
4. Se $\overline{P}_l - P_l \leq \delta$ (limiar), o processo de reestimação é finalizado. Caso contrário, retorna-se ao passo 2.

As atribuições iniciais dos parâmetros do modelo devem obedecer regras simples, de forma a satisfazer as restrições do modelo *letf-right*, de acordo com as equações (4.8) a (4.10). Para a matriz $\mathbf{B} = [b_i(k)]$, assume-se que todos os símbolos nos estados são equiprováveis e $b_j(k)$ inicia com $1/M$ para todo $j$, $k$, por simplificação.

Para reestimação de Baum-Welch são utilizadas as seguintes equações (RABINER, 1989; LEVINSON et al, 1983):

1. $\overline{a_{ij}} =$ (número esperado de transições do estado $q_i$ para o estado $q_j$)/ (número esperado de transições do estado $q_i$).
2. $\overline{b_j(k)} =$ (número esperado de vezes no estado $j$ observando o símbolo $w_k$)/(número esperado de vezes no estado $j$).

ou seja,

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_t(i)} \qquad 1 \le i \le Ne, \;\; 1 \le j \le N. \tag{4.16}$$

$$\overline{b_j(k)} = \frac{\sum_{t=1, O_t = w_k}^{T} \alpha_t(i) \beta_t(j)}{\sum_{t=1}^{T} \alpha_t(j) \beta_t(j)}, \qquad 1 \le j \le Ne, \quad 1 \le k \le M. \tag{4.17}$$

Com

$$\sum_{j=1}^{N} a_{ij} = 1; \sum_{k=1}^{M} b_j(k) = 1; \sum_{i=1}^{N} \pi_i = 1, \qquad a_{ij} \ge 0; b_j(k) \ge 0; \pi_i \ge 0. \tag{4.18}$$

Cada parâmetro $b_j(O_t), \; 1 \le j \le Ne$ e $1 \le t \le T$, é obtido a partir da comparação (em relação a um dado estado $j$ e variando $t$) com os valores da matriz $[b_j(k)]$ referentes ao índice $k$ do símbolo associado ao vetor $O_t$ no mesmo estado $j$. Atribui-se a $b_j(O_t)$ o valor de $b_j(k)$ correspondente ao referido símbolo $w_k$, no estado $j$.

A probabilidade $\alpha_t(i)$ é denominada probabilidade de avanço (*forward probability*), pois está associada à ocorrência de uma dada seqüência de observações $O^l = \{O_1, O_2, ..., O_T\}$, segundo o tempo crescente (iniciando em $t$=1 indo até $t$=$T$), sendo formulada como (RABINER, 1989):

1. Inicialização:

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le Ne. \tag{4.19}$$

2. Indução:

$$\alpha_{t+1}(j) = \left\{ \sum_{i=1}^{N} \alpha_t(i) a_{ij} \right\} b_j(O_{t+1}), \quad 1 \le t \le T-1, \quad 1 \le j \le Ne. \tag{4.20}$$

A probabilidade $P_l$, associada ao modelo $\lambda_l = (A, B, \pi)$ é determinada por (RABINER, 1989):

$$P_l = \text{Prob}(O^l \mid \lambda_l) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \tag{4.21}$$

para algum $t$, $1 \le t \le T$.

Fazendo-se $t = T-1$, obtém-se

$$P_l(O^l \mid \lambda_l) = \sum_{i=1}^{N} \alpha_T(i), \tag{4.22}$$

sendo

$$\alpha_T(i) = P_l(O_1, ..., O_T, q_T = i \mid \lambda_l). \tag{4.23}$$

O cálculo das probabilidades de avanço (*forward probability*) se inicia atribuindo ao estado $q_i$ o vetor inicial $O_1$.

De forma similar, $\beta_t(i)$ é denominada probabilidade de retrocesso (*backward probability*), pois está associada à ocorrência da seqüência de observações $O^l = \{O_1, O_2, ..., O_T\}$ segundo o tempo decrescente. A probabilidade de ocorrência do instante t+1 até o instante T, sendo definida como (RABINER, 1989):

$$\beta_t(i) = P_l(O_{t+1}, O_{t+2}, ..., O_T \mid q_t = i \mid \lambda_l). \tag{4.24}$$

ou seja, a probabilidade da seqüência de observações parcial do instante de tempo $t+1$ até o fim, dado o estado $q_i$ no instante de tempo $t$ e o modelo $\lambda_l$. Assim pode-se obter $\beta_t(i)$ indutivamente da seguinte forma (RABINER, 1989):

1. Inicialização:

$$\beta_T(i) = 1, \quad 1 \le i \le Ne. \tag{4.25}$$

2. Indução:

$$\beta_T(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, ..., 1, \ 1 \le i \le Ne. \tag{4.26}$$

O passo 1 define, arbitrariamente, $\beta_T(i) = 1$ para todo $i$. O passo 2 mostra que para ter ocorrido o estado $q_i$ no instante de tempo $t$, levando-se em conta a seqüência de observações no instante de tempo $t$+1, é necessário considerar todos os possíveis estados $q_j$ no instante $t$+1, considerando a transição de $q_i$ para $q_j$ (o termo $a_{ij}$), como também a observação $O_{t+1}$ no estado $j$ (O termo $b_j(O_{t+1})$).

No método iterativo proposto por Baum e Welch (RABINER, 1986) escolhe-se $\lambda_l$ tal que $P_l(O^l \mid \lambda_l)$ seja localmente máxima. O modelo reestimado $\bar{\lambda}_l = (\bar{A}, \bar{B}, \pi)$ (em modelos do tipo *letft-right* $\pi$ não precisa ser reestimado) é melhor ou igual ao modelo estimado anteriormente $\lambda_l$, desde que $P_l(O^l \mid \bar{\lambda}_l) \ge P_l(O^l \mid \lambda_l)$. Assim, utiliza-se $\bar{\lambda}_l$ no lugar de $\lambda_l$ repetindo o processo de reestimação para uma dada seqüência observada, $O^l$, até que seja atingido um número de iterações desejado ou o valor de probabilidade escolhido, para finalizar o processo. O resultado final ou estimado é denominado estimação de máxima verossimilhança do HMM, obtendo-se assim os HMMs de referência, um para cada um dos L sinais de voz (RABINER, 1989; FECHINE, 2000).

- **Problema 2** – Reconhecimento

Na fase de reconhecimento é realizada a estimação da probabilidade de ocorrência de uma dada seqüência de observações $O^l = \{O_1, O_2, ..., O_T\}$, associada a cada modelo $\lambda_l = (A, B, \pi)$, obtido durante a fase de treinamento $(1 \leq l \leq L)$.

Uma vez que os HMMs tenham sido treinados para a patologia, a estratégia de identificação é direta, em analogia ao problema de reconhecimento de locutor (RABINER 1985; SAVIC & GUPTA, 1990; FECHINE, 2000). Para que o sinal de entrada seja classificado como patológico ou não, é obtida a seqüência de observações $O^l = \{O_1, O_2, ..., O_T\}$ e gerada a tabela de códigos associada à seqüência, pela quantização vetorial. Em seguida, é calculada a probabilidade associada a cada modelo de referência $\lambda_l = (A, B, \pi)$ (obtido durante a fase de treinamento). Após o cálculo da probabilidade, por meio de uma regra de decisão, o sinal é aceito ou rejeitado pelo sistema (como patológico ou não-patológico).

O procedimento para cálculo da probabilidade $P(O^l \mid \lambda_l)$ é o mesmo já apresentado anteriormente, descrito a seguir.

Fazendo $t = T - 1$, obtém-se (RABINER, 1985):

$$P_l(O^l \mid \lambda_l) = \sum_{i=1}^{N} \alpha_T(i), \tag{4.27}$$

$$P_l(O^l \mid \lambda_l) = \sum_{i=1}^{N} \alpha_T(i). \tag{4.28}$$

Sendo:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq Ne, \tag{4.29}$$

$$\alpha_{t+1}(j) = \left\{ \sum_{i=1}^{N} \alpha_t(i) a_{ij} \right\} b_j(O_{t+1}), \quad 1 \leq t \leq T-1, \ 1 \leq j \leq Ne. \tag{4.30}$$

Os coeficientes $a_{ij}$ e $\pi$ correspondem, exatamente, aos valores de referência da matriz $\mathbf{A}$ e vetor $\pi$, respectivamente.

Os coeficientes $b_j(O_t)$ são obtidos a partir da matriz $B=[b_j(k)]$, da seguinte forma: a cada vetor $O_t$ de uma $l$-ésima elocução corresponde, após a quantização vetorial, um índice do quantizador vetorial (símbolo $w_k$). Cada coeficiente $b_j(k)$ representa a probabilidade de ocorrência de um dado símbolo $w_k$, no estado *j*. Assim, cada coeficiente $b_j(O_t)$ corresponde ao valor da probabilidade do símbolo associado àquele estado $j$.

Similarmente a sistemas de reconhecimento de palavras ou verificação de locutor, a elocução que apresentar o maior valor de probabilidade é a elocução identificada pelo sistema (ou aceita), desde que ela seja maior que um dado limiar, caso contrário, a elocução é rejeitada. A aceitação ou rejeição dar-se-á de acordo com que base de dados com a qual o sistema foi treinado. Se o sistema for treinado com voz patológica, apenas, por exemplo, a aceitação significará que o sinal de entrada é patológico. Caso contrário, o sinal é dito não-patológico.

- **Problema 3** – Decodificação

O algoritmo de Viterbi é uma solução ótima recursiva ao problema de estimar a seqüência de estados de um processo de Markov discreto no tempo (COSTA, 1994; FORNEY, 1973).

Em sua forma mais geral, o algoritmo de Viterbi pode ser visto como uma solução ao problema de maximizar a estimação da probabilidade a *posteriori* da seqüência de estados de um processo de Markov discreto no tempo. Em outras palavras, dada uma seqüência de observações de um processo de Markov discreto no tempo, o algoritmo de Viterbi fornece a seqüência de estados, $\mathbf{Q}s=\{q_i,q_j,...\}$, para a qual a probabilidade a posteriori, $\mathrm{P}(Qs|\mathbf{O})$, seja máxima.

Considerando a cadeia de Markov do tipo esquerda-direita (*left-to-right*) com cinco estados, mostrada na Figura 4.4, o algoritmo de Viterbi pode ser melhor compreendido associando a ela uma descrição mais redundante chamada treliça (Figura 4.6) (COSTA, 1994).

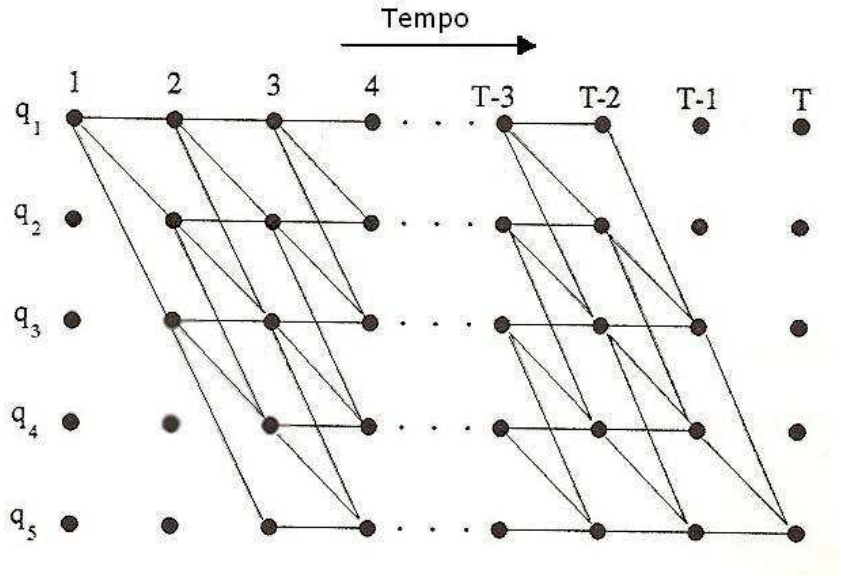



**Figura 4.6 – Estrutura em treliça associada à cadeia de Markov da Figura 4.5.**

Na estrutura em treliça, cada nó corresponde a um estado distinto da cadeia de Markov em um dado instante de tempo e cada ramo representa uma transição para um novo estado no instante de tempo imediatamente posterior. A treliça começa e termina em estados bem definidos, ou seja, nos estados inicial e final da cadeia de Markov, respectivamente. A propriedade mais importante, inerente a essa estrutura, é que para cada seqüência de estados possível, $Q$, corresponde um único caminho na treliça e vice-versa (FORNEY, 1973; COSTA, 1994).

Observando a Figura 4.6, pode-se notar que, para vários instantes de tempo diferentes, existe mais de um caminho parcial chegando em cada nó (estado), cada um com determinado comprimento (valor de probabilidade). O segmento de caminho mais curto, ou seja, aquele que apresenta maior valor de probabilidade, é chamado de sobrevivente correspondente a cada nó. Em outras palavras, para cada instante de tempo existe um número de sobreviventes igual ao número de nós na treliça.

No último instante de tempo deve existir apenas um único sobrevivente, pois a cadeia de Markov deve terminar em um estado bem determinado. Nesse ponto, o caminho total (de $t = 1$ até $t = T$) representa o menor caminho percorrido, ou seja, apresenta o maior valor de probabilidade. Percorrendo de volta a seqüência de estados desse caminho, determina-se a seqüência de

estados associada que fornece o caminho mais provável, ou seja, a seqüência de estados ótima.

Definindo a variável $\delta_t(i)$ como o maior valor de probabilidade ao longo de um único caminho até o instante de tempo *t*, ou seja, considerando as *t* primeiras observações que terminam no estado $q_i$, tem-se por indução que

$$\delta_{t+1}(j) = [\max_i \delta_t a_{ij}] b_j(O_{t+1}), \quad 1 \le i \le Ne. \tag{4.31}$$

Para se obter a seqüência de estados, é necessário reter a trilha do argumento que maximiza a Equação (4.31), para cada *t* e $j$. Para tanto, define-se a variável $\psi_t(j)$. O método para se encontrar a seqüência de estados ótima é dado por (RABINER, 1989; COSTA, 1994):

1. Inicialização:

$$\delta_1(j) = \pi_i b_i(O_1) \quad 1 \le i \le Ne, \tag{4.32}$$

$$\psi_1(i) = 0. \tag{4.33}$$

2. Recursividade:

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \tag{4.34}$$

$$\psi_t(j) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], \quad 2 \le t \le T, \ 1 \le j \le Ne. \tag{4.35}$$

$$\psi_t(j) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], \quad 2 \le t \le T, \ 1 \le j \le Ne, \tag{4.36}$$

3. Término:

$$P^* = \max_{1 \le i \le N} [\delta_T(i)], \tag{4.37}$$

$$q_T^* = \arg\max_{1 \le i \le N} [\delta_T(i)]. \tag{4.38}$$

4. Seqüência de estados ótima:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, ..., 1. \tag{4.39}$$

O algoritmo descrito tem, portanto, a propriedade de determinar a seqüência de estados, que maximiza a probabilidade $P(O^l \mid \lambda_l)$, para a $l$-ésima palavra (RABINER, 1989; FECHINE, 2000). Assim, esse algoritmo pode ser usado para o ajuste da etapa de reconhecimento e determinação da seqüência de estados ótima do modelo.

A solução desses três problemas permite a elaboração de um sistema de reconhecimento automático da fala, utilizando HMM.

- **Uso de Múltiplas Sequências de Observações**

O maior problema associado ao HMM do tipo *left-right* reside no fato de que não se pode usar uma única seqüência de observações para treinar o modelo (isto é, para a reestimação dos parâmetros do modelo) (RABINER, 1989; COSTA, 1994). Isso se deve à natureza transitória dos estados dentro do modelo, permitindo apenas um pequeno número de observações para qualquer estado (até que uma transição seja feita para um estado sucessor). Assim, a fim de se obter dados suficientes para se fazer estimativas confiáveis de todos os parâmetros do modelo, deve-se usar múltiplas seqüências de observações.

