Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Estudo de Técnicas para Classificação de Vozes Afetadas por Patologias

João Vilian de Moraes Lima Marinus

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Modelos Computacionais e Cognitivos

Herman Martins Gomes (Orientador)
Joseana Macêdo Fechine (Orientadora)

Campina Grande, Paraíba, Brasil

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

M339e Marinus, João Vilian de Moraes Lima.
Estudo de técnicas para classificação de vozes afetadas por patologias / João Vilian de Moraes Lima Marinus. — Campina Grande, 2010.
114 f. : il.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.
Orientadores: Profº. Ph.D Herman Martins Gomes, Profª. D.Sc Joseana Macêdo Fechine.
Referências.

1. Processamento Digital de Sinais. 2. Classificação de Vozes. 3. Redes Neurais MLP. 4. Quantificação Vetorial. 5. Modelo de Misturas de Gaussianas. I. Título.

CDU 004.383.3 (043)

"ESTUDO DE TÉCNICAS PARA CLASSIFICAÇÃO DE VOZES AFETADAS POR PATOLOGIAS"

JOÃO VILIAN DE MORAES LIMA MARINUS

DISSERTAÇÃO APROVADA EM 29.11.2010

HERMAN MARTINS GOMES, Ph.D
Orientador(a)

JOSEANA MACÊDO FECHINE, D.Sc
Orientador(a)

JOSÉ EUSTAQUIO RANGEL DE QUEIROZ, D.Sc
Examinador(a)

BENEDITO GUIMARAES AGUIAR NETO, Dr.
Examinador(a)

CAMPINA GRANDE - PB

Dedico esse trabalho aos meus pais, Dr. Nustenil e Maria, e aos meus irmãos Marinus e Nustenil Segundo.

"A alegria está na luta, na tentativa, no sofrimento envolvido.
Não na vitória propriamente dita."
M. Gandhi

Resumo

Nos últimos anos, várias pesquisas na área de processamento digital de voz estão sendo feitas, no sentido de criar técnicas que auxiliem o diagnóstico preciso por um especialista de patologias do trato vocal de maneira não invasiva, fazendo com que o paciente se sinta confortável na hora do exame. Este trabalho trata da investigação de técnicas para a classificação de vozes afetadas por patologias da laringe, em especial edema de Reinke, visando a construção de um sistema de apoio ao especialista. O sistema de auxílio ao diagnóstico de patologias da laringe, proposto nesta dissertação, é constituido de 3 etapas principais: pré-processamento do sinal de voz, extração de características e classificação. A etapa de pré-processamento consiste na aquisição do sinal de voz, na aplicação de um filtro de pré-ênfase para a minimização dos efeitos da radiação dos lábios e da variação da área da glote, seguido da segmentação e janelamento do sinal. Também foi investigada a não utilização da pré-ênfase nessa etapa. Na fase de extração de características, são utilizados coeficientes obtidos a partir da análise por predição linear (coeficientes LPC), coeficientes cepstrais, coeficientes delta-cepstrais e um vetor de características combinando coeficientes LPC e coeficientes cepstrais. A etapa de classificação é dividida em duas partes: classificação entre voz normal e voz afetada por patologia, sem especificar qual patologia, e caso o sinal seja classificado como voz afetada por patologia, tem-se uma segunda parte, a qual é realizada a classificação entre voz afetada por edema de Reinke e voz afetada por outra patologia. Para as duas partes, foram testados 3 diferentes classificadores: Redes Neurais Multilayer Perceptron - MLP, Modelos de Misturas de Gaussianas e Quantização Vetorial. Para diferenciar entre voz normal e voz afetada por patologia, os melhores resultados foram obtidos utilizando Redes Neurais. Para diferenciar entre voz afetada por edema e voz afetada por outra patologia, os melhores resultados foram obtidos utilizando Quantização Vetorial. Em ambos os casos, os melhores resultados foram obtidos ao se utilizar coeficientes cepstrais e sem utilização da pré-ênfase.

Palavras-chave: Processamento digital de sinais de voz, Classificação de vozes afetadas por patologias, Redes Neurais MLP, Quantização Vetorial, Modelo de Misturas de Gaussianas.

Abstract

In recent years, several studies in digital voice processing are being made in order to create techniques to support a noninvasive accurate diagnosis of vocal tract diseases by a specialist, making the patient feel comfortable during examination. This work deals with the investigation of techniques for classification of voices affected by laryngeal pathologies, especially Reinke's edema, aiming to build a support system to the specialist. The system for the diagnosis of laryngeal pathologies, proposed here, consists of three main steps: preprocessing the speech signal, feature extraction and classification. Preprocessing corresponds the acquisition of voice signal, the application of a pre-emphasis filter for minimizing the radiation effects from the lips and from variation in glottal area, and the signal segmentation and windowing. The non-use of pre-emphasis was also investigated at this point. In the feature extraction step, we use coefficients obtained from the linear prediction analysis (LPC coefficients), cepstral coefficients, delta-cepstral coefficients, and a feature vector combining LPC and cepstral coefficients. The classification is divided into two parts: classification of normal voice versus voice affected by pathology, without specifying which pathology, and if the signal is classified as voice affected by pathology, second part happens, which is performed by the classification between voice affected by Reinke's edema and voice affected by other pathology. For both parties, 3 different classifiers were tested: Neural Networks Multilayer Perceptron - MLP, Gaussian Mixture Models and Vector Quantization. To differentiate between normal voice and voice affected by pathology, the best results were obtained using Neural Networks. To differentiate between voice affected by edema and voice affected by pathology, the best results were obtained using vector quantization. In both cases, the best results were obtained when using cepstral coefficients and without use of pre-emphasis.

Keywords: Digital voice signal processing, voices affected by pathologies classification , Neural Networks MLP, Vector Quantization, Gaussian Mixtures Model.

Agradecimentos

Em primeiro lugar, gostaria de agradecer a Deus, pela minha existência e por ter me ajudado, dando-me paciência e perseverança para superar os obstáculos da vida.

Aos meus pais, Dr. Nustenil e Maria de Moraes, pela minha vida, por terem lutado com perseverança para educar seus filhos, enfrentando e superando as diversas dificuldades que a vida nos proporciona, por todo amor, carinho e confiança e por toda dedicação dada à criação de seus filhos, sempre os colocando em primeiro lugar.

Aos meus irmãos, Marinus e Nustenil Segundo, meus melhores amigos, companheiros de jornada, por sempre estarem presentes em todos os momentos da minha vida, nos melhores para comemorar comigo e nos piores para me ajudar.

Aos professores Herman Martins Gomes e Joseana Macêdo Fechine, pela orientação deste trabalho, pela dedicação e estímulos constantes, que foram fundamentais em todos os momentos do mestrado.

À professora Silvana Luciene do Nascimento Cunha Costa, pelo apoio, incentivo e pela valorosa contribuição.

A todos da Copin, em especial a Aninha, por sempre ajudar os alunos com um sorriso no rosto.

A todos os amigos que fiz durante a graduação e o mestrado aqui em Campina Grande, por todo apoio dado, e pelos momentos de descontração.

A todos os companheiros de laboratório, do *iPhotoBot* e do LVC, por sempre ajudarem em momentos de dúvida.