A modificação do método de reestimação é direta e apresentada a seguir (RABINER, 1989).

Seja o conjunto de U seqüências de observações, representado por

$$O = [O^{(1)}, O^{(2)}, ..., O^{(u)}] \tag{4.40}$$

Assume-se que as seqüências de observações são independentes e o objetivo é o ajuste dos parâmetros do modelo $\lambda$ que maximizam a expressão

$$P(O \mid \lambda) = \prod_{u=1}^{U} P(O^{(u)} / \lambda) = \prod_{u=1}^{U} P_u \tag{4.41}$$

Uma vez que as fórmulas de reestimação são baseadas em freqüências de ocorrências de eventos, para as múltiplas seqüências de observações essas fórmulas são modificadas adicionando-se as freqüências de ocorrências individuais de cada seqüência. Assim, as fórmulas de reestimação modificadas são (RABINER, 1989):

$$\overline{a_{ij}} = \frac{\sum_{u=1}^{U} \frac{1}{P_u} \sum_{t=1}^{T_u - 1} \alpha_t^u(i) a_{ij} b_j(O_{t+1}^{(u)}) \beta_{t+1}^u(j)}{\sum_{u=1}^{U} \frac{1}{P_u} \sum_{t=1}^{T_u - 1} \alpha_t^u(i) \beta_t^u(i)}, \quad 1 \le i \le Ne, \quad 1 \le j \le Ne, \qquad (4.42)$$

$$\overline{b_j(k)} = \frac{\sum_{u=1}^{U} \frac{1}{P_u} \sum_{t=1, s.t. O_t = w_k}^{T_u} \alpha_t^u(j) \beta_t^u(j)}{\sum_{u=1}^{U} \frac{1}{P_u} \sum_{t=1}^{T_u} \alpha_t^u(j) \beta_t^u(j)} \quad 1 \le j \le Ne, \quad 1 \le k \le M. \qquad (4.43)$$

## 4.5 Discussão

O uso da Quantização Vetorial e dos Modelos de Markov Escondidos vem sendo abordado em problemas de reconhecimento de fala e de locutor, com excelentes resultados apresentados na literatura.

Neste trabalho, o foco de interesse é modelar uma patologia nas dobras vocais, a partir da análise acústica do sinal patológico, com o objetivo de desenvolver uma técnica não-invasiva para o diagnóstico da presença com desordens provocadas por essa patologia.

A associação de técnicas paramétricas com técnicas estatísticas pode representar um bom caminho para a observação do comportamento dos parâmetros LPC, cepstrais e seus derivados em vozes desordenadas afetadas por patologias nas dobras vocais.

Além da redução da dimensionalidade dos dados, a quantização vetorial, associada a uma medida de distância, é usada para uma pré-classificação, com um classificador individual para cada parâmetro. Assim, o peso de cada parâmetro pode ser avaliado, de acordo com a eficiência obtida nos resultados dessa pré-classificação.

Os Modelos de Markov Escondidos são utilizados como uma etapa de refinamento do processo de classificação, com o objetivo de melhorar os resultados obtidos na etapa de pré-classificação. Os resultados obtidos estão relatados e discutidos no Capítulo 5.

# Capítulo 5
# Apresentação e Análise dos Resultados obtidos

## 5.1 Introdução

A caracterização acústica do sinal de voz é uma etapa fundamental para o processo de detecção de uma determinada patologia, ou para discriminação entre patologias. Neste trabalho, o objetivo da caracterização visa, por meio de análise paramétrica e não-paramétrica, avaliar o comportamento acústico da patologia de interesse (edemas nas dobras vocais). Assim, espera-se ter uma visão clara da eficiência e importância dos parâmetros estudados em um processo de classificação/discriminação de patologias na laringe.

O grau de confiabilidade e eficiência de um processo de discriminação de vozes patológicas depende muito de quais características ou parâmetros são utilizados pelo classificador escolhido. A questão fundamental é saber o quanto uma determinada medida do sinal representa bem as variações impostas pela patologia (AGUIAR NETO et al, 2007a; AGUIAR NETO et al, 2007b; COSTA et al, 2008a; COSTA et al, 2008b).

Neste capítulo são apresentados os resultados obtidos para a caracterização acústica de vozes afetadas por patologias da laringe, mais especificamente, patologias nas dobras vocais, como nódulos, cistos e edemas.

Para esse propósito, mostra-se o desempenho da análise da codificação por predição linear (LPC), como também da análise cepstral baseada na análise LPC. Serão empregados, para análise, os coeficientes cepstrais, cepstrais ponderados, delta cepstrais e delta cepstrais ponderados. E, ainda, uma análise não-paramétrica do sinal, usando coeficientes mel-cepstrais.

Emprega-se um quantizador individual para cada parâmetro, usando a medida de distância do erro médio quadrático, para uma pré-classificação dos sinais para efeito de caracterização acústica. Um refinamento no processo de classificação é realizado por meio da modelagem usando Modelos de Markov Escondidos.

Todo o processo e os resultados obtidos são apresentados e discutidos neste capítulo.

## 5.2 Base de dados

A base de dados usada neste trabalho foi desenvolvida pelo *Massachusetts Eye and Ear Infirmary* (MEEI) *Voice and Speech L*ab (KAY ELEMETRICS, 1994). A base de dados de vozes desordenadas (*Disordered Voice Database*, Model 4337), inclui mais de 1400 amostras de vozes de aproximadamente 700 sujeitos (isto é, vogal sustentada /ah/ e os primeiros 12 segundos da "*Rainbow Passage*") de aproximadamente 700 pessoas. Esta base de dados foi desenvolvida como um auxílio na análise acústica e perceptual de vozes desordenadas para aplicações clínicas ou de pesquisa. Ela inclui amostras de pacientes com uma larga variedade de desordens vocais por causas orgânicas, neurológicas, traumáticas, psicogênicas entre outras. Todas as amostras foram coletadas em um ambiente controlado com as seguintes características: baixo nível de ruído, distância constante do microfone, taxas de amostragem de 25 kamostras/s (sinais patológicos) ou 50 kamostras/s (sinais normais), com 16 bits por amostra[7].

Os seguintes casos da base de dados foram considerados:

- Vozes patológicas:
    - Edemas nas dobras vocais: 44 vozes - 33 mulheres, na faixa de 17 a 85 anos e 11 homens, na faixa de 23 a 63 anos, a maioria com edema bilateral (32 casos);
    - Outras patologias nas dobras vocais: 23 casos contendo vozes de pessoas afetadas por cistos, nódulos e paralisia na faixa etária de 18 a 80 anos (8 homens entre 43 a 75 anos e 15 mulheres entre 18 e 80), sendo a maioria acima de 40 anos. A faixa etária mais jovem está nas mulheres (4 de 18 a 21 anos), com nódulos nas dobras vocais.
- Vozes normais: 53 casos de vozes normais, sendo 32 mulheres (de 26 a 59 anos) e 21 homens (22 a 52 anos).

Para abreviar, usa-se a expressão "Outras Patologias" para os casos que não forem edemas (cistos, nódulos e paralisia).

[7] http://www.kayelemetrics.com/Product%20Info/CSL%20Options/4337/4337.htm

## 5.3 Metodologia

O processo de discriminação de vozes patológicas está dividido em duas etapas: treinamento e teste. Para a etapa de treinamento (Figura 5.1), foram usados 20 sinais de vozes femininas e 5 de vozes masculinas, afetados por edemas nas dobras vocais. O restante dos sinais foi usado na etapa de teste: 19 vozes com edemas, 23 com sinais incluindo as outras patologias mencionadas e 53 vozes normais, totalizando 95 sinais. Entre vozes normais e patológicas, foram usados, no total, 120 sinais.

Na Figura 5.1 é ilustrado o procedimento utilizado desde a aquisição do sinal até a obtenção dos parâmetros a partir da quantização vetorial, cujo embasamento teórico está descrito nos capítulos 3 e 4.

Considera-se, para cada parâmetro, um diagrama similar ao da Figura 5.1, já que é aplicado um classificador individual para cada um (LPC, cepstral, delta-cepstral, cepstral ponderado, delta-cepstral ponderado e mel-cepstral).

Para cada tipo de parâmetro, foram extraídos 12 coeficientes, a partir do sinal segmentado em 20 ms com sobreposição de 50% (a cada 10 *ms*), pré-enfatizado e janelado (janela de *Hamming* – 0,95).

Os sinais de vozes patológicas têm duração média de 1 segundo e taxa de amostragem de 25 kamostras/s, enquanto que os sinais de vozes normais têm duração média de 3 segundos e taxa de amostragem de 50 kamostras/s.

A Quantização Vetorial foi aplicada com 64 níveis ($Nq$) e dimensão 12 ($K$), gerando um dicionário (*codebook*) para cada parâmetro do sinal em análise obtendo-se, assim, os padrões de referência.

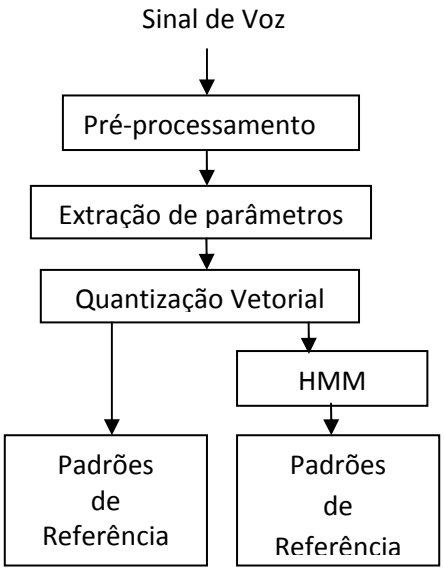



**Figura 5.1 – Fase de treinamento do processo de discriminação de vozes patológicas.**

Na Figura 5.2 é ilustrada, em diagrama em blocos, a fase de teste/classificação, a partir da quantização vetorial dos parâmetros do sinal em análise. De forma semelhante à fase de treinamento, o sinal é pré-processado antes da aquisição dos parâmetros. Após a extração dos parâmetros, os padrões de teste são obtidos pela geração do dicionário por quantização vetorial ($N$=64 e $k$=12).

Uma etapa de pré-classificação é realizada logo após a quantização vetorial, em que os padrões de teste são comparados com os padrões de referência obtidos na fase de treinamento. Para tanto, é utilizada a medida de distância do erro médio quadrático mínimo. A resposta obtida, a partir dos valores da distorção, indica voz patológica ou não-patológica, de acordo com um limiar pré-estabelecido, que proporcione a melhor separação entre as classes.

Os casos que não foram classificados corretamente na etapa de pré-classificação são submetidos a uma etapa de classificação final, após um refinamento proporcionado por um classificador baseado em Modelos de Markov Escondidos do tipo discreto. O modelo aplicado é do tipo esquerda-direita (*left-to-right*), com cinco estados (vide Capítulo 4). A escolha de cinco estados é conveniente ao propósito deste trabalho, que utiliza como sinal de fala a vogal sustentada /a/, o que não exige um grande número de estados para representar as mudanças ocorridas no sinal de voz em análise.

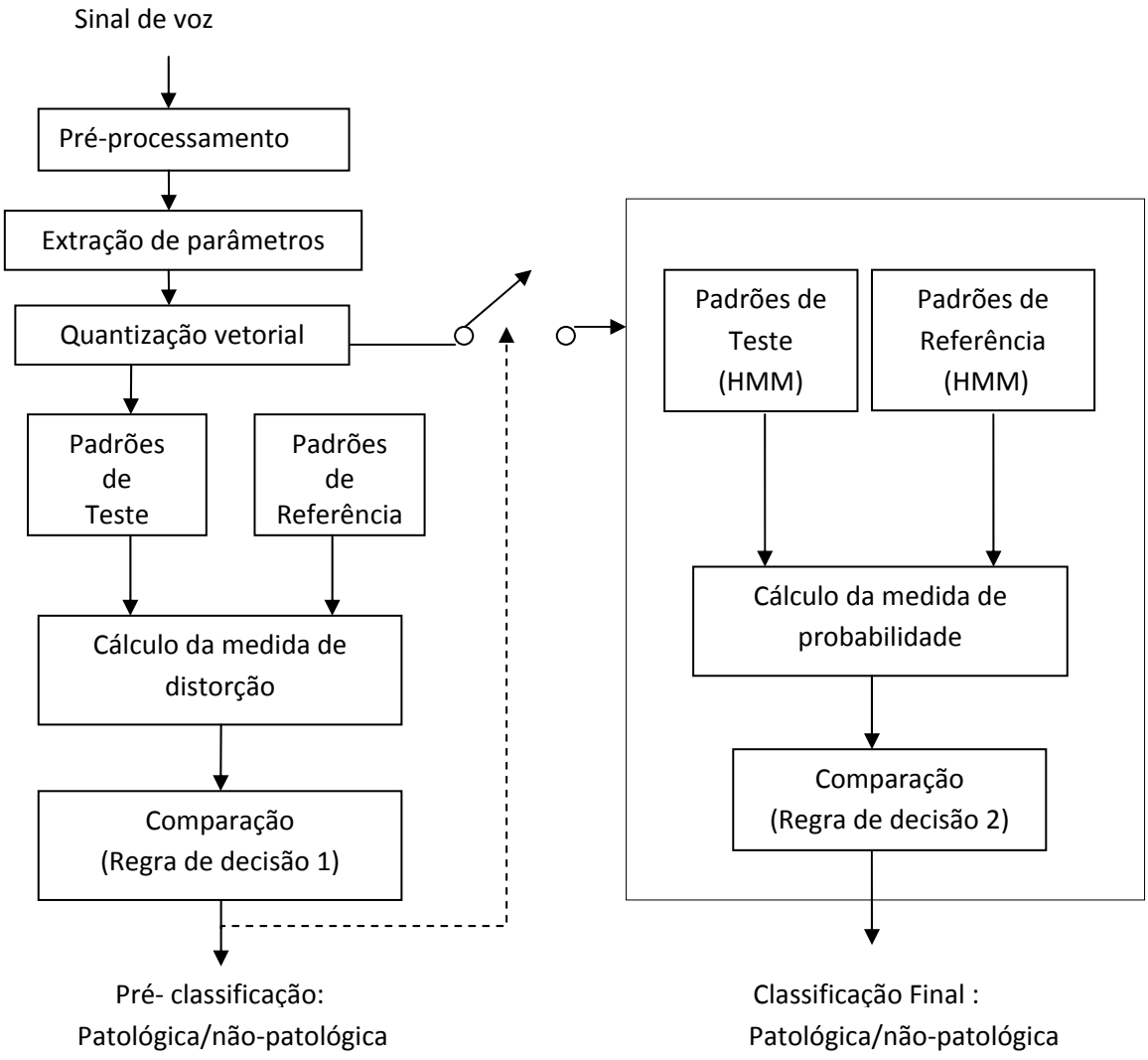



**Figura 5.2 - Fase de teste do processo de discriminação de vozes patológicas.**

A análise de desempenho é realizada, inicialmente, usando as seguintes medidas (GODINO-LLORENTE et al, 2006, AGUIAR NETO et al, 2007a; COSTA et al, 2008b):

- Correta aceitação (*CA*) - A presença da patologia é detectada quando a patologia está realmente presente, também chamada de Verdadeiro Positivo.
- Correta Rejeição (*CR*) - É detectada a correta ausência da patologia (Verdadeiro Negativo).
- Falsa Aceitação (*FA*) - É detectada a presença da patologia quando ela não está presente (Falso Alarme ou Falso Positivo).
- Falsa Rejeição (*FR*) - A presença da patologia é rejeitada quando, de fato, ela está presente (Falso Negativo).

- Especificidade (*SP*) - Representa a probabilidade de que a patologia seja rejeitada quando ela está ausente, ou seja, representa a proporção de pessoas sem a doença, cujo teste dá negativo. Indica quão bom é o método empregado na identificação dos indivíduos com vozes não-patológicas. É dada por

$$SP = \frac{CR}{CR + FA} \text{x} 100. \tag{5.1}$$

- Sensibilidade (*SE*) - representa a probabilidade de que a patologia seja detectada quando a mesma estiver realmente presente, ou seja, proporção de pessoas com a patologia de interesse que têm o resultado do teste positivo. Indica quão bom é o teste para identificar os indivíduos com a patologia, dada por

$$SE = \frac{CA}{CA + FR} \text{x} 100. \tag{5.2}$$

- Eficiência (E) - representa a taxa de classificação correta de uma dada classe, quando ela está presente, dada por

$$E = \frac{CR + CA}{CA + CR + FA + FR} \text{x} 100. \tag{5.3}$$

Um ponto de limiar deve ser escolhido para a classificação e a avaliação de desempenho do método empregado. O limiar deve ser definido de tal forma que se obtenha a melhor separação entre as classes.

Além disso, é preciso avaliar cuidadosamente a importância relativa da sensibilidade e especificidade do teste para o ponto de transição do diagnóstico mais adequado. A estratégia geral é a seguinte:

a) Se a principal preocupação é evitar resultado de falsa aceitação ou falso positivo, então o ponto de corte, ou limiar escolhido, deve objetivar o máximo de especificidade.
b) Se a preocupação maior é evitar resultado de falsa rejeição ou falso-negativo, então o ponto de corte, ou limiar escolhido, deve objetivar o máximo de sensibilidade.

Para efeito da avaliação de desempenho e escolha do limiar, são utilizados gráficos representativos da distribuição dos sinais de acordo com a medida de distorção Euclidiana, além das curvas ROC e curvas DET. São apresentados também tabelas e gráficos comparativos entre os métodos empregados.

A Curva ROC (*Receive Operator Characteristic Curve*) é uma maneira adequada de estabelecer o ponto de corte, otimizando a sensibilidade e especificidade do teste diagnóstico. É um método simples e robusto, muito utilizado para avaliação de desempenho em testes de diagnósticos médicos. O traçado da curva é feito levando em conta as probabilidades de ocorrência de Correta Aceitação (*CA*) em função da probabilidade da ocorrência de Falsa Aceitação (*FA*) para cada ponto de operação notado na curva, variando-se valores de corte. Geometricamente, a curva ROC é um gráfico de pares (x,y) (que correspondem à especificidade e à sensibilidade, respectivamente num plano designado por plano ROC unitário. A designação de plano ROC unitário deve-se ao fato das coordenadas deste gráfico representarem medidas de probabilidade e, por conseguinte, variarem entre zero e um.

O resultado ideal do teste é aquele que alcança a extremidade mais superior e esquerda do gráfico. Uma das vantagens deste método é que as curvas de diferentes testes diagnósticos podem ser comparadas; quanto melhor o resultado da comparação, mais perto está a curva do canto superior esquerdo do gráfico[8]. Na Figura 5.3 são ilustrados exemplos de curva ROC.

A Curva DET (*Detection-Error Tradeoff Curve*)[9] tem sido amplamente usada para a avaliação do desempenho de detecção em tarefas de identificação de locutor. A curva DET traça um gráfico com as taxas de erro em ambos os eixos, dando tratamento uniforme aos tipos de erro (Falsa Aceitação e Falsa Rejeição). Quanto mais à esquerda e na porção inferior estiver a curva, melhor o desempenho do método empregado (MARTIN et al, 1997; GODINO-LLORENTE et al, 2006).

Na Figura 5.4 estão ilustrados exemplos de curva DET (WESSEL et al, 2001), utilizadas para avaliar desempenho de medidas de desempenho em reconhecimento de voz contínua com grandes vocabulários.

[8] http://www.unifesp.br/dmed/cardio/ch/utiliza.htm
[9] http://www.nist.gov/speech/publications/papers/

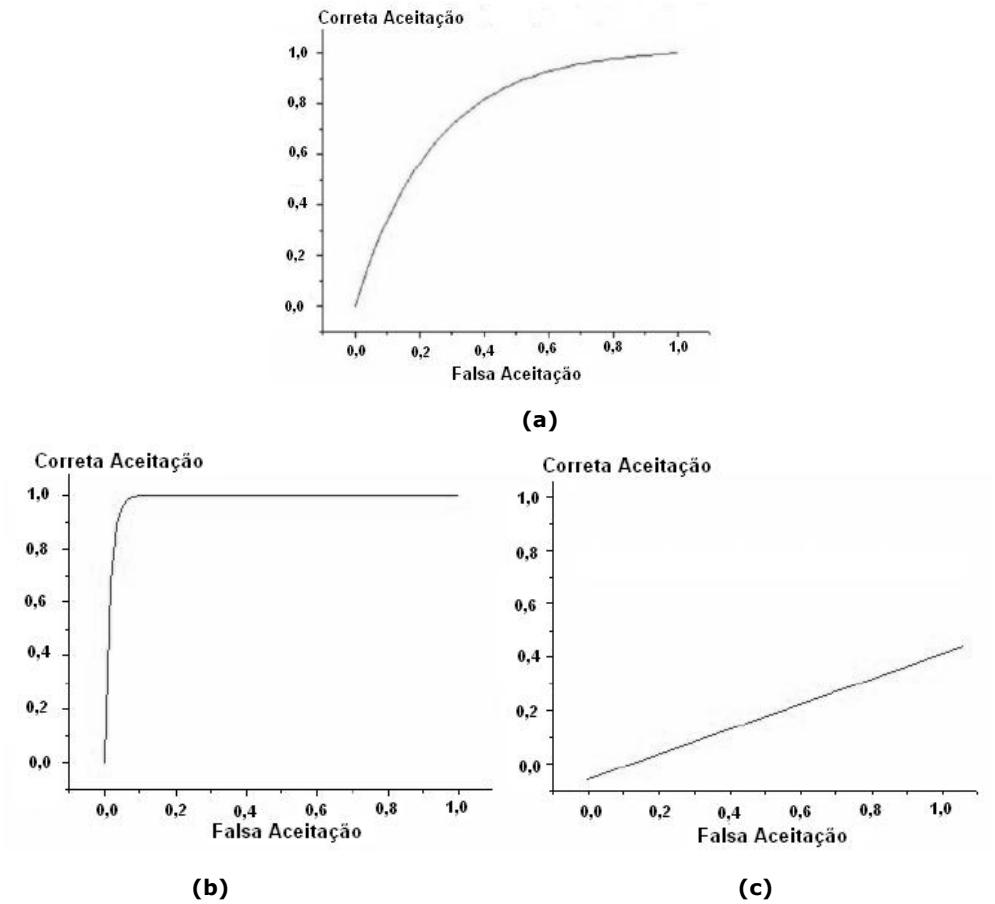



**Figura 5.3 – Exemplos de Curva ROC: (a) Traçado de uma curva ROC típica; (b) Curva ROC para um bom desempenho; (c) Curva ROC para um desempenho ruim.**

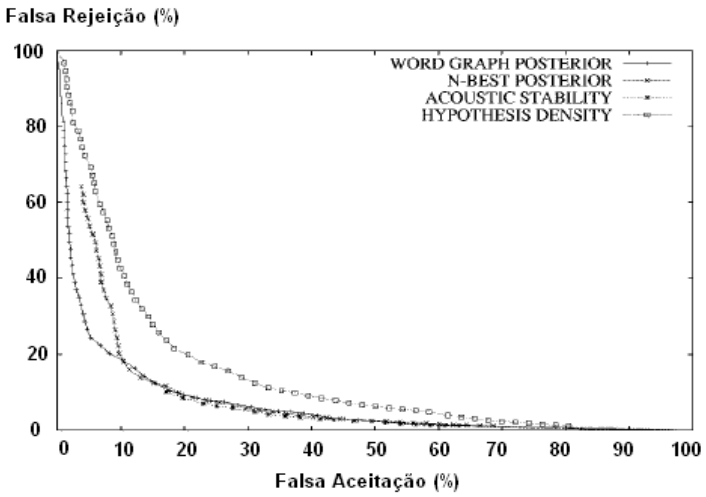



**Figura 5.4 – Exemplos de Curva DET.**
FONTE: (WESSEL et al, 2001)

A seguir, serão apresentados os resultados obtidos para cada um dos métodos empregados, além da comparação de desempenho entre os métodos.

Nas Figuras 5.5 a 5.10, são apresentadas as distribuições dos valores das distorções obtidas para os sinais da base de dados, em cada método de análise. Os resultados obtidos nessa fase foram usados como pré-classificação dos sinais. Após esta etapa, os resultados em que ocorram erros de classificação, passam pela etapa de refinamento, sendo modelados usando modelos de Markov escondidos discretos, obtendo-se, então a classificação final.

## 5.4 Resultados obtidos - Pré-Classificação

A seguir, são apresentados os resultados obtidos pela Análise LPC (coeficientes LPC) e pelas Análises Cepstral (coeficientes cepstrais, delta-cepstrais, cepstrais ponderados e delta-cepstrais ponderados) e Mel-cepstral.

### 5.4.1 Análise LPC

No método da análise por predição linear, foram obtidos os coeficientes LPC, após o pré-processamento do sinal (segmentação, janelamento e pré-ênfase). Foi empregado um preditor de ordem 12 ($p$=12). Os coeficientes LPC foram obtidos usando o método da autocorrelação pelo algoritmo de Levinson-Durbin (RABINER and SCHAFER, 1978).

Na Figura 5.5 é mostrada a distribuição dos sinais de vozes: normal, afetados por edema nas dobras vocais e por outras patologias (nódulos, cistos e paralisia) de acordo com a distorção do euclidiana, empregada na quantização vetorial. Por simplicidade, vozes afetadas por edemas nas dobras vocais, são denominadas nos gráficos e tabelas como “Edema”. As outras patologias, como nódulos, cistos e paralisia estão incluídas na mesma classe, a partir daqui denominada por “Outras Patologias (OP)” e vozes normais, por “Normal” (AGUIAR NETO et al, 2007a; AGUIAR NETO et al, 2007b; COSTA et al, 2008a; COSTA et al, 2008b).

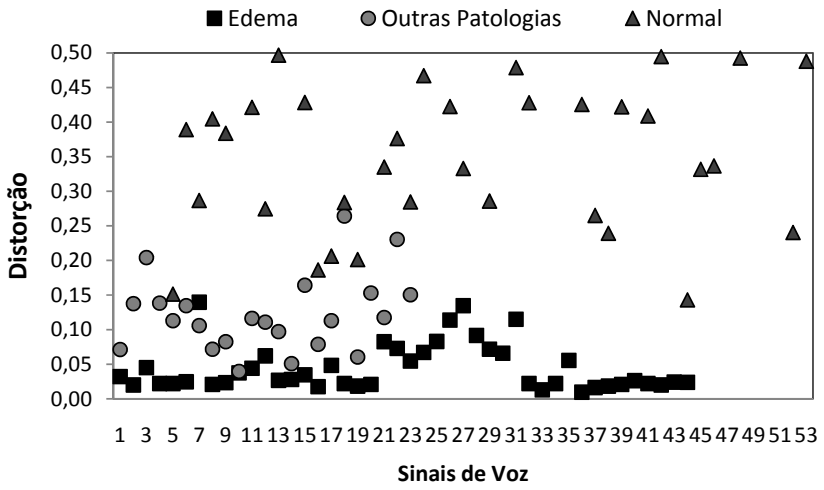



**Figura 5.5 – Comportamento da distorção para vozes normais, vozes afetadas por edema nas dobras vocais e por Outras Patologias – Método LPC.**

Pela observação da Figura 5.5, é clara a separação entre as classes Edema e Normal, havendo uma pequena confusão de Edema com as outras patologias. No entanto, a proximidade dos valores obtidos para as outras patologias causa dificuldades em discriminar entre edemas e outras patologias. Há que se destacar que as patologias escolhidas afetam as dobras vocais, o que sugere certa similaridade, especialmente nos casos de cistos, nódulos e edemas.

## 5.4.2 Análise Cepstral

Na análise cepstral são abordados e analisados os resultados obtidos para os coeficientes cepstrais, delta-cepstrais, cepstrais ponderados e delta-cepstrais ponderados.

### 5.4.2.1 Coeficientes Cepstrais (CEP)

Na Figura 5.6 é ilustrada a distribuição dos sinais de voz de acordo com a medida de distorção, aplicada após a quantização vetorial dos parâmetros cepstrais. Também para esse caso, há uma separação entre as classes Edema e Normal, que permite a distinção entre as mesmas de forma mais clara que entre as Outras Patologias e Edema.

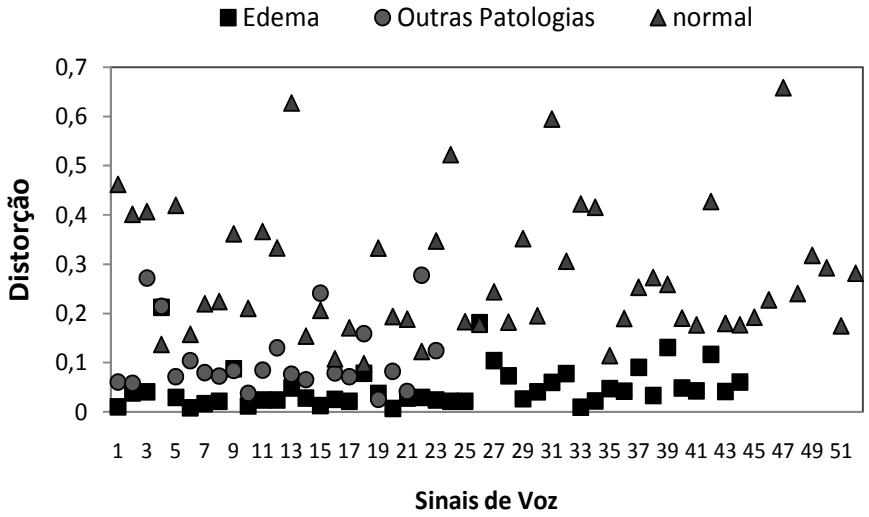



**Figura 5.6 – Comportamento da distorção para vozes normais, vozes afetadas por edema nas dobras vocais e por Outras Patologias – Método LPC.**

### 5.4.2.2 Coeficientes delta-cepstrais (DCEP)

Na Figura 5.7 é apresentado o comportamento, em relação à distorção, dos sinais de vozes normais, com Edema e com Outras Patologias. A separação entre as classes parece não ser tão clara como nos casos dos coeficientes LPC e Cepstrais.