Conteúdo

Lista de Siglas e Abreviaturas

$\Delta\Delta MCep$ - *Coeficientes Delta-Delta-Mel-cepstrais*

ΔCep - *Coeficientes Delta-Cepstrais*

$\Delta CepP$ - *coeficientes Delta-cepstrais ponderados*

$\Delta MCep$ - *Coeficientes Delta-Mel-cepstrais*

BDA - *Base de dados Artificial*

Cep - *Coeficientes Cepstrais*

CepP - *coeficientes Cepstrais ponderados*

ceW - *coeficientes de energia Wavelet*

cenW - *coeficientes de entropia Wavelet*

DFA - *Detrended Fluctuation Analysis*

DLG - *discriminador linear Gaussiano*

DME - *Decomposição de modo empírico*

E - *Energia*

EM - *Expectation-Maximization*

EM - *espectro de modulação*

eR - *entropia Relativa*

eS - *entropia de Shannon*

eT - *entropia de Tsallis*

FFT - *Fast Fourier transform*

FIR - *Finite Impulse Response*

GERABTA - *Laboratório G.E. da Universidade de Los Angeles e RABTA Hospital de Tunis*

GMM - *Gaussian Misture Models*

GPV - *Grau de Paradas de Voz*

HMM - *Hidden Markov Models*

KVP - *K-vizinhos mais próximos*

LBG - *Linde, Buzo e Gray*

LPC - *Linear Predictive Coding*

LVQ - *Learning Vector Quantization*

MCep - *Coeficientes Mel-cepstrais*

MEEI - *Massachussetts Eye and Ear Infirmary*

MLP - *Multilayer Perceptron*

NLS - *Nervo Laríngeo Superior*

NLR - *Nervo Laríngeo Recorrente*

PMR - *Perturbação Média Relativa*

PW - *Pacotes Wavelets*

QPA3pt - *quociente de perturbação da amplitude de 3 pontos*

QPA11pt - *quociente de perturbação da amplitude de 11 pontos*

QPP5pt - *quociente de perturbação do periodo de 5 pontos*

QV - *Quantização Vetorial*

RCSVHP - *Republican Center of Speech, Voice and Hearing Pathologies*

RNA - *Rede Neural Artificial*

RPDE - *Return Period Density Entropy*

SPAPL - *Speech Processing and Auditory Perception Laboratory*

SVM - *Support Vector Machine*

TIMIT - *TIMIT continuous speech corpus*

TMRAP - *Tri Mean Relative average pertubation*

TR - *Taxa de ruído*

TWC - *Transformada Wavelet Contínua*

TWD - *Transformada Wavelet Discreta*

VQ - *Vector Quantization*

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

As patologias nas dobras vocais normalmente se caracterizam por apresentar como sintoma principal ou secundário a disfonia (FUKUDA, 2003). Disfonia é qualquer dificuldade da emissão vocal que modifique a produção normal da voz, afetando a qualidade da produção vocal, causada por alterações orgânicas ou funcionais da laringe (ALBERNAZ; GANANA; FUKUDA, 1997). As principais causas da disfonia são inadaptações fônicas, alterações psicoemocionais e o uso incorreto da voz (FUKUDA, 2003), ou doenças neuro-degenerativas (DAVIS, 1979; QUEK et al., 2002).

Dentre as patologias que afetam a laringe, podem-se citar nódulos nas dobras vocais, pólipos, cistos, carcinomas e paralisia dos nervos laríngeos. Estas patologias podem ser tratadas por meio de terapia vocal, cirurgia ou, em último caso, radioterapia (MARTINEZ; RUFINER, 2000).

A ocorrência destas patologias tem crescido devido, principalmente, a hábitos sociais não saudáveis[1] e ao abuso vocal[2]. Particularmente, em função do abuso vocal, o grupo profissional em que se registra mais ocorrências são os professores (COSTA et al., 2000; GOTAAS; STARR, 1993; KOUFMAN; ISAACSON, 1991; SAPIR; KEIDAR; MATHERS-SCHMIDT, 1993).

Existem várias técnicas para a avaliação da qualidade vocal de um paciente. Uma técnica bastante utilizada consiste na escuta da voz do paciente e definí-la como afetada por patologia ou não. O problema desta técnica é que ela possui caráter subjetivo, varia conforme

[1]Ex: fumar; beber bebidas alcoolicas; usar drogas; tomar café em excesso; dormir pouco; entre outros.

[2]Ex: Uso excessivo da voz; falar muito alto; uso da voz por muito tempo sem hidratação; falar excessivamente em ambientes frios, como salas com ar condicionado; falar em ambiente ruidoso; entre outros.

o profissional que a está aplicando e é dependente da sua experiência. Portanto, pode estar sujeita a erros. Técnicas laboratoriais são aplicadas no sentido de contornar esse problema, tais como videolaringoscopia (exame com um instrumento de fibra ótica), videoestroboscopia (iluminação estroboscópica da laringe, útil para visualização dos movimentos), eletromiografia (observação indireta do estado funcional da laringe) e videofluoroscopia (técnica radiográfica a partir da qual o paciente ingere uma determinada quantidade de substância rádio-opaca para avaliar a deglutição) (MARTINEZ; RUFINER, 2000).

A partir destas técnicas, pode-se diagnosticar com precisão as mais diversas patologias das dobras vocais. O problema é que elas são invasivas, causando desconforto ao paciente, o que gera resistência por parte deste, no momento em que se efetua o exame, podendo causar distorções nos dados obtidos e, assim, produzir falsos diagnósticos (ADNENE; LAMIA, 2003; ALONSO et al., 2001), além de necessitar de instrumentos caros e sofisticados, e de serem consideradas de alto risco, tendo que ser executadas em condições controladas por profissionais especializados (MANFREDI, 2000; ESPINOSA; FERNáNDEZ-REDONDO; GOMEZ, 2000).

1.1 Motivação

Nos últimos anos, várias pesquisas na área de processamento digital de sinais de voz vêm sendo levadas a efeito, no sentido de criar técnicas que auxiliem o diagnóstico preciso por um especialista em patologias do trato vocal de maneira não invasiva, fazendo com que o paciente se sinta mais confortável na hora do exame .

Essas técnicas podem ser aplicadas no auxílio às técnicas laboratoriais citadas anteriormente permitindo, assim, diminuir a regularidade dos exames invasivos, além de poderem ser usadas como forma de prevenção, verificando a qualidade vocal dos profissionais que trabalham com a voz, e pré-detectando possíveis patologias.

Diferentes abordagens para extração de características da voz foram propostas. Inicialmente, foram usados características como frequência fundamental (correlato perceptual - pitch), jitter (perturbação da frequência fundamental), shimmer (perturbação em amplitude), quociente de perturbação de amplitude, quociente de perturbação do pitch, relação sinal-ruído, energia de ruído normalizada (MANFREDI, 2000), dentre outros (ROSA; PEREIRA;

GRELLET, 2000; WALLEN; HANSEN, 1996). Atualmente, têm sido registradas na literatura da área pesquisas baseadas no modelo de predição linear (LPC), análise cepstral e no modelo de percepção auditiva (COSTA, 2008; AGUIAR-NETO; COSTA; FECHINE, 2008; GODINO-LLORENTE; GÓMEZ-VILDA, 2004; GODINO-LLORENTE; GÓMEZ-VILDA; BLANCO-VELASCO, 2006).

Dentre as técnicas mais utilizados na construção de classificadores para o diagnóstico de patologias das dobras vocais, estão a Quantização Vetorial (AGUIAR-NETO; COSTA; FECHINE, 2008), o Modelo de Misturas de Gaussianas (GODINO-LLORENTE; GÓMEZ-VILDA; BLANCO-VELASCO, 2006), as Máquinas de Suporte Vetorial (KUKHARCHIK et al., 2007), os Modelos de Markov Escondidos (COSTA, 2008) e Redes Neurais Artificiais, com destaque para as Redes Multilayer Perceptron (MARTINEZ; RUFINER, 2000; GODINO-LLORENTE; GÓMEZ-VILDA, 2004).

A maioria destas abordagens objetiva diferenciar voz normal de voz afetada por patologia, sem especificar a patologia. Ultimamente, foram realizados esforços no sentido de destacar uma patologia das demais, como, por exemplo, o edema de Reinke (COSTA, 2008). Ainda assim, as pesquisas ainda não são conclusivas, não tendo sido obtido um conjunto de características que melhor represente um sinal de voz com patologia e um classificador mais preciso para esse conjunto de características. Aliado a isto, destaca-se que ainda é uma tarefa árdua diferenciar uma patologia das demais, devido ao caráter ruidoso dos sinais de voz com patologias.

1.2 Objetivos

O objetivo principal deste trabalho é o estudo de técnicas para a classificação de um sinal de voz em uma das três classes: voz normal, voz afetada por edema de Reinke, e voz afetada por outra patologia. O foco do trabalho está no estudo de diferentes classificadores, de forma a propor um método de classificação satisfatório no sentido de discriminar edema de Reinke face a outras patologias.

Diversos trabalhos anteriores utilizaram características de longa duração para análise do sinal de voz, como pitch, jitter, entre outros. Entretanto, muitas dessas características são baseadas na extração da frequência fundamental, o que pode ser uma tarefa complexa devido

à característica ruidosa do sinal afetado pela patologia (BOYANOV et al., 1993; MANFREDI; PIERAZZI; BRUSCAGLIONI, 1999).