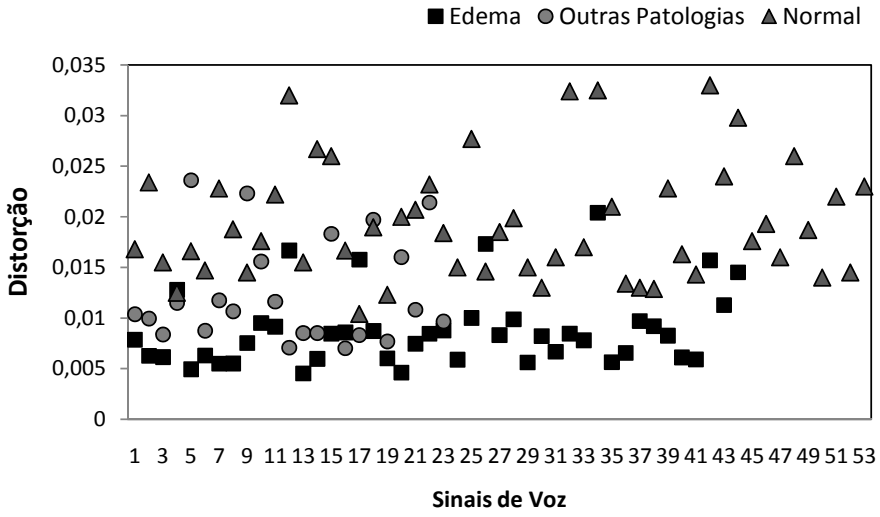



**Figura 5.7 – Comportamento da distorção para vozes normais, vozes afetadas por edema nas dobras vocais e por Outras Patologias – Método DCEP.**

### 5.4.2.3 Coeficientes cepstrais ponderados (CEPP)

Na Figura 5.8 é apresentado o gráfico da separação entre as classes para o método que emprega os coeficientes cepstrais ponderados, de acordo com a distorção medida, para a etapa de comparação e tomada de decisão. Os resultados para os três casos considerados para cada método são apresentados e analisados a seguir.

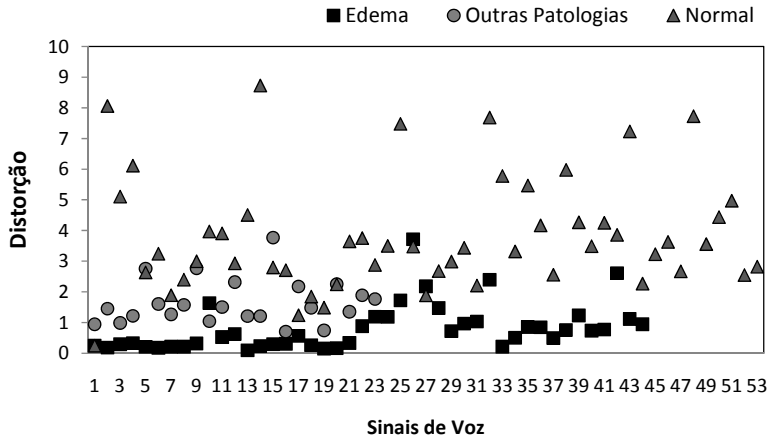



**Figura 5.8 – Comportamento da distorção para vozes normais, vozes afetadas por edema nas dobras vocais e por Outras Patologias – Método CEPP.**

### 5.4.2.4 Coeficientes delta-cepstrais ponderados (DCEPP)

Na Figura 5.9 é mostrado o comportamento dos sinais de vozes afetados por edemas nas dobras vocais, por Outras Patologias (nódulos, cistos e paralisia) e vozes normais, em relação à distorção obtida no processo de classificação para o processo de decisão (comparação).

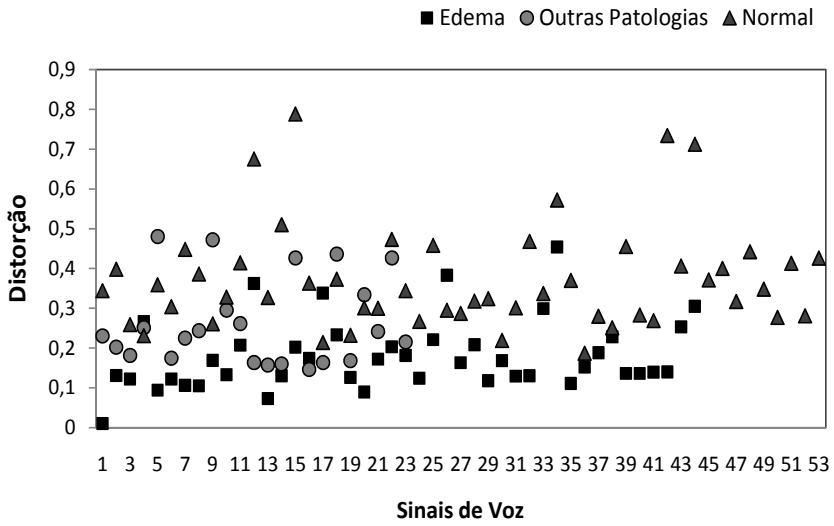



**Figura 5.9 – Comportamento da distorção para vozes normais, vozes afetadas por edema nas dobras vocais e por Outras Patologias – Método DCEPP.**

### 5.4.2.5 Coeficientes Mel-cepstrais (MEL)

Na Figura 5.10 é apresentada a concentração das classes em relação aos valores de distorção, para escolha do limiar mais adequado que separe melhor as classes, proporcionando uma melhor discriminação entre elas.

Os coeficientes mel-cepstrais foram obtidos pelo método no domínio da freqüência (*Mel Frequency Cepstral Coefficients*). Foi utilizado um algoritmo de uma ferramenta para processamento de sinais, a *Voicebox - Speech Processing Toolbox for MATLAB* [10]. Para tanto, foram calculados 12 coeficientes por segmento do sinal de voz. O número de filtros utilizados para compor o banco de filtros na escala mel, conforme método descrito no Capítulo 3 foi de 30 (3ln(*Fa*), *Fa* = 25 kamostras/s).

[10] www.ee.ic.ac.uk/hp/staff/dmb/voicebox

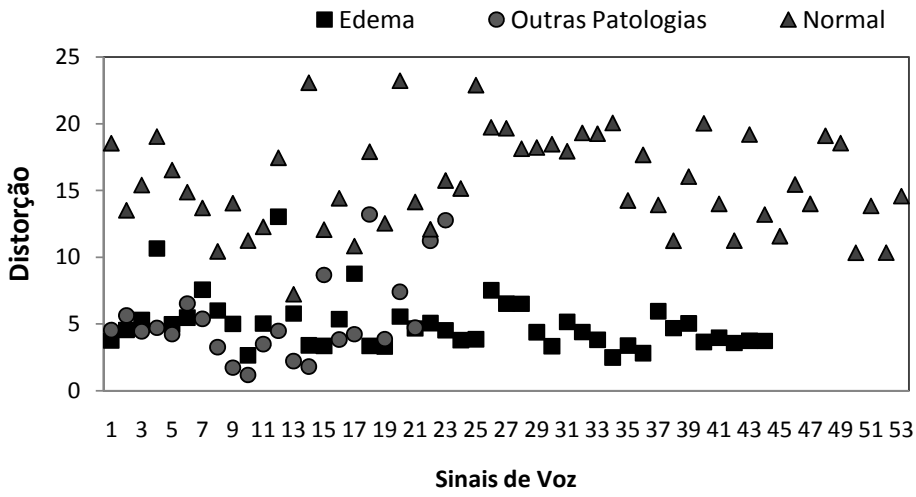



**Figura 5.10 – Comportamento da distorção para vozes normais, vozes afetadas por edema nas dobras vocais e por Outras Patologias – Método MEL.**

Para o processo de classificação, foi empregado um valor de limiar que proporciona uma melhor separação entre as classes, obtido pela análise das curvas ROC e DET, como também pela análise dos valores de Eficiência obtidos para cada método. O valor de limiar que proporciona a melhor taxa de classificação correta foi escolhido.

São abordados três casos específicos para cada método:

- **Caso 1:** sinais de teste com edemas (Classe Edema) e sinais normais (classe Normal);
- **Caso 2:** sinais de teste com edemas (Edema) e Outras Patologias (OP), em classes diferentes;
- **Caso 3:** sinais de teste com edema e outras patologias na mesma classe (Edema +OP) e sinais de vozes normais (Normal).

## 5.5 Comparação de desempenho entre os métodos empregados na etapa de pré-classificação

Os resultados referentes aos seis métodos aplicados para a caracterização acústica de sinais de vozes afetados por edemas nas dobras vocais são

apresentados em tabelas e gráficos resumindo os dados obtidos para cada um dos três casos, detalhados nas seções anteriores. Dessa forma, é possível comparar o desempenho entre os métodos.

**Caso 1:**

Na Tabela 5.1 estão os resultados dos seis métodos para este caso, em que os sinais de teste são sinais com edema e sinais de vozes normais.

**Tabela 5.1: Medidas de desempenho para os seis métodos, em função de limiares de distorção, avaliando vozes com edema e vozes normais (Edema x Normal).**

| MÉTODO | CR (%) | FA (%) | CA (%) | FR (%) | SP (%) | SE (%) | E (%) |
|---|---|---|---|---|---|---|---|
| LPC | 98 | 2 | 100 | 0 | 98 | 100 | 99 |
| CEP | 89 | 11 | 91 | 9 | 89 | 91 | 90 |
| CEPP | 94 | 6 | 86 | 14 | 94 | 86 | 90 |
| DCEP | 98 | 2 | 86 | 14 | 98 | 86 | 92 |
| DCEPP | 91 | 9 | 82 | 18 | 91 | 82 | 87 |
| MEL | 98 | 2 | 95 | 5 | 98 | 95 | 97 |

Os métodos LPC e MEL forneceram os melhores resultados, proporcionando as melhores taxas de falsa aceitação e falsa rejeição, sendo o pior caso obtido para os coeficientes delta-cepstrais ponderados. Os coeficientes cepstrais proporcionaram a maior taxa de falsa aceitação. Dos sinais de vozes normais testados, 11% foram confundidos com Edema, enquanto que nos métodos LPC, DCEP e MEL, esta taxa é de apenas 2%.

Na Figura 5.11 estão ilustradas as curvas ROC para os seis métodos empregados e na Figura 5.12 são apresentadas as curvas DET, para melhor efeito de comparação entre os métodos.

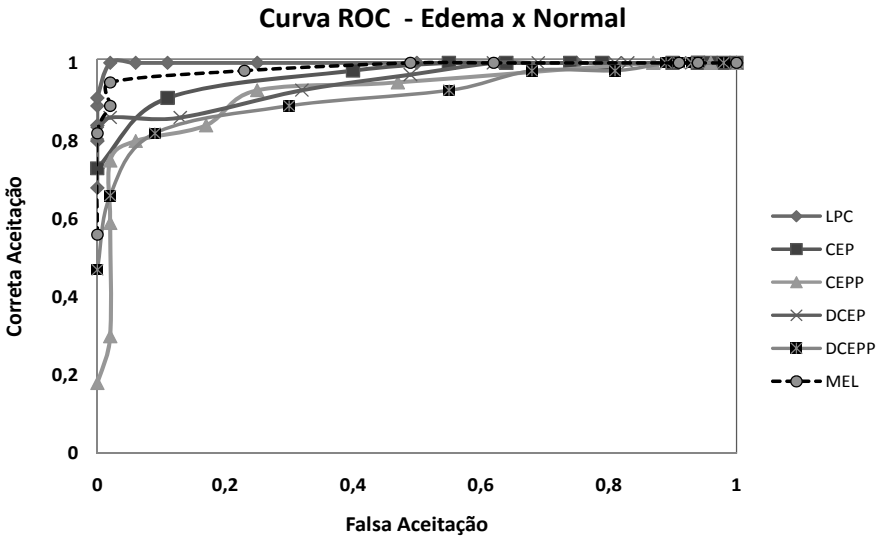



**Figura 5.11 – Curvas ROC para os métodos LPC, CEP, CEPP, DCEP, DCEPP e MEL, para o Caso 1 (Edema x Normal).**

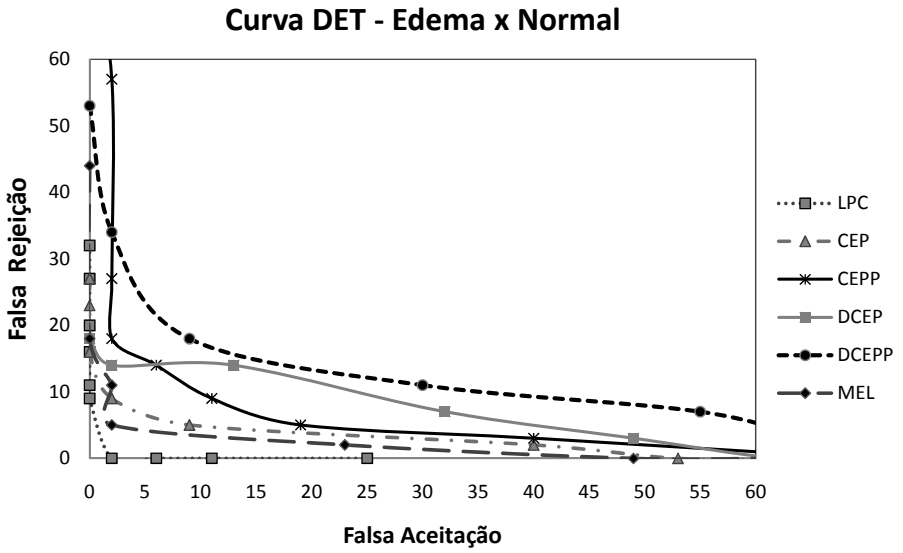



**Figura 5.12 – Curvas DET para os métodos LPC, CEP, CEPP, DCEP, DCEPP e MEL, para o Caso 1.**

Os resultados das curvas confirmam a superioridade do MEL e do LPC. Observa-se, nas curvas DET, que esses métodos proporcionaram as menores taxas de Falsa Rejeição e Falsa Aceitação, seguido dos cepstrais, sendo o delta-cepstral ponderado confirmado como o pior caso.

**Caso 2:**

Na Tabela 5.2, estão os resultados dos seis métodos para o caso em que os sinais de teste são sinais com Edema e sinais de vozes sob Outras Patologias (nódulos, cistos e paralisia). O sistema é treinado com a patologia edema. O método LPC proporcionou a maior eficiência (83%), mesmo apresentando uma taxa de Falsa Aceitação de 14%, bem mais alta do que os delta-cepstrais ponderados (4%). Esse último método, entretanto, apresentou uma taxa de Falsa Rejeição muito alta (53%), diminuindo a sua eficiência.

Para que os métodos proporcionem um melhor desempenho é preciso um refinamento no processo de classificação, que permita acompanhar as pequenas variações entre as patologias.

**Tabela 5.2: Medidas de desempenho para os seis métodos, em função de limiares de distorção, avaliando vozes com edema e vozes sob Outras Patologias (Edema x OP).**

| MÉTODO | CR (%) | FA (%) | CA (%) | FR (%) | SP (%) | SE (%) | E (%) |
|---|---|---|---|---|---|---|---|
| **LPC** | 86 | 14 | 80 | 20 | 86 | 80 | 83 |
| **CEP** | 87 | 13 | 73 | 27 | 87 | 73 | 80 |
| **CEPP** | 78 | 22 | 82 | 18 | 78 | 82 | 80 |
| **DCEP** | 56 | 44 | 82 | 18 | 56 | 82 | 69 |
| **DCEPP** | 96 | 4 | 47 | 53 | 96 | 47 | 72 |
| **MEL** | 65 | 35 | 56 | 44 | 65 | 56 | 61 |

Na Figura 5.13 são apresentadas as curvas ROC, comparando os seis métodos, para esse caso. As curvas ROC apontam para LPC e CEP como os melhores parâmetros na relação entre Falsa Aceitação e Falsa Rejeição. Os métodos MEL e DCEPP geraram os piores resultados.

O resultado se confirma pelas curvas DET (Figura 5.14), que também trazem os melhores resultados para LPC e CEP.