Costa (2008) realizou uma comparação entre diferentes vetores de características, de modo a verificar os que melhor representam um sinal de voz com patologia. O trabalho realizou um estudo comparativo entre características obtidas a partir da Codificação por Predição Linear, da Análise Cepstral e derivados e de coeficientes mel-cepstrais para classificar os sinais de voz nas três classes citadas anteriormente. Para tanto, utilizou-se um sistema de classificação envolvendo Quantização Vetorial e Modelos de Markov Escondidos para realizar a comparação. Nesta dissertação, optou-se por utilizar os coeficientes obtidos a partir da Codificação por Predição Linear e da Análise Cepstral, por terem proporcionado os melhores resultados no trabalho de Costa (2008).

1.3 Abordagem Proposta

O processo de classificação do sinal de voz é dividido em três etapas principais: pré-processamento do sinal de voz, extração de características e treinamento/classificação.

No pré-processamento do sinal de voz, é aplicado um filtro de pré-ênfase no sinal, de maneira a diminuir o seu ruído. Em seguida, o sinal é segmentado e é aplicada uma função janela a cada segmento. Também é investigada a não aplicação do filtro de pré-ênfase.

Na etapa de extração de características, são obtidos coeficientes a partir da Codificação por Predição Linear (coeficientes LPC - *Linear Predictive Coding*) de cada janela, e também os coeficientes a partir da Análise Cepstral (coeficientes Cepstrais e coeficientes Delta-cepstrais). Os coeficientes Cepstrais são obtidos a partir de uma abordagem paramétrica derivada dos coeficientes LPC. Também é utilizada uma combinação de coeficientes LPC e coeficientes Cepstrais como vetor de características. Os vetores de características obtidos nessa etapa são utilizados para treinamento dos classificadores, e também para classificação dos sinais.

Os experimentos de treinamento e classificação foram divididos em 2 etapas: classificação entre voz normal e voz afetada por patologia, sem especificar a patologia; e caso o sinal possua patologia, detectar a ocorrência de edema de Reinke, ou outra patologia.

Para ambas as etapas, foram testadas Redes Neurais Multilayer Perceptron (MLP), Quan-

tização Vetorial e Modelos de Misturas Gaussianas (*Gaussian Mixture Models* - GMM).

Para a classificação utilizando Redes Neurais utilizou-se inicialmente a Quantização Vetorial (algoritmo de Linde, Buzo e Gray - LBG (LINDE; BUZO; GRAY, 1980)) nos vetores de características obtidos na etapa anterior, para redução da dimensionalidade do vetor de características, e normalização do número de vetores obtidos, já que os sinais podem possuir tamanhos distintos. Em seguida, foi realizado o treinamento, utilizando uma Rede Neural MLP. Para classificação, o sinal de voz de teste foi conduzido por todas as etapas anteriores, inclusive a Quantização Vetorial, sendo o resultado aplicado a uma Rede Neural já treinada para classificar como (i)voz normal e voz afetada por patologia; ou (ii)voz afetada por edema e voz afetada por outra patologia.

Para a classificação utilizando Quantização Vetorial, foi criado um padrão de referência da classe edema utilizando o algoritmo LBG , então foram obtidas medidas de distorção entre os vetores de características e o padrão de referência, e comparado a um limiar. Caso essa medida fosse igual ou inferior ao limiar, o sinal era classificado como voz afetada por patologia e caso fosse superior, classificado como voz normal. Para classificação entre voz afetada por edema e voz afetada por outra patologia foi obtido outro limiar. Caso a medida de distorção fosse igual ou abaixo desse limiar, o sinal é classificado como afetado por edema, caso contrário, como afetado por outra patologia.

Para classificação utilizando GMM, foi criado um modelo de apenas uma classe (classe normal para voz normal *versus* voz afetada por patologia e classe edema para voz afetada por edema *versus* voz afetada por outra patologia), e foi calculada a probabilidade a priori de os vetores de características pertencerem ao modelo. Se a probabilidade fosse igual ou superior a um limiar estabelecido, o sinal era classificado como sendo da classe modelada; se estivesse abaixo, era classificado como sendo pertencente à outra classe.

A base de dados utilizada neste trabalho foi obtida do Massachussetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab (KAY-ELEMETRICS, 1994). Esta base de dados contém 1400 amostras de voz obtidas a partir de, aproximadamente, 700 pessoas, sendo que aproximadamente 700 das amostras são da vogal sustentada /a/, uma para cada pessoa. Os sinais foram obtidos com baixo nível de ruído, distância constante do microfone, tamanho da amostra de 16 bits e taxa de amostragem de 25 ou 50 Kamostras/s, com uma resolução de 16 bits/amostra.

Dentre estas amostras, foram utilizadas 53 sinais de voz normal, sendo 21 do gênero masculino e 32 do gênero feminino, 43 sinais de indivíduos com edema, sendo 32 mulheres e 11 homens, e 21 sinais de voz afetada por outra patologia (nódulos, cistos, e paralisia), com 14 vozes femininas e 7 masculinas.

Para a fase de treinamento do sistema, visando a classificação entre voz normal e voz com patologia, utilizou-se 27 sinais de voz normal e 33 dos sinais de voz afeta por patologia e os demais sinais foram usados na fase de teste nas Redes Neurais. Para a Quantização Vetorial, foram utilizados 25 sinais de voz afetada por edema para criação do padrão de referência e o restante foi utilizado para testes. Para o treinamento no GMM, foram utilizados 25 sinais de voz normal para geração do modelo, e os sinais restantes foram utilizados para teste. Para classificação entre voz afetada por edema e voz afetada por outra patologia, foram utilizados 11 sinais de voz afetada por edema e 11 sinais de voz afetada por outra patologia para treinamento e o restante para testes nas Redes Neurais. Para a Quantização Vetorial e para para o GMM, foram utilizados 25 sinais de voz afetada por edema para a criação do padrão de referência, sendo os sinais restantes utilizados para testes. Os sinais de voz normal foram desconsiderados para classificar entre voz afetada por edema e voz afetada por outra patologia.

A base de dados foi obtida com o software Multi-Speech - Signal Analysis Workstation, Modelo 3700, da Kay Elemetrics, USA (KAY-ELEMETRICS, 1994). A obtenção dos vetores de características foi efetuada por meio de implementação em MATLAB 7.0, MathWorks (MATLAB, 1998), as Redes Neurais foram simuladas no WEKA (HALL et al., 2009), a Quantização Vetorial foi implementada em MATLAB 7.0 e para o GMM foi utilizada a *toolbox* GMM/GMR 2.0 do MATLAB (CALINON, 2009).

1.4 Estrutura do Trabalho

O restante da dissertação é organizado como se segue. No Capítulo 2, é apresentada a fundamentação teórica do trabalho, com apresentação do Estado da Arte da área, incluindo a descrição da fisiologia de produção da voz e de diversas patologias que afetam a laringe. Também é apresentada a descrição dos algoritmos empregados no sistema proposto de classificação de voz, os algoritmos de pré-processamento, de extração de características (co-

eficientes LPC, Cepstrais e Delta-cepstrais) e dos classificadores utilizados (Quantização Vetorial, Redes Neurais MLP e GMM).

No Capítulo 3, é apresentada uma descrição mais detalhada da abordagem proposta, incluindo informações sobre a base de dados utilizada nos experimentos.

No Capítulo 4, são mostrados e analisados os resultados obtidos no processo de classificação e no Capítulo 5 são apresentadas as considerações finais e sugestões para trabalhos futuros.

Em anexo, apresentam-se tabelas com a identificação dos sinais de voz utilizados na base de dados, além de algumas características dos locutores, como gênero e faixa-etária.

Capítulo 2

Fundamentação

2.1 Introdução

O objetivo deste capítulo é proporcionar um embasamento teórico sobre o problema abordado no trabalho, a discriminação automática de patologias da fala.

Neste capítulo, é mostrado um breve estudo sobre o processo de produção da voz, com a descrição do funcionamento dos principais órgãos, e algumas patologias que afetam a laringe, atingindo as dobras vocais. As patologias abordadas neste capítulo são edema de Reinke, cistos vocais, nódulos vocais e paralisia. Na seção seguinte, um estudo sobre o processamento digital do sinal de voz, desde a aquisição do sinal, passando por pré-processamento, extração de características, e concluindo no treinamento e uso de classificadores.

As características abordadas neste capítulo são os coeficientes LPC, coeficientes Cepstrais e coeficientes Delta-cepstrais. Os classificadores vistos neste capítulo são Redes Neurais Multilayer Perceptron, Quantização Vetorial e Modelos de Misturas de Gaussianas.

2.2 Fisiologia e Patologias da Voz

O sistema produtor de voz humana consiste na junção de partes de dois sistemas do corpo humano: sistema digestivo e sistema respiratório, tendo esse último maior importância na produção da voz humana. Do sistema respiratório, o pulmão, a traquéia, a laringe e as cavidades nasais participam da produção da voz, enquanto no sistema digestivo, apenas a

faringe e a cavidade oral participam. Nas próximas subseções, será descrita detalhadamente cada parte do sistema de produção da fala, e as patologias que afetam a laringe.

2.2.1 Sistema de Produção da Fala

O estudo do sistema de produção da fala é fundamental para as modelagens matemáticas que servem de base para o desenvolvimento de sistemas de reconhecimento, síntese e codificação de voz, bem como de sistemas de terapia de voz.

O sistema de produção de fala (Figura 2.1) é formado pelo pulmão, traquéia, laringe, faringe, e pelas cavidades nasais e oral. A produção de voz começa com a respiração, a partir da qual é inalado ar nos pulmões. Após, o ar é expelido dos pulmões, passando pela traquéia, que consiste num tubo que liga a laringe aos pulmões, e chega na laringe. A laringe é o órgão fundamental na produção da voz. Nela estão localizadas as dobras vocais, que podem vibrar com a passagem do ar vindo dos pulmões, excitando assim o trato vocal, que é constituído pela faringe, cavidade oral, e cavidades nasais. Essas duas últimas só são excitadas em casos de sons nasais (RIBEIRO, 2003).

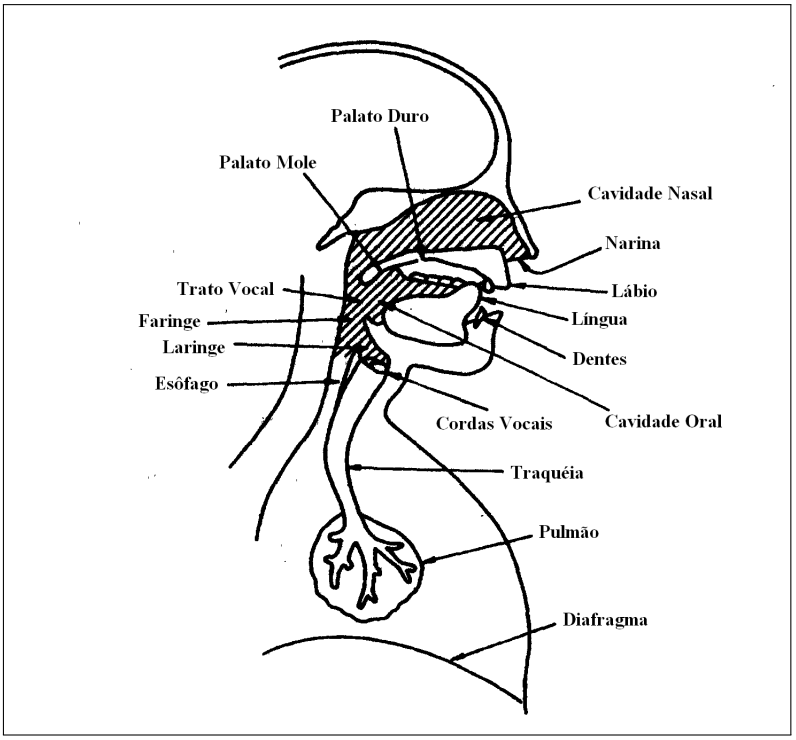

Figura 2.1: Sistema de produção de fala.

Adaptado de (FURUI, 2001).

A laringe (Figura 2.2) é um órgão em formato de tubo alongado situado na parte anterior do pescoço, que realiza a ligação entre a traquéia e a faringe (DAJER, 2006). A principal função da laringe é conduzir o ar durante a respiração. Outras funções são proteger as vias aéreas inferiores, deglutir, eliminar secreções e corpos estranhos e apoiar os mecanismos de esforços, como por exemplo defecação e parto (IMAMURA; TSUJU; SENNES, 2002). Uma outra função da laringe de extrema importância aos seres humanos é a fonação, ou seja, a produção de voz. A voz humana é resultado de uma interação de fatores genéticos, anatômicos, características pessoais, como por exemplo, a personalidade, além de fatores econômicos, socio-culturais, e ainda de aspectos emocionais únicos de cada indivíduo (BEHLAU et al., 2001).

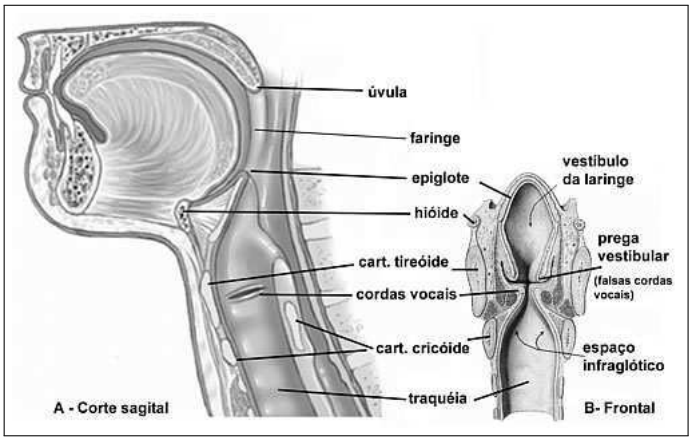

Figura 2.2: Laringe.

Fonte: http://www.viaaereadificil.com.br/anatomia/anatomia.htm

As dobras vocais (Figura 2.3) são estruturas multilaminadas compostas por músculo e mucosa. A mucosa pode ser dividida em epitélio e lâmina própria (subdividida em camadas superficial, intermediária e profunda) (GRAY, 1991). O mecanismo vibratório das dobras vocais origina-se de ondulações nos sentidos horizontal, vertical e longitudinal, determinando o ciclo glótico com suas fases de abdução e adução. O comprimento das dobras vocais varia de tamanho no controle da frequência da voz, e a pressão subglótica varia durante o controle da intensidade da voz. O Trato vocal é modificado totalmente durante o controle da qualidade vocal (TUMA et al., 2005).

A qualidade vocal depende, dentre outros, do modo de fechamento e abertura da glote, e da vibração das dobras vocais. Os principais fatores que determinam a vibração das dobras

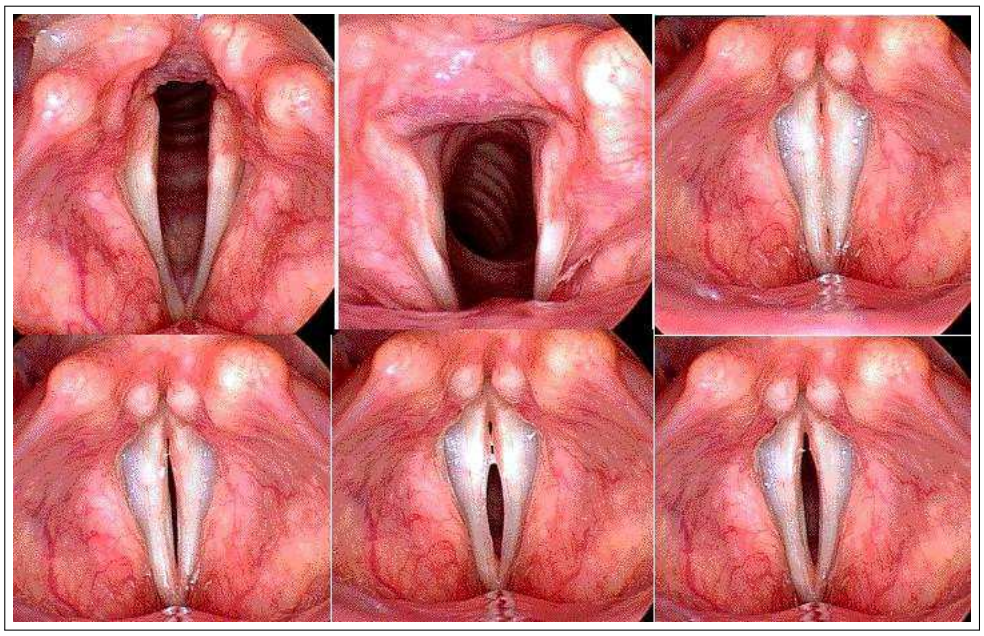

Figura 2.3: Movimento das dobras vocais.