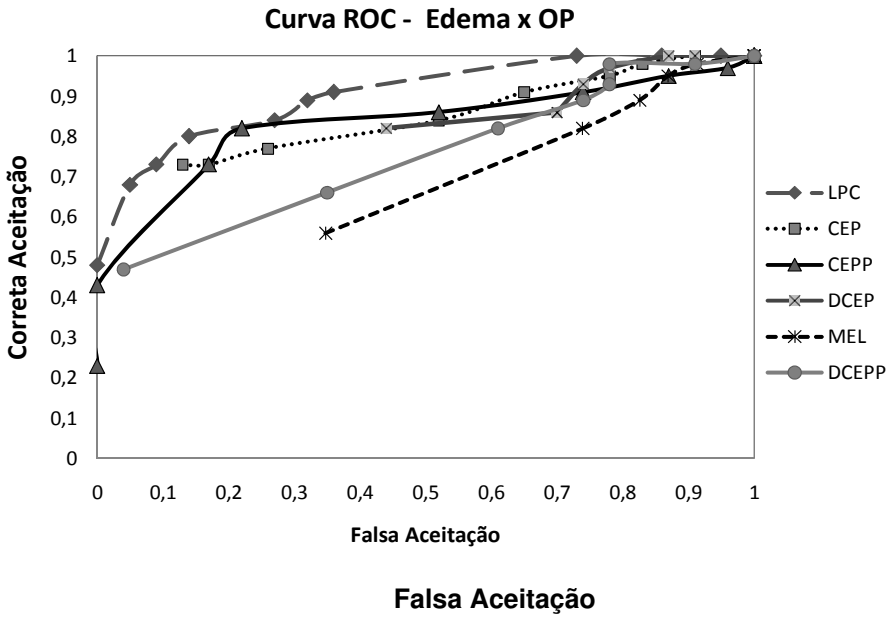



**Figura 5.13 – Curvas ROC para os métodos LPC, CEP, CEPP, DCEP, DCEPP e MEL, para o Caso 2 (Edema x OP).**

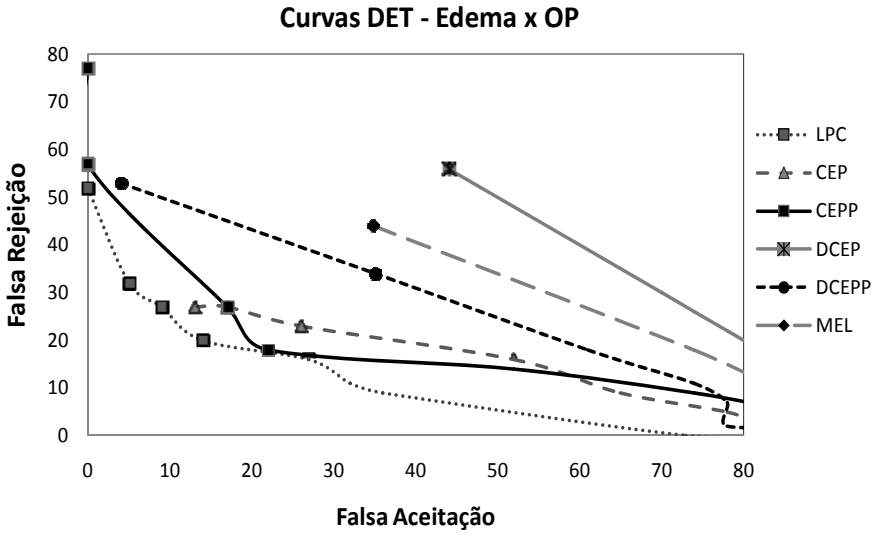



**Figura 5.14 – Curvas DET para os métodos LPC, CEP, CEPP, DCEP, DCEPP e MEL, para o Caso 2.**

**Caso 3:**

Na Tabela 5.3 estão os dados relacionados ao Caso 3, para todos os métodos, nos quais Edema e Outras Patologias estão na mesma classe (Edema +OP) x Normal. Os métodos LPC e MEL proporcionaram os melhores resultados, seguido dos coeficientes cepstrais. Esses métodos, além da maior eficiência, apresentaram as melhores relações entre FA e FR, resultado que se confirma nas curvas ROC (Figura 5.15) e DET (Figura 5.16).

**Tabela 5.3: Medidas de desempenho para os seis métodos, em função de limiares de distorção, avaliando vozes com edema e vozes sob Outras Patologias na mesma classe - (Edema + OP) x Normal.**

| MÉTODO | CR (%) | FA (%) | CA (%) | FR (%) | SP (%) | SE (%) | E (%) |
|---|---|---|---|---|---|---|---|
| LPC | 94 | 6 | 96 | 4 | 94 | 96 | 95 |
| CEP | 91 | 9 | 93 | 7 | 91 | 93 | 92 |
| CEPP | 89 | 11 | 85 | 15 | 89 | 85 | 87 |
| DCEP | 98 | 2 | 79 | 21 | 98 | 79 | 89 |
| DCEPP | 91 | 9 | 75 | 25 | 91 | 75 | 83 |
| MEL | 100 | 2 | 93 | 7 | 98 | 93 | 95 |

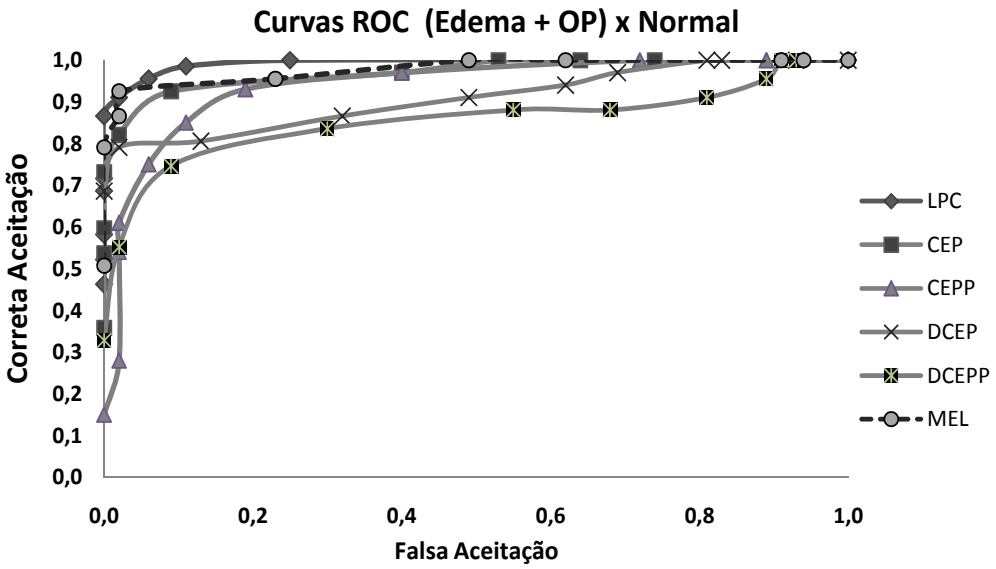



**Figura 5.15 – Curvas ROC para os métodos LPC, CEP, CEPP, DCEP, DCEPP e MEL, para o Caso 3 - (Edema + OP) x Normal.**

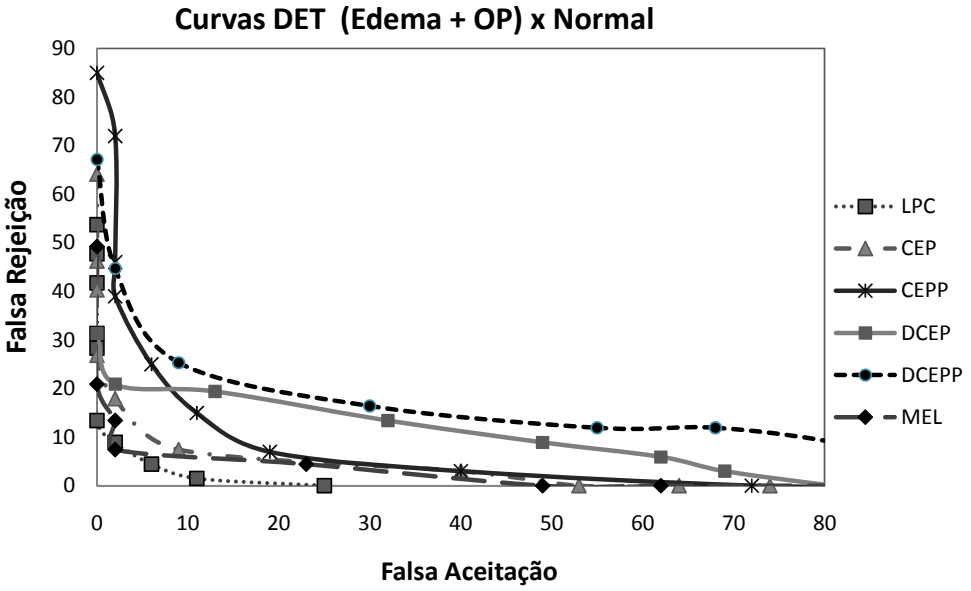



**Figura 5.16 – Curvas DET para os métodos LPC, CEP, CEPP, DCEP, DCEPP e MEL, para o Caso 3.**

Nessa etapa de pré-classificação, em que se usa a medida de distorção do erro médio quadrático mínimo, foram obtidos excelentes resultados para os casos em que é feita a discriminação entre vozes patológicas e vozes normais. Nesses casos (Casos 1 e 3), foram obtidas taxas de discriminação correta igual ou maior que 95%, pelos métodos LPC (99% no Caso 1 e 95% no Caso 3) e MEL (97% no Caso 1 e 95% no Caso 3). Esses métodos, no entanto, apresentaram baixa eficiência no Caso 2, em que foram colocados sinais de vozes afetados por outras patologias (nódulos, cistos e paralisia). Para o método dos coeficientes Mel-cepstrais, obteve-se o pior desempenho (61%).

Dessa forma, é interessante procurar uma forma de melhorar o desempenho desses métodos, buscando-se um refinamento no processo para que haja uma melhor discriminação entre as patologias em estudo. Neste trabalho, foi escolhido o modelamento por meio de Modelos de Markov, por ser um método que já apresenta bons resultados em sistemas de reconhecimento de padrões. Os resultados obtidos são apresentados a seguir.

## 5.6 Resultados obtidos na etapa de refinamento - Classificação usando HMM

Os resultados obtidos na fase de pré-classificação, que não forneceram resultado correto, são testados pelo classificador da etapa de refinamento, baseado em Modelos de Markov Escondidos do tipo discreto. Os padrões de referência já armazenados, obtidos dos vetores de observação de cada um dos métodos empregados, são também separados por casos, como na etapa de pré-classificação. Os resultados são apresentados e analisados a seguir.

Na Tabela 5.4 são apresentados os resultados para o Caso 1, em que são comparadas as classes Edema e Normal, para os métodos empregados.

**Tabela 5.4: Medidas de desempenho obtidas na etapa de refinamento para o Caso 1 (Edema x Normal).**

| Método | CR (%) | FA (%) | CA (%) | FR (%) | SP (%) | SE (%) | E (%) |
|---|---|---|---|---|---|---|---|
| LPC | 100 | 0 | 100 | 0 | 100 | 100 | 100 |
| CEP | 100 | 0 | 93 | 7 | 100 | 93 | 97 |
| DCEP | 100 | 0 | 98 | 2 | 100 | 98 | 99 |
| CEPP | 98 | 2 | 100 | 0 | 98 | 100 | 99 |
| DCEPP | 100 | 0 | 93 | 7 | 100 | 93 | 97 |
| MEL | 100 | 0 | 98 | 2 | 100 | 98 | 99 |

Com a etapa de refinamento, aumenta a eficiência dos métodos empregados, chegando a 100%, no método LPC. As maiores taxas de falsa rejeição foram obtidas nos métodos CEP e DCEP, ficando, no entanto, abaixo de 10%.

Na Tabela 5.5, estão apresentados os dados relativos à avaliação de desempenho dos métodos empregados, com o modelamento por HMM, para o caso 2, em que se considera Edema e Outras Patologias como classes diferentes.

Nota-se que o método LPC continua sendo superior, seguido, nesse caso, do método dos coeficientes cepstrais. Este último apresentando menor taxa de falsa aceitação do que o método LPC. No entanto, o método LPC classificou corretamente 100% dos sinais com Edema. O método DCEP, e o método MEL apresentaram os piores desempenhos, para esse caso. Apesar disso, os resultados são um pouco melhores do que os resultados da pré-classificação.

**Tabela 5.5: Medidas de desempenho obtidas na etapa de refinamento para o Caso 2 (Edema x OP).**

| Método | CR (%) | FA (%) | CA (%) | FR (%) | SP (%) | SE (%) | E (%) |
|---|---|---|---|---|---|---|---|
| LPC | 91 | 9 | 100 | 0 | 91 | 100 | 96 |
| CEP | 96 | 4 | 93 | 7 | 96 | 93 | 95 |
| DCEP | 57 | 43 | 98 | 2 | 57 | 98 | 78 |
| CEPP | 87 | 13 | 100 | 0 | 87 | 100 | 94 |
| DCEPP | 96 | 4 | 93 | 7 | 96 | 93 | 94 |
| MEL | 48 | 52 | 98 | 2 | 48 | 98 | 73 |

Na Tabela 5.6 são apresentados os resultados obtidos para o Caso 3, após a etapa de refinamento. Nesse caso, apenas o método dos coeficientes cepstrais ponderados apresentou taxa de falsa aceitação de apenas 2%. A maior taxa de falsa rejeição foi para o método dos coeficientes delta-cepstrais ponderados (16%). Na maior parte dos métodos, a Eficiência ultrapassou 95%. Esses resultados mostram que os métodos empregando a análise cepstral são eficientes em discriminar vozes patológicas de vozes normais.

**Tabela 5.6: Medidas de desempenho obtidas na etapa de refinamento para o Caso 3 ((Edema+OP)xNormal).**

| Método | CR (%) | FA (%) | CA (%) | FR (%) | SP (%) | SE (%) | E (%) |
|---|---|---|---|---|---|---|---|
| LPC | 100 | 0 | 97 | 3 | 100 | 97 | 99 |
| CEP | 100 | 0 | 91 | 9 | 100 | 91 | 96 |
| DCEP | 100 | 0 | 97 | 3 | 100 | 97 | 99 |
| CEPP | 98 | 2 | 99 | 1 | 98 | 99 | 99 |
| DCEPP | 100 | 0 | 84 | 16 | 100 | 84 | 92 |
| MEL | 100 | 0 | 96 | 4 | 100 | 96 | 98 |

## 5.7 Comparação entre os resultados obtidos em QV e HMM

Na Tabela 5.7 são apresentados os resultados da eficiência obtidos para os casos 1, 2 e 3. Nota-se que há um considerável aumento no desempenho dos métodos empregados em todos os casos, inclusive para o caso 2. A eficiência para esse caso, na etapa de pré-classificação, não chegava a 85%. Com a etapa de refinamento, a eficiência chegou a 96%, para esse caso, no método LPC, apresentando um aumento de 13% em relação à etapa de pré-classificação. Para

o método dos coeficientes delta-cepstrais ponderados, houve um aumento na eficiência de 22%. Para os casos 1 e 3, a eficiência dos métodos empregados ultrapassou 95%, para a maioria dos métodos.

**Tabela 5.7: Valores obtidos para a Eficiência pelos métodos empregados, para a etapa de pré-classificação e para a classificação final, usando HMM.**

| | Eficiência | | | | | |
|---|---|---|---|---|---|---|
| | Caso 1 | | Caso 2 | | Caso 3 | |
| MÉTODO | QV | HMM | QV | HMM | QV | HMM |
| LPC | 99% | 100% | 83% | 96% | 95% | 99% |
| CEP | 90% | 97% | 80% | 95% | 92% | 96% |
| CEPP | 90% | 99% | 80% | 94% | 87% | 99% |
| DCEP | 92% | 99% | 69% | 78% | 89% | 99% |
| DCEPP | 87% | 97% | 72% | 94% | 83% | 92% |
| MEL | 97% | 99% | 61% | 73% | 95% | 98% |

## 5.8 Discussão

Os resultados apresentados neste trabalho mostram que os métodos empregados são eficientes na discriminação entre vozes normais e vozes patológicas.

O método LPC apresenta as mais altas taxas de correta aceitação em todos os casos. Como o trato vocal é saudável, as alterações na voz representadas pelos coeficientes LPC, refletem as variações sofridas pela voz na laringe, mas precisamente nas dobras vocais não-saudáveis. No presente estudo, as dobras vocais atingidas por patologias como edema, nódulos, cistos e paralisia atingem a voz, provocando alterações nos coeficientes LPC e naqueles oriundos da análise cepstral, permitindo a discriminação da voz por esses parâmetros.

Para os casos em que vozes afetadas por Edema e por Outras Patologias (nódulos, cistos e paralisia) são consideradas numa mesma classe, os resultados são similares aos obtidos quando se tenta discriminar vozes com edema e vozes normais, apenas.

Os métodos empregados não se mostraram robustos em discriminar vozes com edemas das vozes com outras patologias (nódulos, cistos e paralisia), quando submetidos à comparação pela medida de distorção após a quantização

vetorial, apenas. Há que se considerar que o sistema foi treinado apenas com vozes afetadas por edema nas dobras vocais. A semelhança da patologia edema com outras patologias interfere no processo de discriminação entre elas. Portanto, para uma maior eficiência do processo de discriminação, torna-se necessário investigar essas similaridades e buscar identificar eventuais características diferenciadoras. O treinamento do sistema com os dados das vozes das Outras Patologias consideradas pode ser um caminho a ser seguido.

Melhores resultados, entretanto, para os casos considerados de baixa eficiência, são alcançados utilizando o classificador baseado em HMM, como refinamento no processo de decisão, para os três casos em análise.

No caso dos coeficientes mel-cepstrais, para discriminar vozes com edema de Outras Patologias, o método apresentou uma eficiência relativamente baixa, quando comparada com os outros casos. Isso pode ser justificado pelo fato de que os coeficientes mel-cepstrais representarem os aspectos perceptuais da fala. No caso das patologias em estudo, há muita similaridade nos aspectos perceptuais como, por exemplo, no aspecto ruidoso, provocando uma voz rouca e soprosa, com dificuldades em sustentação da fala.

Pelos resultados obtidos pelos métodos considerados, observa-se que a caracterização acústica de sinais de vozes patológicas pelo modelo linear de produção de fala é viável e ainda pode ser bastante explorada.

Os métodos empregados, bastante utilizados em sistemas de reconhecimento de padrões, como LPC e Análise Cepstral, podem ser aplicados eficientemente na discriminação de vozes patológicas.

# Capítulo 6
# Considerações finais e Sugestões para Trabalhos Futuros

## 6.1. Introdução

Métodos tradicionais para diagnosticar patologias da laringe, tais como laringoscopia, são considerados invasivos e desconfortáveis para o paciente. Métodos baseados na análise acústica dos sinais de voz têm sido investigados com a finalidade de diminuir o número de exames laringoscópicos e servir como ferramenta auxiliar para terapia vocal, podendo ser utilizados em pré-diagnósticos médicos, avaliações pós-cirúrgicas e tratamentos farmacológicos.

Esses métodos têm sido utilizados para avaliação da qualidade vocal devido à sua simplicidade e natureza não-invasiva nos procedimentos de medição.

Assim, técnicas de análise acústica podem ser úteis para auxiliar a tomada de decisão quanto à presença de uma dada patologia em um sinal de voz. Sistemas de discriminação de vozes patológicas têm sido propostos na literatura, com a utilização de várias técnicas. Algumas das técnicas usadas baseiam-se nas medidas acústicas dependentes diretamente da obtenção correta da freqüência fundamental ou *pitch* (*Jitter*, *Shimmer*, etc.). No entanto, para algumas vozes patológicas fica difícil e, em alguns casos, até impossível a obtenção correta do *pitch*, dada a gravidade da patologia. Dependendo da natureza e grau da patologia, em vozes severamente ruidosas, por exemplo, a detecção do *pitch* fica prejudicada, comprometendo as respectivas medidas.

O grau de confiabilidade e a eficiência no processo de discriminação entre vozes patológicas e vozes normais dependem muito das características e parâmetros da voz usados para a modelagem acústica correspondente e do classificador empregado.

A presença de patologias na laringe, como edemas, nódulos, pólipos e cistos, por exemplo, causam um aumento de massa nas dobras vocais, provocando uma vibração irregular. Isso provoca, ainda, um mau fechamento da glote levando a uma modificação significativa na voz, em especial nos sons sonoros, pela produção de componentes ruidosas adicionais.

Neste trabalho, foi apresentada uma proposta para um método de classificação de vozes patológicas, afetadas por patologias na laringe, mais especificamente por Edema nas dobras vocais.

## 6.2. Resumo da Pesquisa

O processo de classificação empregado utiliza técnicas de predição baseadas no modelo linear de produção de fala. Os parâmetros empregados na análise do sinal de voz são os coeficientes LPC e os coeficientes obtidos pela análise cepstral: coeficientes cepstrais, cepstrais ponderados, delta-cepstrais ponderados e delta-cepstrais. Além disso, utilizou-se uma abordagem não-paramétrica usando os coeficientes mel-cepstrais, obtidos no domínio da freqüência (MFCC – *Mel Frequency Cepstral Coefficients*), que representam bem a falta de fechamento da glote e as vibrações irregulares, devido à presença da patologia.

Os resultados obtidos utilizando os coeficientes LPC, cepstrais e seus derivados, além dos coeficientes mel cepstrais para modelar a patologia em estudo (Edema) foram bastante eficientes, já que a taxa de classificação correta para discriminar voz patológica de voz normal ficou acima de 90%.

A Análise por predição linear e a análise cepstral são indicadas, neste trabalho, para acompanhar as variações que ocorrem na voz, devido à sua passagem pela laringe e pelo trato vocal. Com base no modelo linear de produção da voz, as alterações observadas no comportamento do sinal na saída, embora provocadas por problemas relacionados às questões da excitação, são também decorrentes da ação do trato vocal. Considerando que o trato vocal esteja saudável, o comportamento irregular do sinal em análise é atribuído à patologia presente na laringe, afetando o sinal, portanto, pela modificação da excitação.

Neste trabalho, foram utilizados os coeficientes LPC e os coeficientes cepstrais e mel-cepstrais para analisar o comportamento do sinal de voz patológica provocada por edemas nas dobras vocais, comparando esses parâmetros com os de vozes normais. No processo de avaliação são utilizadas ainda vozes afetadas por outras patologias na laringe como nódulos, cistos e paralisia por serem mais comumente encontradas.

O processo de discriminação das vozes patológicas foi levado a efeito em duas etapas: pré-classificação e classificação final. A etapa de pré-classificação foi realizada utilizando uma medida de distorção (Erro médio quadrático) associada a uma regra de decisão, pela escolha de um limiar que proporcionasse a melhor separação entre as classes (Normal, Edema e Outras Patologias).

Já na etapa de classificação final, foi empregado um classificador baseado em HMM do tipo discreto, esquerda-direita, com cinco estados. Nessa etapa, são classificados apenas os sinais que não foram classificados corretamente na etapa de pré-classificação. O classificador baseado em HMM atua, portanto, como um refinamento no processo de decisão. Uma medida de probabilidade é associada a cada sinal e comparada com um limiar para a tomada de decisão na classificação: patológica ou não-patológica.

Pelos resultados obtidos, os métodos propostos demonstraram uma boa eficiência em discriminar entre vozes afetadas por patologias na laringe de vozes normais. Foram realizados, ainda, testes com as outras patologias citadas. Na etapa de pré-classificação, usando quantização vetorial associada a uma medida de distorção, foram obtidos resultados de até 99% de classificação correta em se tratando da discriminação entre vozes normais e vozes patológicas. Entretanto, os resultados, nessa etapa, indicam que o sistema não se mostrou tão eficiente em discriminar vozes afetadas por edema das vozes afetadas por nódulos, cistos e paralisia (Outras Patolologias).

Os coeficientes LPC e mel-cepstrais proporcionaram os melhores resultados, seguidos dos coeficientes cepstrais, na discriminação de vozes patológicas e vozes normais. Os coeficientes LPC se destacam pela eficiência nos três casos em análise, apresentando os maiores valores para as taxas de correta aceitação e correta rejeição. As mudanças que a patologia provoca na excitação foram bem traduzidas pelos coeficientes LPC, dado que, considerando o trato vocal como saudável, as mudanças provenientes da excitação atingem o sinal de voz, provocando um sinal de saída desordenado ou degradado. A desordem é captada por esses coeficientes de forma bem significativa.

Os resultados obtidos na etapa de pré-classificação obtiveram um melhoramento significativo na etapa de refinamento. No Caso 1, obteve-se um melhoramento de até 10% na taxa de classificação correta (Eficiência). No Caso 2, houve um aumento de até 22% (delta-cepstrais ponderados) na taxa de classificação correta e no Caso 3 de até 12%. É importante destacar que as

referidas taxas nos casos 1 e 3, em sua maioria ficaram acima de 95% de classificação correta.

Os coeficientes mel-cepstrais apresentaram excelentes resultados nos Casos 1 e 3. Ou seja, mostraram-se eficientes em discriminar entre voz normal e voz patológica. Não foram muito eficientes, no entanto, no caso em que se colocou o edema e as outras patologias em classes diferentes. Isso é compreensível, devido à característica inerente destes coeficientes em capturar os aspectos perceptuais da voz humana, dado que estes aspectos são bem similares nas patologias em estudo (rouquidão, soprosidade, dificuldade de manter a vogal sustentada, voz ruidosa).

A análise cepstral mostra-se eficiente em discriminar voz patológica e voz normal. O desempenho para o Caso 2, que é o caso mais crítico da pesquisa em questão, foi melhor para os coeficientes cepstrais, cepstrais ponderados e delta-cepstrais ponderados. Esses coeficientes conseguem capturar bem as desordens vocais provocadas pelas patologias e capturam melhor as diferenças entre Edema e as Outras Patologias, tanto na etapa de pré-classificação quanto no uso do HMM, quando os resultados melhoraram significativamente para esses parâmetros.

## 6.3. Contribuições

As técnicas de análise paramétrica (LPC e Cepstral) e não-paramétrica (Mel cepstral – MFCC), empregadas neste trabalho são técnicas já usuais em sistemas de reconhecimento de voz e de locutor. Para a discriminação de vozes patológicas, no entanto, esses métodos têm sido pouco empregados, não tendo sido explorado, ainda, o seu potencial discriminativo de desordens vocais. A análise acústica da patologia edema nas dobras vocais, usando o modelo linear de produção de fala, com os coeficientes LPC, cepstrais e seus derivados, não é abordada como nesta pesquisa.

A partir da análise acústica realizada (Capítulo 3), constataram-se diferenças significativas no espectro LPC, no cepstro, na energia, freqüência fundamental, formantes e energia segmental dos sinais de vozes patológicas comparadas com vozes normais. Além disso, a dificuldade de obtenção do *pitch*,

para sinais com patologias severas, sugere a busca de métodos que não dependam da obtenção desta medida. Os métodos usualmente empregados em análise acústica de vozes patológicas utilizam medidas dependentes do *pitch*. Essas características dos sinais foram bem documentadas neste trabalho e analisadas de tal forma que justifica bem o uso das Análises LPC e Cepstral como um bom caminho para a implementação de um sistema automático de discriminação de vozes patológicas que utilizem estas medidas.

Não foi encontrado, na literatura, trabalhos relacionados à discriminação de vozes patológicas empregando a técnica de Quantização Vetorial, associada a HMM como etapa de refinamento do processo de classificação. Além disso, a caracterização acústica da patologia Edema nas dobras vocais não é apresentada em detalhes como nesta pesquisa. A maior parte dos trabalhos limita-se a discriminar entre voz normal e vozes patológicas, sem investir numa patologia específica.

As observações de que as patologias que apresentam aspectos similares precisam de um tratamento mais refinado para a classificação, são relevantes. Isso foi verificado pelos resultados do Caso 2, destacado no estudo. As pesquisas na área de detecção automática de patologias da laringe precisam levar esse fato em consideração, para que proporcionem um bom desempenho.

## 6.4. Sugestões para Trabalhos Futuros

A partir da análise dos resultados obtidos, sugere-se:

- A ampliação da base de dados com vozes afetadas por outras patologias nas dobras vocais para tentar melhorar o desempenho do sistema na discriminação entre patologias;
- Utilizar o método apresentado para a discriminação de outras patologias que não Edema, bastando para isso o treinamento do sistema com sinais afetados por patologias de interesse;
- O desenvolvimento de um sistema automático de detecção de patologias na laringe utilizando a análise LPC e Cepstral para discriminar vozes patológicas, com a patologia Edema de vozes normais, com uma interface gráfica interativa;

- A investigação do desempenho de outros métodos de análise para as mesmas patologias em estudo, além de outras similares, utilizando a análise dinâmica não-linear e teoria do caos;
- O uso de outros classificadores, como os baseados em Modelos de Misturas Gaussianas (GMM), *Support Vector Machine* (SVM), Redes Neurais Artificiais, entre outros.

# Referências Bibliográficas


1. ADNENE, C., LAMIA, B., MOUNIR, M. Analysis of Pathological Voices by Speech Processing. *Proceedings of the Seventh International Symposium on on Signal Processing and Its Aplications*, IEEE, 365- 367, Vol.1, July, 2003.

2. AGUIAR NETO, B. G., COSTA, S. C., FECHINE, J. M., MUPPA, M., Acoustic Features of Disordered Voices Under Vocal Fold Pathology. *19th International Congress on Acoustics* (ICA'07), Madrid, September 2007a. (http://www.sea-acustica.es/WEB_ICA_07/fchrs/papers/cas-03-003.pdf).

3. AGUIAR NETO, B. G., *Signal Aufbereitung in Digitalen Sprachübertragungssystemen*. Doctor-Thesis, Technische Universität Berlin, Germany, 1987.

4. AGUIAR NETO, B. G., FECHINE, J. M., COSTA, S. CUNHA, MUPPA, M., Feature Estimation for Vocal Fold Edema Detection Using Short-Term Cepstral Analysis. *Proceedings of the 7th International Conference on Bioinformatics and Bioengineering*, 14-17 Oct., page(s) 1158-1162, 2007b.

5. AGUIAR NETO, B. G., COSTA, S. C., FECHINE, J. M., LPC Modelling and Cepstral Analysis Applied to Vocal Fold Pathology Detection. International Journal of Functional Informatics and Personalised Medicine, Vol. 1, No 2, pp 156-170, september, 2008.

6. ANDREÃO, R. V. *Implementação em Tempo Real de Um Sistema de Reconhecimento de Dígitos Conectados*. Universidade Estadual de Campinas - Faculdade de Engenharia Elétrica e de Computação. Dissertação de Mestrado. Janeiro 2001.

7. ARAÚJO, A. M. L. *Jogos Computacionais Fonoarticulatórios para Crianças com Deficiência Auditiva*. Tese de doutorado – Universidade Estadual de Campinas. São Paulo, 2000.

8. BAHOURA, M., PELLETIER, C. Respiratory Sound Classification using Cepstral Analysis and Gaussian Mixture. *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, CA, USA, September, 2004.

9. BEHLAU, M., PONTES, P. *Exame laringológico. In: Avaliação e tratamento das disfonias*. São Paulo. Lovise; 1995.

10. BENJAMIN, B. *Cirurgia Endolaríngea*. Editora Revinter: Rio de Janeiro, 2000.

11. BENTO, R.F. and MINITI, A. Doenças Benignas da Laringe e Alterações da Voz e Fonocirurgias – *Revistas Seminários em Otorrinolaringologias*, Ano 3, n. 7 . São Paulo, 1997.