Fonte: http://www.rc.unesp.br/pef/2003_projetos/Pedro/trabalho.htm

vocais são (HANSEN; GAVDIDIA-CEBALLOS; KAISER, 1998):

- Posição da dobra vocal, ou a extensão em que as dobras vocais são aduzidas ou abduzidas;
- Mioelasticidade[1], que é determinada pela posição e grau de tensão do músculo vocal durante a contração;
- Nível de pressão do ar através das dobras vocais.

Em determinadas condições que podem ser causadas por diferentes patologias vocais, as vibrações das dobras vocais podem se tornar aperiódicas ou irregulares. São elas: assimetria, interferência na homogeneidade, flacidez, tono oscilante e força inconsistente (COSTA, 2008).

2.2.2 Patologias da Laringe

As patologias da laringe normalmente produzem mudanças assimétricas nas dobras vocais, causando diferentes tipos de vibrações, e mudanças nas características de massa, elasticidade e tensão das dobras (DAVIS, 1979), sendo causadas por alterações morfológicas nas estruturas

[1] grau de elasticidade das dobras vocais

do trato vocal, ou por seu mau funcionamento (COSTA, 2008). Quando ocorre alguma mudança negativa da voz, como rouquidão, rudeza, aspereza, estridência, entre outras, da-se a essa mudança o nome de disfonia (DANIEL; BOONE; MCFARLANE, 1994). As disfonias podem ser classificadas como sendo de três tipos (BRANCO; ROMARIZ, 2006; BEHLAU, 2001):

- Disfonias Funcionais - são causadas por uso abusivo de voz e/ou inadaptações vocais e alterações psicogênicas. Elas são caracterizadas por alterações na emissão da voz;
- Disfonias Organofuncionais - são disfonias funcionais que apresentam lesões orgânicas secundárias. Normalmente, estão relacionadas a diagnósticos tardios de disfonias funcionais. Entre as disfonias organofuncionais, tem-se o edema de Reinke, patologia de estudo desse trabalho, a qual será apresentada posteriormente;
- Disfonias Orgânicas - são disfonias que apresentam alterações anatômicas, e que podem possuir causas independentes do uso da voz.

A presença da patologia na laringe pode ser detectada a partir de sintomas relatados aos médicos por seus pacientes, 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, tais como a rouquidão, garganta rascante ou tremor na voz. A seguir, estão descritos oito dos principais sintomas associados a patologias da laringe. Esses sintomas normalmente não ocorrem individualmente, mas sim de forma combinada (COLTON; CASPER, 1996).

- Rouquidão - ocorrência de uma vibração aperiódica das dobras vocais. Outros termos também utilizados: voz “rouca”, “áspera” ou “raspada”;
- Fadiga vocal - cansaço após fala prolongada, e exigência de muito esforço para fala contínua;
- Soprosidade - incapacidade de pronunciar sentenças completas sem pausa para inspiração de ar;
- Extensão fonatória reduzida - ocorrência principalmente em cantores, que se queixam da dificuldade em produzir notas que anteriormente eram produzidas facilmente. Em

geral, estas notas ocorrem na extremidade superior da extensão de canto (tons mais agudos);

- Afonia - ausência de voz. O paciente fala sussurrando e pode, às vezes, ter uma variedade de sintomas associados, como secura na garganta, dor e uma grande dificuldade para tentar falar;

- Quebras de frequência - saltos periódicos de voz e quebra de voz. A voz parece fora de controle e o paciente não sabe que som sairá. Ocorre pelo uso inadequado de falsete ou puberfonia[2]. É relatado por jovens adolescentes que utilizam uma frequência inapropriadamente aguda como voz habitual ao invés da voz masculina típica de frequência mais grave;

- Voz tensa/comprimida - dificuldades ao falar. Pode incluir também inabilidade de iniciar ou manter a vocalização. Exigência de esforço para falar por causa da tensão, e ocorrência de fadiga devido ao esforço envolvido;

- Tremor - voz cambaleante ou trêmula. Incapacidade de produzir voluntariamente um som estável sustentado.

A seguir, estão descritas com mais detalhes as patologias edema de Reinke, cistos vocais, nódulos vocais e paralisia. Essas são as patologias presentes na base de dados utilizada no trabalho.

Edema de Reinke

Edema é um acúmulo de fluido em alguma parte da dobra vocal, podendo ocorrer em camadas superficiais ou em locais mais profundos (PARRAGA, 2002). O edema de Reinke (Figura 2.4) é um edema que ocorre na camada superficial da lâmina própria. Ele pode ocorrer em ambas as dobras vocais ou somente de um lado, principalmente em estágios iniciais. O acúmulo de fluido ocorre na parte mucosa do espaço de Reinke, fazendo com que a cobertura da dobra vocal fique menos rígida e vibre como uma estrutura mais massiva (HIRANO, 1981).

[2]Uso de voz aguda acima da idade em que a voz masculina deveria ter mudado

Figura 2.4: Edema de Reinke unilateral.

Fonte: http://www.voicemedicine.com/reinkes_edema.htm

O edema de Reinke também é conhecido como cordite polipóide, degeneração polipóide ou polipose difusa bilateral. O primeiro anatomista a descrever a estrutura fina das dobras vocais foi Reinke, que teve seu nome dado à camada superficial da lâmina própria (BENJAMIN, 2000), local de ocorrência do edema de Reinke.

A principal causa do edema de Reinke é o fumo, seguido por uso abusivo da voz (KLEINSASSER, 1997). A alergia também pode ser considerada como um dos fatores etiopatogênicos[3]. A hipersensibilidade a diferentes alergenos inalantes podem tornar a parede mucosa da laringe mais susceptível à ação de diversos outros fatores, como mau uso vocal, refluxo gastro-esofágico, fumo, fatores irritantes climáticos, entre outros (HOCEVAR-BOLTEZAR; RADSEL; ZARGI, 1997). O consumo do álcool, obstrução nasal e infecção das vias aéreas superiores também podem ser fatores constitucionais (PAPARELLA; SHUMRICK, 1982). Todos estes fatores agindo juntos causam lesões na mucosa laringea (ABREU, 1999).

Inicialmente, o paciente afetado pelo edema só percebe uma diminuição do *pitch* da voz. Gradualmente, o edema de Reinke vai se desenvolvendo, a voz vai se tornando mais áspera e é necessário um aumento do esforço para falar. Dessa maneira, as pessoas que sofrem de edema normalmente só procuram um especialista quando ocorrem grandes mudanças na voz (SCALASSARA, 2009).

Cistos vocais

O cisto vocal, ilustrado na Figura 2.5, é uma patologia benigna que produz uma alteração funcional da voz, caracterizada por disfonia e fadiga vocal. O grupo de maior incidência é o de gênero feminino, com idade entre 20 e 50 anos. Existem duas possíveis causas para o

[3]Causas de uma doença

cisto: adquirido ou mal formação congênita (BOUCHAYER et al., 1985).

Figura 2.5: Cisto vocal.

Fonte: http://www.voicemedicine.com/cyst.htm

Os cistos podem ser classificados em três grandes grupos:

- Cisto Intracordal de Retenção - também conhecido como Cisto Intracordal Mucoso, Cisto Mucoso, Cisto de Retenção Mucosa, Cisto de Inclusão epidermóidica ou Cisto do tipo Anexial. São pequenas lesões benignas da dobra vocal, que causam disfonia e esforço na produção da voz. Supõem-se que o cisto seja resultado de uma obstrução do duto da glândula mucosa, provavelmente sendo resultado de algum processo inflamatório, possuindo no seu interior fluido viscoso (MONDAY et al., 1983; BOUCHAYER et al., 1985).

- Cisto Epidermóide - também conhecido como Cisto Intracordal Epidermóide verdadeiro, Cisto Aberto, Cisto Fechado ou Cisto tipo Epidérmico. Pode ter causa congênita ou adquirida. O cisto congênito é derivado de uma má formação que ocorre durante a vida intra-uterina, no decorrer da formação do quarto e do sexto arcos branquiais[4] ou durante a formação da laringe (BOUCHAYER et al., 1985), podendo também ser proveniente de uma mudança de formação do epitélio de revestimento (STEFFEN; MOSCHETTI; ZAFFARI, 1995). O cisto adquirido é causado por um processo em que as dobras vocais estão hiperreativas e qualquer agressão de ordem física ou mecânica ocasiona edema, levando à inflamação. Esta reação leva a uma invaginação do epitélio de revestimento da superfície da lâmina própria, formando-se uma fenda que depois se fecha e vira um cisto (MONDAY et al., 1983).

[4]Estruturas do embrião que dão origem a Laringe

- Pseudocisto - também conhecido como Cisto de Inclusão Epidermóidica ou Cisto Intra-epitelial. O pseudocisto se assemelha ao cisto de retenção pela localização, e pode ser confundido com nódulos. Sua parede é fina e translúcida, contendo no seu interior um líquido viscoso (MONDAY et al., 1981). O pseudocisto não pode ser visto como um cisto verdadeiro por não ter a sua parede completa (MONDAY et al., 1983).

Dentre as patologias existentes em relação à voz, esta se diferencia pelo fato de precisar de um minucioso procedimento de diagnóstico clínico, com vistas de evitar, dentro do possível, alguma confusão com outra patologia (NEGREIROS, 1997).

Nódulos vocais

Os nódulos vocais (Figura 2.6) são protuberâncias esbranquiçadas ou cinzentas (DANIEL; BOONE; MCFARLANE, 1994). São as lesões benignas mais superficiais da lâmina própria, frequentemente acompanhadas de edema (BEHLAU; PONTES, 1995). Ocorrem geralmente em ambas as dobras vocais de maneira simétrica, mas também podem ser encontrados unilateralmentes (CASE, 1996; GONZáLES, 1990; WILSON, 1993).

Figura 2.6: Nódulos vocais.

Fonte: http://www.voicemedicine.com/nodules.htm

Os nódulos vocais são a patologia relacionada a voz que mais afeta as crianças em idade escolar, sendo a causa de distúrbios de voz em 80% das crianças. Os nódulos são mais comuns em crianças agitadas, em especial do gênero masculino, que cometem abusos vocais (COSTA, 2008) como gritos, fala excessiva, falar competindo com o ruído do ambiente (GREEN, 1989; HERSAN, 1991; CASE, 1996; WILSON, 1993), ataque vocal brusco, uso impróprio do *pitch* (GREEN, 1989; CASE, 1996; WILSON, 1993) e do *loudness*, fonação invertida, vocalizações explosivas (WILSON, 1993), berros, vocalizações tensas (CASE, 1996;

WILSON, 1993), choro prolongado (GREEN, 1989; HERSAN, 1991), pigarro (HERSAN, 1991; CASE, 1996; WILSON, 1993), tosse, rir excessivamente, imitar outras vozes (HERSAN, 1991; CASE, 1996), cantar de modo abusivo, falta de hidratação e falar com apoio respiratório inadequado (CASE, 1996).

Entre os adultos, a incidência de nódulos vocais é maior entre as mulheres, normalmente em ocupações que requerem uso frequente de voz, como cantores, locutores, operadores de telefonia ou telemarketing, professores e outros profissionais que usam a voz exageradamente. Pessoas afetadas por nódulos vocais apresentam qualidade da voz reduzida e soprosidade com vários graus de ruído. Geralmente, a voz apresenta irregularidades como rouquidão e instabilidade (HAMMARBERG, 1998).

Paralisia

Paralisia (Figura 2.7) é a perda da capacidade de movimentos voluntários de um músculo por causa de lesão ou doença que afetem a unidade motora e o sistema nervoso central. No caso da paralisia das dobras vocais, inicialmente os músculos tornam-se flácidos e, mais tarde, apresentam vários graus de atrofia e fibrose[5] (GREENE, 1989).

Figura 2.7: Paralisia após uma operação na tireóide.
Fonte: http://www.voicemedicine.com/unilateral.htm

Dentre os nervos que controlam todos os músculos da laringe, o principal é o nervo vago. O nervo vago possui duas ramificações, os nervos laríngeos superior (NLS) e recorrente (NLR). O superior controla os músculos cricotireóideos[6], e o recorrente os músculos restantes da laringe (PARRAGA, 2002). Uma lesão no nervo laríngeo superior poderá causar algumas alterações, como pigarrear ou tossir, fadiga ao falar e queda no tom de voz. A

[5]Formação ou desenvolvimento em excesso de um tecido ou um órgão.
[6]Músculos da laringe responsáveis por elevar a altura da voz

dobra vocal fica discretamente flácida e arqueada. Quando a lesão ocorre no nervo laríngeo recorrente, os sintomas são fonatórios ou obstrutivos, e pode ser unilateral ou bilateral, e esta ultima pode ser do tipo adutor ou abdutor. A dobra vocal apresenta flacidez no início da paralisia. Após algumas semanas, a massa muscular da dobra vocal torna-se mais rígida (PINTO, 1997).

Em adultos, as principais causas de paralisia unilateral de dobra vocal são: viral ou idiopático, pós-cirúrgico, doença maligna do pescoço, pós-intubação endotraqueal, trauma cervical, entre outras. Em se tratando da paralisia bilateral, a principal causa em adultos tem sido a tireoidectomia[7]. Entre outras causas, ainda podem ser citadas doenças malignas do pescoço, pós-intubação endotraqueal, traumacervical, doença neurológica ou alguma causa desconhecida (COSTA, 2008).

2.3 Processamento Digital de Sinais de Voz

O processamento digital de sinais de voz, incluindo processamento de voz e detecção de patologias, é dividido em 3 etapas principais (Figura 2.8), descritos a seguir:

1. Pré-processamento: consiste na aquisição, pré-ênfase e janelamento do sinal de voz;
2. Extração de características: etapa em que são extraídas características que representem o sinal de voz, para posterior classificação;
3. Treinamento e Classificação: na fase de treinamento, um classificador é treinado a partir de um conjunto de características. Na fase de classificação, dado um vetor de características de entrada, o classificador irá retornar como saída a classe a qual esse vetor pertence.

As três etapas estão descritas nas subseções a seguir.

2.3.1 Pré-processamento

Esta etapa é dividida em 3 sub-etapas: aquisição, pré-ênfase e janelamento do sinal de voz. Na fase de aquisição, o sinal é obtido por uma entrada de áudio (e.g.: microfone), discreti-

[7]Cirurgia para retirada da tireóide

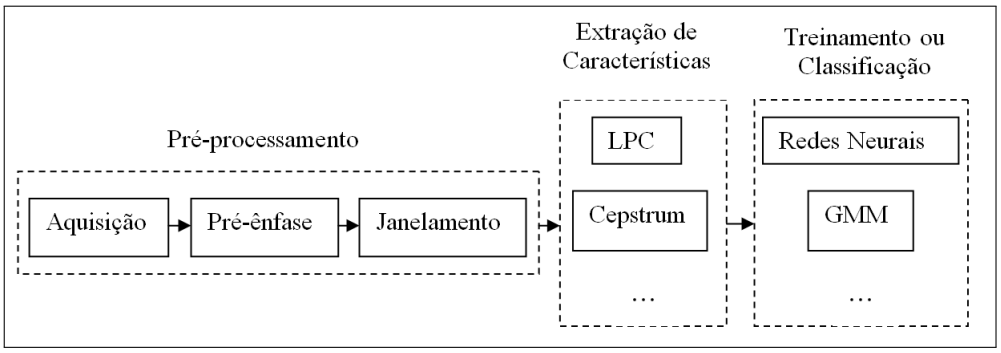

Figura 2.8: Etapas do processamento digital de sinais de voz.

zado para armazenamento em um meio físico (e.g.: computador).