12. BOONE, D. *Sua Voz Está Traindo Você? Como Encontrar e Usar sua Voz Natural*. Artes Médicas. Porto Alegre, 1996.

13. BOYANOV, B., IVANOV, T., HADJITODOROV, S., and CHOLLET, G., Robust hybrid pitch detector. *Electronic Letters*, vol. 29, no. 22, pp. 1924-1926, 1993.

14. COLTON. R. & CASPER, J. Conduta médica e cirúrgica dos distúrbios vocais. In: *Compreendendo os Problemas de Voz*, Porto Alegre, Artes médicas, 1996.

15. COSTA , S. C., AGUIAR NETO, B. G., FECHINE, J. M. and MUPPA, M. Short-Term Cepstral Analysis Applied to Vocal Fold Edema Detection. *Proceedings of BIOSIGNALS 2008 – International Conference on Bio-inspired Systems and Signal Processing*. Portugal, January, 2008a.

16. COSTA , S. C., AGUIAR NETO, B. G., FECHINE, CORREIA, S. Parametric Cepstral Analysis for Pathological Voice Assessment. *Proceedings of The 23rd ACM Symposium on Applied Computing 2008 (ACM SAC' 2008). Computer Applications in Health Care Track,* Pages 1410-1414, Fortaleza, Ceará, Brazil, March 16-20, 2008b.

17. COSTA, S. C., FALCÃO, H., ALMEIDA, N., AGUIAR NETO, B. G., FECHINE, CORREIA, S. AGUIAR NETO, B. G., FECHINE. Pathological Voice Discrimination based on Entropy Measurements. *Proceedings of The 23rd ACM Symposium on Applied Computing 2008 (ACM SAC' 2008).* Computer Applications in Health Care Track, Pages 1424-1426, Fortaleza, Ceará, Brazil, March 16-20, 2008c.

18. COSTA, S. C., AGUIAR NETO, B. G., FECHINE, J. M. Pathological Voice Discrimination using Cepstral Analysis, Vector Quantization and Hidden Markov Models, *8th International Conference on Bioinformatics and Bioengineering (BIBE'08)*, October, 2008.

19. COSTA, W. C. de A. *Reconhecimento de Fala Utilizando Modelos de Markov Escondidos (HMM's) de Densidades Contínuas*. Universidade Federal da Paraíba- Dissertação de Mestrado, Junho 1994.

20. DAJER, M. E. *Padrões Visuais de Sinais de Voz através de Técnica de Análise de Não-Linear*. Dissertação. Bioengenharia, Escola de Engenharia de São Carlos, São Paulo, 2006.

21. DANIEL, R., BOONE, S., McFARLANE, C. *A Voz e a Terapia Vocal*. Artes Médicas. Porto Alegre, 1994.

22. DAVIS, S. B. Acoustic Characteristics of Normal and Pathological Voices. *Speech and Language: Advances in Basic Research and Practice*, Vol. 1, pp. 271–335, 1979.

23. DELLER Jr. R., PROAKIS, J. G., and HANSEN, J. H. L. *Discrete-time Processing of Speech Signals*. Macmillan Publishing Co., 1993.

24. DIAS, R. de S. F. *Normalização de Locutor em Sistema de Reconhecimento de Fala*. Universidade Estadual de Campinas – Faculdade de Engenharia Elétrica e de Computação. *Dissertação de Mestrado*, Novembro 2000.

25. DIBAZAR, A. A., BERGER, T.W., and NARAYANAN, S. S. Pathological Voice Assessment*. Proceedings of the 28th IEEE EMBS Annual International Conference*, New York, USA, pp. 1669-1673, August, 2006.

26. DIBAZAR, A. A., BERGER, T.W., NARAYANAN, S. S. Pathological Voice Assessment*. Proceedings of the 28th IEEE EMBS Annual International Conference*, New York, USA, Aug., pp. 1669-1673, 2006.

27. DIBAZAR. A. A., NARAYANAN. "Feature Analysis for Automatic Detection of Patological Speech". *Proceedings of the Second Joint Conference on EMBS/BMES*. Houston, Volume 1, pages 182- 183. October 2002.

28. ESPINOSA, C. H., M. F., GOMEZ, V. P. Diagnosis of Vocal and Voice Disorders by the Speech Signal. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, Volume: 4, pages 253-258, vol.4, 2000.

29. FAGUNDES, R. D. R. and ALENS, N. Reconhecimento de Voz, Linguagem Contínua, Usando Modelos de Markov. *11º Simpósio Brasileiro de Telecomunicacões* - SBT, Setembro 1993.

30. FECHINE, J. M. *Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística*. Tese de Doutorado. Universidade Federal da Paraíba, 2000.

31. FEIJOO, S. and HERNÁNDEZ, C., Short-Term Stability Measures for the Evaluation of Vocal Quality. *J. Speech, Hearing Res.*, Vol. 33, pages 324–334, Jun. 1990.

32. FONSECA, E. S., GUIDO, R. C., SILVESTRE, A. C. and PEREIRA, J. C.et al. Discrete Wavelet Transform and Support Vector Machine Applied to Pathological Voice Signals Identification. *Proceedings of the Seventh IEEE International Symposium on Multimedia* (ISM'05), Dec., 2005.

33. FORNEY, G. D. The Viterbi Algorithm. Proceedings of the IEEE, Vol. 61, pages 268-278, 1973.

34. FREDOUILLE, C. Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). *9th European Conference on Speech Communication and Technology* (Interspeech), Lisboa, Portugal, September 2005.

35. FUKS, L. and SUNDBERG, J. Using respiratory inductive plethysmography for monitoring professional reed instrument performance. *Medical Problems of Performing artists.* Hanley 7 Belfus, Inc., Philadelphia, PA, 1999.

36. GARCIA, B. et al. Multiplatform Interface Adapted to Pathological Voices. *IEEE International Symposium on Signal Processing and Information Technology*, p. 912-917, Dec., 2005.

37. GARCIA, B., VICENTE, J, RUIZ, I., and ALONSO, A. Multiplatform Interface Adapted to Pathological Voices. *IEEE International Symposium on Signal Processing and Information Technology*, pages 912-917, December, 2005.

38. GAVIDIA-CEBALLOS, L. and HANSEN, J. H. L. Direct Speech Feature Estimation Using an Interactive EM Algorithm for Vocal Fold Pathology Detection, *IEEE Trans. on Biomedical Engineering*, Vol. 43, No. 4, April.

39. GODINO-LLORENTE, J. I., GÓMEZ-VILDA, P., BLANCO VELASCO, M. Dimensionality Reduction of a Patholgical Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters. *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 10, October, 2006.

40. HIRANO, M. *Laryngeal Histopathology*. In COLTON, R., CASPER, J. Understanding Voice Problems. A Physiological Perspective of the Diagnosis and Treatment. 2th Ed. Baltimor: Williams & wilkins, 1996.

41. HIRANO, M., BLESS, D.M. *Videoestroboscopic Examination of the Larynx*. San Diego: Singular Publishing Group Inc., 1993.

42. JIANG, J. J., ZHANG, Y. and McGILLIGAN, C. Chaos in Voice, From Modeling to Measurement. *Journal of Voice*, Vol. 20, No. 1, pages 2-17, 2006.

43. KAY ELEMETRICS, Kay Elemetrics Corp. Disordered Voice Database1.03 ed. 1994.

44. KROM G. de. A cepstrum-based Technique for Determining a Harmonics-to-noise Ratio in Speech Signals. *J. Speech, Hearing Res*., vol. 36, n. 2, pages 254–266, Apr. 1993.

45. KUHL, I. *Manual Prático de Laringologia*. Livro Texto/II. Editora da Universidade – UFRGS. Porto Alegre, 1982

46. LEVINSON, S. E., RABINER, L. R., and SONDHI M. M. An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition. *The Bell System Technical Journal*, Vol. 62, No. 4, pages 1035-1068, April 1983.

47. LI, Tao, JO, C. Discrimination of Severely Noisy Pathological Voice with Spectral Slope and HNR. *Proceedings of the 7th International Conference on Speech Processing of IEEE* (ICSP'04). Vol. 3, pages 2218- 2221, Sept. 2004.

48. LINDE, Y., BUZO, A., and GRAY, R.M. An Algorithm for Vector Quantizer Design. *IEEE Transactions on Communications*, Vol. COM - 28, No. 1, pages 84-95, January 1980.

49. MAKHOUL, J., ROUCOS, S., and GISH, H. Vector Quantization in Speech Coding. *Proceedings of the IEEE*, Vol. 73, No. 11, pages 1551-1588, November 1985.

50. MAMMONE, R. J., ZHANG, X., and RAMACHANDRAN, R. P. Robust Speaker Recognition - A Feature-Based Approach. *IEEE Signal Processing Magazine*, Vol. 13, No. 5, pages 58-71, September 1996.

51. MANFREDI, C. Adaptive Noise Energy Estimation in Pathological Speech Signals. *IEEE Transactions on Biomedical Engineering*, Vol. 47, No. 11, November 2000.

52. MANFREDI, C., PIERAZZI, L., and BRUSCAGLIONI, P. Pitch Estimation For Noise Retrieval in Time and Frequency Domain. *Med. Biol. Eng. Comput.*, vol. 37, n. 2, I, pages 532–533, 1999.

53. MARINAKI, M., CONTROPOULOS, C., PITAS, I. and MAGLAVERAS, N. Automatic Detection of Vocal Fold Paralysis and Edema". Proc. Of 8th Conf. Spoken Language Processing (Interspeech 2004), Jeju, Korea, October, 2004.

54. MARINAKI, Maria et al. Automatic Detection of Vocal Fold Paralysis and Edema. *Proc. Of 8th Conf. Spoken Language Processing* (Interspeech 2004), Jeju, Krea, October, 2004.

55. MARKEL, J.D. and GRAY, A. H. Jr. *Linear Prediction of Speech*, Springer Verlag, Berlin, 1976.

56. MARTIN, A., DODDINGTON, G., KAMM, T., ORDOWSKI, M. and PRZYBOCKI, M. The DET Curve in Assessment of Detection Task Performance, *Proceedings of Eurospeech*, Vol. 4, pages 1895-1898, 1997.

57. MARTINEZ, C. E., RUFINER, H. L., Acoustic Analysis of Speech for Detection of laryngeal Pathologies. *Proceedings of the 22th Annual EMBS Conference,* pages 2369-2372, Chicago, July, 2000.

58. MERGELL, P., HERZEL, H. and TITZE, I. R. Irregular Vocal-fold Vibration – High-speed observation and modeling. Journal Acoustic Society America, 108 (6), December, 2000.

59. MICHAELIS D., GRAMSS, T., and STRUBE, H. W., Glottal-to-noise Excitation Ratio – A new measure for describing pathological voices. Acustica/Acta Acustica, vol. 83, pages 700–706, 1997.

60. MURPHY, P. J. and AKANDE, OLATUNJI O., 2007. Noise Estimation in Voice Signals Using Short-term Cepstral. *Journal of the Acoustical Society of America*, pages 1679-1690, Vol. 121, No. 3, March, 2007.

61. NOLL, A. M. Cepstrum Pitch Determination. *Journal of the Acoustic Society America*. Vol. 41, pages 293-309, 1967.

62. O'SHAUGHNESSY, D. *Speech Communications: Human and Machine*, 2nd Edition, NY, IEEE Press, 2000.

63. PAPARELLA, M.M., SHUMRICK, D.A - *Otorrinolaringologia - Cabeza y CueIlo* – 2ª ed. Buenos Ares: Panamericana, 1982.

64. PARRAGA, A. *Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas*. Universidade Federal do Rio Grande do Sul. Dissertação de Mestrado, 2002.

65. PARSA, V. and JAMIESON, D. G. Acoustic Discrimination of Pathological Voice: Sustained Vowels versus Continuous Speech. *Journal of Speech, Language, and Hearing Research*, Vol. 44, April, pp 327–339.

66. Qi, Y. and HILLMAN, R. E., Temporal and Spectral Estimations of Harmonics-to-Noise Ratio in Human Voice Signals, *Journal of the Acoustic Society America*, Vol. 102, n. 1, pages 537–543, 1997.

67. QUEK, F., HARPER, M. Y., HACIAHMETOGLOU, L. C., and RAMING, L. O. Speech pauses and gestural holds in Parkinson's disease. *Proceedings of International Conference on Spoken Language Processing*, pp. 2485-2488, 2002.

68. RABINER L. R. and JUANG B. H. *Fundamentals of Speech Recognition.* Englewood Ciffs, New Jersey: Prentice Hall, 1993.

69. RABINER, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, Vol. 77, No. 2, pages 257-286, February, 1989.

70. RABINER, L. R. and SCHAFER, R. W. *Digital Processing of Speech Signals*. Prentice Hall, Upper Saddle River, New Jersey, 1978.

71. RABINER, L. R., JUANG, B. H., LEVINSON, S. E., and SONDHI, M. M. Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities. *AT & T Technical Journal*, Vol. 64, No. 6, pages 1211-1234, July-August 1985.

72. RABINER, L. R., JUANG, B. H., LEVINSON, S. E., and SONDHI, M. M. Recognition of Isolated Digits Using Hidden Markov Models with Continuos Mixture Densities. *AT & T Technical Journal*, Vol. 64, No. 6, pages 1211-1234, July-August 1985.

73. RABINER, L. R., LEVINSON, S. E., and SONDHI, M. M. On the Application of Vector Quantization and Hidden Markov Models to Speaker-independent, Isolated Word Recognition. *The Bell System Technical Journal*, Vol. 62, No. 4, pages 1075-1105, April 1983.

74. ROSA M. de O., PEREIRA, J. C., and Grellet, M. Adaptive estimation of Residue Signal for Voice Pathology Diagnosis, *IEEE Trans. on Biomedical Engineering*, Vol. 47, No. 1, pp. 96-104, 2000.

75. RUSSO, I. C. P. & BEHLAU, M. *Percepção da fala: análise acústica do português brasileiro*. São Paulo: Lovise, 1993.

76. SATISH, L. and GURURAJ, B. I. Use of Hidden Markov Models for Partial Discharge Pattern Classification. *IEEE Transactions on Electrical Insulation*, Vol. 28, No. 2, pages 172-182, April 1993.

77. SAVIC, M. and GUPTA, S. K. Variable Parameter Speaker Verification System Based on Hidden Markov Modeling. *Proceedings of the IEEE International*

*Conference on Acoustics, Speech, and Signal Processing* (ICASSP'90), pages 281-284, 1990.

78. SHAMA, K., KRISHNA, A., and CHOLAYYA N. U., Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology, EURASIP Journal on Advances in Signal Processing, Vol. 2007.

79. SONDHI, M. M. New Methods of Pitch Extraction. *IEEE Transactions on Audio and Electroacoustics*. AU-16(2):262-266, Jun. 1968.

80. TEIXEIRA, J. P. R., *Modelização Paramétrica de Sinais para Aplicação em Sistemas de Conversão texto-fala*. Dissertação de Mestrado. Universidade do Porto, faculdade de Engenharia, 1995.

81. UMAPATHY, K. KRISHNAN, S., PARSA, V., and JAMIESON, D. G. Discrimination of Pathological Voices Using a Time-Frequency Approach. *IEEE Transactions On Biomedical Engineering*, Vol. 52., No. 3, March, 2005.

82. WESSEL, F.; SCHLUTER, R.; MACHEREY, K. and NEY, H. Confidence Measures for Large Vocabulary Continuous Speech Recognition. *Speech and Audio Processing, IEEE Transactions on*, Volume 9, Issue 3, pages 288-298, 2001.

83. WINHOLTZ, W. Vocal Tremor Analysis with the Vocal Demodulator. J*. Speech, Hearing Res*., no. 35, pp. 562–563, 1992.

84. YIN, T., CHIU, N., Discrimination between Alzheimer's Dementia and Controls by AUTOMATED ANALYSIS of Statistical Parametric Maps of Satistical Parametric Maps of 99m. Tc-HMPAO-SPECT Volumes. *Proceedings of the Fourth IEEE Symposium on Bioinformatics and Bioengineering*, BIBE'04, pages 183-190, May 2004.