Na pré-ênfase, é aplicado um filtro FIR de primeira ordem no sinal. No janelamento, o sinal de voz é particionado em segmentos com características estacionárias. Detalhes sobre a pré-ênfase e o janelamento podem ser encontrados a seguir.

Pré-ênfase

A etapa da pré-ênfase tem como objetivo atenuar os componentes de baixa frequência do sinal de voz, minimizando o efeito da radiação dos lábios e da variação da área da glote no sinal (SOTOMAYOR, 2003). A função transferência do sistema de pré-ênfase é dada por (JR.; HANSEN; PROAKIS, 2000):

$$L(z) = 1 - a_p^{-1}. \tag{2.1}$$

em que a_p é o fator de pré-ênfase, e tipicamente recebe valores próximos de 1,0 (RABINER; SCHAFER, 1978). O sinal de saída da pré-ênfase $s_p(n)$ está relacionado ao sinal de entrada $s(n)$ pela equação diferença (JR.; HANSEN; PROAKIS, 2000; RABINER; SCHAFER, 1978):

$$s_p(n) = s(n) - 0,95s(n-1). \tag{2.2}$$

Janelamento

Depois que a pré-ênfase é realizada, o sinal de voz é dividido em vários segmentos. Esta segmentação é importante devido ao sinal de voz variar estatisticamente com o tempo (ALENCAR, 2005), e ser aproximadamente estacionário a curtos intervalos de tempo, normalmente

entre 16 ms e 32 ms (SOTOMAYOR, 2003). Trabalhando com segmentos deste período, garante-se um sinal aproximadamente estacionário.

Essa divisão em segmentos é feita a partir da multiplicação do sinal por uma função janela. Essa multiplicação é feita no domínio do tempo. Portanto, o espectro do sinal janelado é a convolução do espectro do sinal com o espectro da função janela (HAYKIN; VEEN, 2002), ou seja, o janelamento modifica o sinal tanto no domínio do tempo quanto no domínio da frequência.

As funções janela mais comuns são as janelas Retangular, Hamming e Hanning (BRAGA, 2006). A janela Retangular (Equação 2.3 e Figura 2.9) consiste em simplesmente dividir o sinal de voz em segmentos de mesmo tamanho N_A. Esse é o janelamento que possui o maior volume de perda espectral.

$$J(n) = \begin{cases} 1 & \text{, se } 0 \leq \text{n} \leq N_A - 1; \\ 0 & \text{, caso contrário.} \end{cases} \tag{2.3}$$

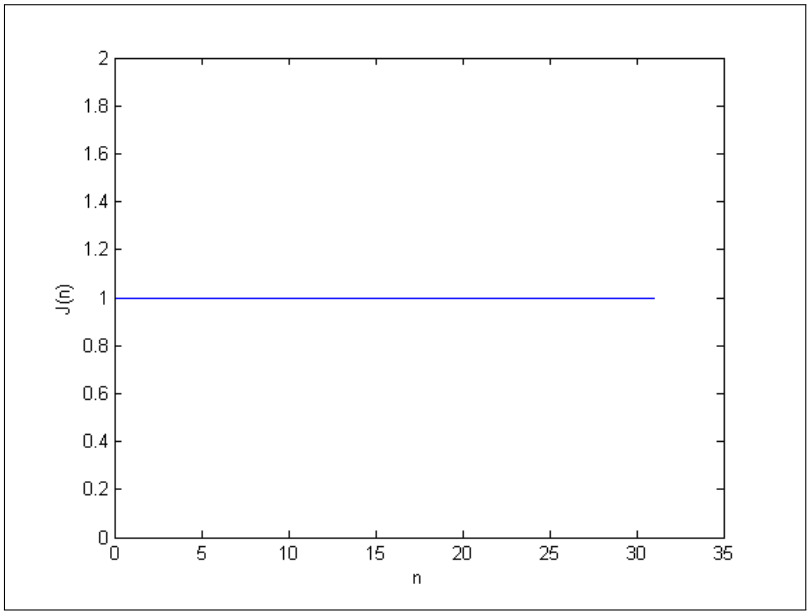

Figura 2.9: Janela Retangular com $N_A = 32$.

O janelamento de Hamming (Equação 2.4 e Figura 2.10) consiste em aplicar uma função janela cossenoidal ao sinal de voz, mantendo as características espectrais do centro do segmento, e eliminando as transições abruptas das extremidades. A janela de Hanning (Equação 2.5 e Figura 2.11) também possui caráter cossenoidal, mas permite uma maior suavização

das extremidades e um reforço menor no centro do segmento (FECHINE, 2000).

$$J(n) = \begin{cases} 0.54 - 0.46cos(\frac{2\pi n}{N_A - 1}) & \text{, se 0} \leq \text{n} \leq N_A - 1; \\ 0 & \text{, caso contrário.} \end{cases} \tag{2.4}$$

$$J(n) = \begin{cases} 0.5 - 0.5cos(\frac{2\pi n}{N_A - 1}) & \text{, se 0} \leq \text{n} \leq N_A - 1; \\ 0 & \text{, caso contrário.} \end{cases} \tag{2.5}$$

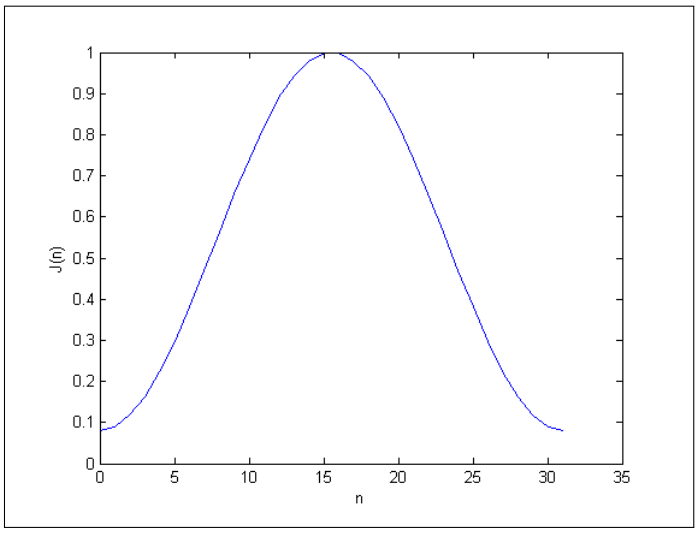

Figura 2.10: Janela de Hamming com $N_A = 32$.

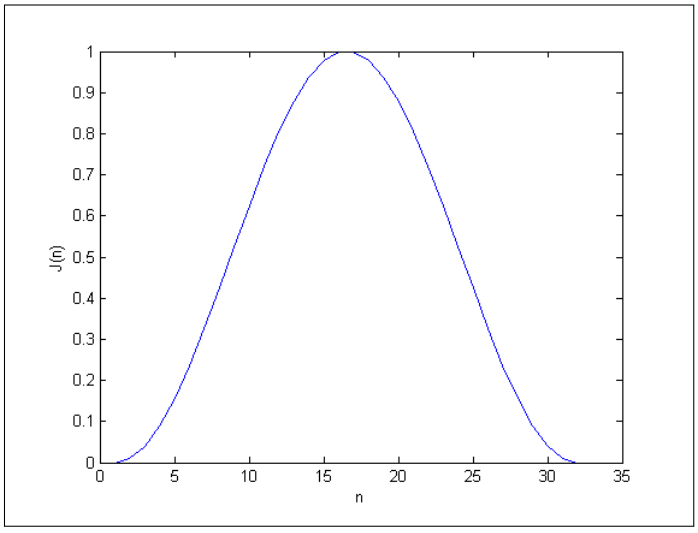

Figura 2.11: Janela de Hanning com $N_A = 32$.

Como as janelas de Hamming e de Hanning proporcionam perdas espectrais nas extremidades, faz-se necessário utilizar-se de sobreposição dos segmentos para que essas características não se percam quando as janelas são utilizadas.

2.3.2 Extração de Características

A etapa de extração de características é importante para proporcionar uma melhor separação entre as classes. Para isto, as características extraídas devem representar o sinal de voz de forma a manter características importantes desse sinal.

As técnicas mais comuns para extração de características do sinal de voz são os métodos banco de filtros, que utilizam Transformada Rápida de Fourier (FFT), análise homomórfica (cepstrum) e os métodos de codificação por predição linear (LPC) (JR.; HANSEN; PROAKIS, 2000; RABINER; SCHAFER, 1978). A seguir, estão descritos métodos de obtenção dos coeficientes LPC e seus derivados (Cepstrais e Delta-cepstrais).