85. YUMOTO, E., GOULD, W. and BAER, T. Harmonics-to-Noise Ratio as an Index of The Degree of Hoarseness, *Journal of the Acoustic Society America*, Vol. 71, no. 6, pages 1544–1550, 1982.

86. ZHANG, Y. AND JIANG, J. J. Chaotic Vibrations of a Vocal Fold Model with a Unilateral Polyp. *Journal of the Acoustic Society America*, 115 (3), March, 2004.

87. ZHANG, Y., JIANG, J,J. and FEROZE, F.A. Wavelet-based Denoising for Improving Nonlinear Dynamic Analisys of Pathological Voices. *Eletronic Letters 4th*, Vol. 42, No. 16, August, 2005.

88. ZHANG, Y., McGILLIGAN C., ZHOU, L., VIG, M. and JIANG, J. Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps. *Journal of the Acoustic Society America*, 115 (5), May, 2004.

89. ZITTA, S. M. Análise Perceptivo-Auditiva e Acústica em Mulheres com Nódulos Vocais. Centro Federal de Educação Tecnológica – CEFET-PR. Curitiba, Paraná, 2005.

90. ZWETSCH, I. C., RIBEIRO, R. D., FAGUNDES, T. R., SCOLARI, D. Processamento Digital de Sinais no Diagnóstico Diferencial de Doenças Laríngeas Benignas. *Scientia Medica*, Porto Alegre: PUCRS, Vol. 16, n. 3, jul./set. 2006.

# ANEXO A

# Informações dos sinais de vozes da base de dados utilizada

**INFORMAÇÕES DE PACIENTES COM VOZES NORMAIS**

| Nº | FILE VOWEL 'AH' | AGE | SEX | SMOKE | NATLANG | ORIGIN |
|---|---|---|---|---|---|---|
| 1 | AXH1NAL.NSP | 29 | F | N | English | White- not Hispanic |
| 2 | BJB1NAL.NSP | 34 | M | N | English | White- not Hispanic |
| 3 | BJV1NAL.NSP | 52 | F | N | English | White- not Hispanic |
| 4 | CAD1NAL.NSP | 31 | F | N | English | White- not Hispanic |
| 5 | CEB1NAL.NSP | 43 | F | N | English | White- not Hispanic |
| 6 | DAJ1NAL.NSP | 26 | F | N | English | White- not Hispanic |
| 7 | DFP1NAL.NSP | 34 | F | N | English | White- not Hispanic |
| 8 | DMA1NAL.NSP | 24 | F | N | English | White- not Hispanic |
| 9 | DWS1NAL.NSP | 32 | M | N | English | White- not Hispanic |
| 10 | EDC1NAL.NSP | 32 | F | N | English | White- not Hispanic |
| 11 | EJC1NAL.NSP | 44 | M | N | English | White- not Hispanic |
| 12 | FMB1NAL.NSP | 28 | M | N | English | White- not Hispanic |
| 13 | GPC1NAL.NSP | 40 | M | N | English | White- not Hispanic |
| 14 | GZZ1NAL.NSP | 47 | M | N | English | White- not Hispanic |
| 15 | HBL1NAL.NSP | 25 | F | N | English | White- not Hispanic |
| 16 | JAF1NAL.NSP | 31 | F | N | English | White- not Hispanic |
| 17 | JAN1NAL.NSP | 30 | F | N | English | White- not Hispanic |
| 18 | JAP1NAL.NSP | 40 | F | N | English | White- not Hispanic |
| 19 | JEG1NAL.NSP | 26 | F | N | English | White- not Hispanic |
| 20 | JMC1NAL.NSP | 45 | M | N | English | White- not Hispanic |
| 21 | JTH1NAL.NSP | 31 | F | N | English | White- not Hispanic |
| 22 | JXC1NAL.NSP | 43 | F | N | English | White- not Hispanic |
| 23 | KAN1NAL.NSP | 55 | M | N | English | White- not Hispanic |
| 24 | LAD1NAL.NSP | 40 | F | N | English | White- not Hispanic |
| 25 | LDP1NAL.NSP | 22 | F | N | English | White- not Hispanic |
| 26 | LLA1NAL.NSP | 30 | F | N | English | White- not Hispanic |
| 27 | LMV1NAL.NSP | 43 | F | N | English | White- not Hispanic |
| 28 | LMW1NAL.NSP | 45 | F | N | English | White- not Hispanic |
| 29 | MAS1NAL.NSP | 37 | M | N | English | White- not Hispanic |
| 30 | MCB1NAL.NSP | 28 | F | Y | English | White- not Hispanic |
| 31 | MFM1NAL.NSP | 28 | M | N | English | White- not Hispanic |
| 32 | MJU1NAL.NSP | 26 | M | N | English | White- not Hispanic |
| 33 | MXB1NAL.NSP | 24 | F | N | English | White- not Hispanic |
| 34 | MXZ1NAL.NSP | 28 | F | N | English | White- not Hispanic |
| 35 | NJS1NAL.NSP | 39 | F | Y | English | White- not Hispanic |
| 36 | OVK1NAL.NSP | 29 | M | N | English | White- not Hispanic |
| 37 | PBD1NAL.NSP | 40 | F | N | English | White- not Hispanic |
| 38 | PCA1NAL.NSP | 36 | M | N | English | White- not Hispanic |
| 39 | RHM1NAL.NSP | 40 | M | N | English | White- not Hispanic |
| 40 | RJS1NAL.NSP | 46 | M | N | English | White- not Hispanic |
| 41 | SCK1NAL.NSP | 33 | F | N | English | White- not Hispanic |

| № | FILE VOWEL 'AH' | AGE | SEX | SMOKE | NATLANG | ORIGIN |
|---|---|---|---|---|---|---|
| 42 | SCT1NAL.NSP | 39 | F | N | English | White- not Hispanic |
| 43 | SEB1NAL.NSP | 37 | F | N | English | White- not Hispanic |
| 44 | SIS1NAL.NSP | 36 | M | N | English | White- not Hispanic |
| 45 | SLC1NAL.NSP | 22 | F | N | English | White- not Hispanic |
| 46 | SXV1NAL.NSP | 38 | M | N | English | White- not Hispanic |
| 47 | TXN1NAL.NSP | 39 | M | Y | English | White- not Hispanic |
| 48 | VMC1NAL.NSP | 44 | F | N | English | White- not Hispanic |
| 49 | DJG1NAL.NSP | 37 | M | N | English | White- not Hispanic |
| 50 | JKR1NAL.NSP | 43 | F | N | English | White- not Hispanic |
| 51 | MAM1NAL.NSP | 39 | F | N | English | White- not Hispanic |
| 52 | WDK1NAL.NSP | 39 | M | N | English | White- not Hispanic |
| 53 | RHG1NAL.NSP | 59 | M | N | English | White- not Hispanic |

**INFORMAÇÕES DE PACIENTES COM EDEMA NAS DOBRAS VOCAIS**

| Nº | PAT_ID | FILE VOWEL'AH' | AGE | SEX | LOCATION | SMOKE | NATLANG | ORIGIN |
|---|---|---|---|---|---|---|---|---|
| 1 | ANA000 | ANA15AN.NSP | 71 | F | bilateral | Y | Armenian | White- not Hispanic |
| 2 | ANB000 | ANB28AN.NSP | 18 | F | bilateral | N | English | White- not Hispanic |
| 3 | CAC000 | CAC10AN.NSP | 49 | F | bilateral | | English | White- not Hispanic |
| 4 | CAK000 | CAK25AN.NSP | 47 | F | unilateral left | Y | English | White- not Hispanic |
| 5 | CER000 | CER16AN.NSP | 45 | F | unilateral left | Y | English | White- not Hispanic |
| 6 | CTB000 | CTB30AN.NSP | 36 | M | | N | English | White- not Hispanic |
| 7 | DBF000 | DBF18AN.NSP | 25 | F | bilateral | Y | English | White- not Hispanic |
| 8 | DJF000 | DJF23AN.NSP | 45 | F | bilateral | N | English | White- not Hispanic |
| 9 | DMG000 | DMG07AN.NSP | 24 | M | bilateral | N | English | White- not Hispanic |
| 10 | DXC000 | DXC22AN.NSP | 43 | M | bilateral | | English | White- not Hispanic |
| 11 | EED000 | EED07AN.NSP | 30 | F | bilateral | Y | English | White- not Hispanic |
| 12 | EXE000 | EXE06AN.NSP | 57 | F | bilateral | Y | English | White- not Hispanic |
| 13 | HLM000 | HLM24AN.NSP | 36 | F | bilateral | Y | English | White- not Hispanic |
| 14 | JAJ000 | JAJ31AN.NSP | 17 | F | bilateral | N | English | White- not Hispanic |
| 15 | JJD000 | JJD29AN.NSP | 23 | M | bilateral | | English | White- not Hispanic |
| 16 | JMC000 | JMC18AN.NSP | 38 | F | bilateral | | English | White- not Hispanic |
| 17 | JMH000 | JMH22AN.NSP | 59 | F | bilateral | Y | English | White- not Hispanic |
| 18 | JXB000 | JXB16AN.NSP | 63 | M | bilateral | Y | English | White- not Hispanic |
| 19 | JXC000 | JXC21AN.NSP | 42 | F | bilateral | | English | White- not Hispanic |
| 20 | JXF001 | JXF11AN.NSP | 34 | F | bilateral | N | English | White- not Hispanic |
| 21 | JXS002 | JXS09AN.NSP | 60 | F | bilateral | N | English | White- not Hispanic |
| 22 | KAB000 | KAB03AN.NSP | 31 | F | bilateral | N | English | White- not Hispanic |
| 23 | KLC000 | KLC09AN.NSP | 46 | F | bilateral | | English | White- not Hispanic |
| 24 | LAC000 | LAC02AN.NSP | 25 | F | bilateral | | English | White- not Hispanic |
| 25 | LAD000 | LAD13AN.NSP | 41 | F | bilateral | | English | White- not Hispanic |
| 26 | LGM000 | LGM01AN.NSP | 32 | F | | N | English | Black- not Hispanic |
| 27 | LXC001 | LXC01AN.NSP | 33 | F | bilateral | N | Cambodian | Asian or Pacific Islander |
| 28 | LXD000 | LXD22AN.NSP | 85 | F | unilateral left | | English | White- not Hispanic |

| Nº | PAT_ID | FILE VOWEL'AH' | AGE | SEX | LOCATION | SMOKE | NATLANG | ORIGIN |
|---|---|---|---|---|---|---|---|---|
| 29 | MCA000 | MCA07AN.NSP | 37 | F | | N | Portuguese | White- not Hispanic |
| 30 | MCW001 | MCW21AN.NSP | 39 | F | bilateral | N | English | White- not Hispanic |
| 31 | NFG000 | NFG08AN.NSP | 49 | F | bilateral | | English | White- not Hispanic |
| 32 | NLC000 | NLC08AN.NSP | 48 | F | bilateral | | English | White- not Hispanic |
| 33 | OAB000 | OAB28AN.NSP | 43 | M | periarytenoid area | | English | White- not Hispanic |
| 34 | PAT000 | PAT10AN.NSP | 33 | M | unilateral left | Y | English | White- not Hispanic |
| 35 | PMF000 | PMF03AN.NSP | 34 | F | bilateral | | English | White- not Hispanic |
| 36 | RCC000 | RCC11AN.NSP | 49 | F | bilateral | N | English | White- not Hispanic |
| 37 | RJL000 | RJL28AN.NSP | 47 | M | | Y | English | White- not Hispanic |
| 38 | RTL000 | RTL17AN.NSP | 39 | M | bilateral | N | English | White- not Hispanic |
| 39 | RXP000 | RXP02AN.NSP | 26 | M | bilateral | N | Frech/Creole | Black- not Hispanic |
| 40 | SLC000 | SLC23AN.NSP | 28 | F | bilateral | | English | White- not Hispanic |
| 41 | SXG000 | SXG23AN.NSP | 70 | F | | N | English | White- not Hispanic |
| 42 | TLP000 | TLP13AN.NSP | 24 | F | bilateral | N | English | White- not Hispanic |
| 43 | VAW000 | VAW07AN.NSP | 39 | F | | N | English | White- not Hispanic |
| 44 | WST000 | WST20AN.NSP | 56 | M | | N | English | White- not Hispanic |

## INFORMAÇÕES DE PACIENTE COM "OUTRAS PATOLOGIAS"

| Nº | PAT_ID | FILE VOWEL'AH' | AGE | SEX | DISEASE | LOCATION | SMOKE | NATLANG | ORIGIN |
|---|---|---|---|---|---|---|---|---|---|
| 1 | AMC000 | AMC14AN.NSP | 48 | M | cyst | unilateral right | Y | Portuguese | White- not Hispanic |
| 2 | DVD000 | DVD19AN.NSP | 52 | M | cyst | unilateral right | | Vietnamese | Asian or Pacific Islander |
| 3 | EAB000 | EAB27AN.NSP | 40 | F | anterior saccular cyst | | Y | English | White- not Hispanic |
| 4 | EAS000 | EAS11AN.NSP | 47 | M | cyst | unilateral right | Y | English | White- not Hispanic |
| 5 | JCC000 | JCC10AN.NSP | 48 | F | cyst | unilateral right | | English | White- not Hispanic |
| 6 | LXC001 | LXC01AN.NSP | 33 | F | cyst | unilateral left | N | Cambodian | Asian or Pacific Islander |
| 7 | PMD000 | PMD25AN.NSP | 45 | F | cyst | unilateral right | | English | White- not Hispanic |
| 8 | SWS000 | SWS04AN.NSP | 26 | F | cyst | unilateral left | Y | English | White- not Hispanic |
| 9 | BSA000 | BSA08AN.NSP | 69 | M | paralysis | unilateral left | Y | Arabic | White- not Hispanic |
| 10 | CAR000 | CAR10AN.NSP | 66 | F | paralysis | unilateral left | Y | English | White- not Hispanic |
| 11 | CTY000 | CTY09AN.NSP | 75 | M | paralysis | unilateral left | Y | English | White- not Hispanic |
| 12 | DJP000 | DJP04AN.NSP | 43 | M | paralysis | unilateral right | N | English | White- not Hispanic |
| 13 | EDG000 | EDG19AN.NSP | 80 | F | paralysis | unilateral right | | English | White- not Hispanic |
| 14 | EJH000 | EJH24AN.NSP | 49 | M | paralysis | unilateral left | N | English | White- not Hispanic |
| 15 | DAC000 | DAC26AN.NSP | 64 | F | paralysis | unilateral left | Y | English | White- not Hispanic |
| 16 | DAG000 | DAG01AN.NSP | 75 | M | paralysis | unilateral left | Y | English | White- not Hispanic |
| 17 | JCC000 | JCC10AN.NSP | 48 | F | vocal nodules | bilateral | | English | White- not Hispanic |
| 18 | KAS000 | KAS09AN.NSP | 18 | F | vocal nodules | bilateral | N | English | White- not Hispanic |
| 19 | KCG000 | KCG25AN.NSP | 39 | F | vocal nodules | unilateral left | | English | White- not Hispanic |
| 20 | MRC000 | MRC20AN.NSP | 40 | F | vocal nodules | bilateral | N | Portuguese | White- not Hispanic |
| 21 | MXN000 | MXN24AN.NSP | 21 | F | vocal nodules | | N | Japanese | Asian or Pacific Islander |
| 22 | NJS000 | NJS06AN.NSP | 21 | F | vocal nodules | bilateral | N | English | White- not Hispanic |
| 23 | SEC000 | SEC02AN.NSP | 21 | F | vocal nodules | bilateral | N | English | White- not Hispanic |