2.3.3 Coeficientes LPC

A codificação por predição linear é o conjunto de técnicas que visa obter uma estimação da voz amostrada a partir de uma combinação linear entre amostras de voz passadas e valores presentes e passados de uma entrada hipotética de um sistema, em que a saída desse sistema é o sinal de voz (COSTA, 1994).

O trato vocal é excitado durante a produção de voz por uma série de pulsos periódicos produzidos pelas dobras vocais no caso dos sons sonoros, e, no caso dos sons não-sonoros, por ar turbulento passando através das constrições do trato (ATAL; HANAUER, 1971).

O conjunto de vetores de características obtidos pela predição linear representa o trato vocal (RABINER; SCHAFER, 1978) cujo o sistema linear (Figura 2.12) é excitado por pulsos quase periódicos (sons sonoros) ou ruído aleatório (sons não-sonoros) (COSTA, 1994). A intensidade do sinal sonoro é determinada por um ganho G. Outra característica importante da codificação por predição linear reside no fato de ela combinar os efeitos da excitação glotal, do trato vocal e da radiação (ATAL; HANAUER, 1971).

As amostras de som $s(n)$ são relacionadas com o termo $Gu(n)$, que pode ser uma fonte de excitação sonora ou surda, conforme mostrado na Figura 2.12, pela Equação:

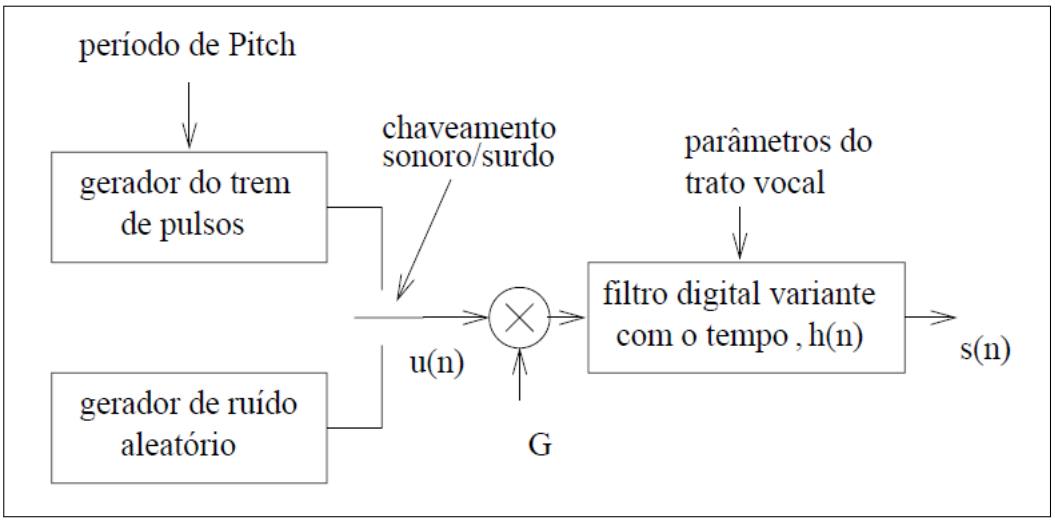

Figura 2.12: Modelo simplificado de produção de voz (FECHINE, 2000).

$$s(n) = \sum_{k=1}^{K} c_k s(n-k) + Gu(n). \tag{2.6}$$

Considerando as K amostras anteriores de $s(n)$, pode-se definir um preditor linear com coeficientes de predição c_k, como um sistema cuja saída é (RABINER; SCHAFER, 1978):

$$\tilde{s}(n) = \sum_{k=1}^{K} c_k s(n-k). \tag{2.7}$$

Dentre os métodos utilizados para resolução dessa Equação, tem-se o método da covariância (ATAL; HANAUER, 1971); o método da autocorrelação (MAKHOUL, 1975); a formulação do filtro inverso (RABINER; SCHAFER, 1978); a formulação da estimação espectral (RABINER; SCHAFER, 1978); a formulação da máxima verossimilhança (RABINER; SCHAFER, 1978) e a formulação do produto interno (MYERS; ALEKSANDER, 1989).

Neste trabalho, foi utilizado o método da autocorrelação, que é baseado na minimização do valor do erro de predição médio quadrático $e(n)$ dado por:

$$e(n) = s(n) = \tilde{s}(n) = s(n) - \sum_{k=1}^{K} c_k s(n-k). \tag{2.8}$$

Para resolução do problema, é selecionado inicialmente um segmento do sinal de voz por meio de uma janela de comprimento finito e igual a N_A. Esta medida visa limitar a extensão do sinal de voz em análise, assegurando sua estacionariedade. Os valores apropriados para N_A compreendem o intervalo entre 16 e 32 ms (RIBEIRO, 2003). O segmento $x(n)$

selecionado pela janela $j(n)$ é:

$$x(n) = j(n).s(n). \tag{2.9}$$

A partir da Equação 2.7, pode-se obter a predição linear do segmento $x(n)$:

$$\tilde{x}(n) = \sum_{k=1}^{K} c_k x(n-k). \tag{2.10}$$

A partir da Equação 2.8, o erro de predição $e(n)$ é obtido por:

$$e(n) = x(n) = \tilde{x}(n) = x(n) - \sum_{k=1}^{K} c_k x(n-k). \tag{2.11}$$

E o erro quadrático ε por:

$$\varepsilon = \sum_{n=-\infty}^{\infty} e(n)^2 = \sum_{n=-\infty}^{\infty} [x(n) - \sum_{k=1}^{K} c_k x(n-k)]^2. \tag{2.12}$$

A minimização do erro quadrático é realizada fazendo-se:

$$\frac{\partial(\varepsilon)}{\partial(c_k)} = 0 \quad , 1 \leq \text{k} \leq \text{K}. \tag{2.13}$$

Substituindo ε, definido na Equação 2.12 na Equação 2.13 e realizando K derivadas parciais, obtem-se:

$$\sum_{k=1}^{K} c_k R_r(|i-k|) = R_r(i) \quad , 1 \leq \text{k} \leq \text{K}, \tag{2.14}$$

com

$$R_r(k) = \sum_{n=0}^{N_A-K-1} x(n)x(n+k). \tag{2.15}$$

As Equações 2.14 e 2.15 também podem ser vistas na forma matricial (VIEIRA, 1989; AGUIAR-NETO, 1987; FECHINE, 2000):

$$\begin{vmatrix} R_r(0) & R_r(1) & ... & R_r(K-1) \\ R_r(1) & R_r(0) & ... & R_r(K-2) \\ R_r(2) & R_r(1) & ... & R_r(K-3) \\ ... & ... & ... & ... \\ R_r(K-1) & R_r(K-2) & ... & R_r(0) \end{vmatrix} \begin{vmatrix} c_1 \\ c_2 \\ c_3 \\ ... \\ c_K \end{vmatrix} = \begin{vmatrix} R_r(1) \\ R_r(2) \\ R_r(3) \\ ... \\ R_r(K) \end{vmatrix} \quad (2.16)$$

Essa matriz apresenta simetria de Toeplitz, ou seja, os elementos da diagonal principal e suas paralelas são constantes. Além disso, é uma matriz simétrica. Para solução do sistema, neste trabalho é utilizado o algoritmo de Levinson-Durbin (VIEIRA, 1989; SILVA, 1992).

2.3.4 Coeficientes Cepstrais

O objetivo da análise cepstral é a obtenção de uma relação linear entre a excitação da energia do sinal com o filtro utilizado (BRAGA, 2006). Os coeficientes cepstrais são usados para descrever a envoltória espectral do sinal de voz em um segmento. Eles podem ser obtidos por meio da FFT ou então a partir dos coeficientes LPC (TOLBA; O'SHAUGHNESSY, 1997). A obtenção dos coeficientes Cepstrais a partir da FFT se dá aplicando-se diretamente ao sinal de voz uma transformada inversa rápida de Fourier. Para obteção a partir dos coeficientes LPC, a transformada z é aplicada no sinal de voz modelado pela análise LPC.

As Equações 2.17 (RABINER; HUANG, 1993; AGUIAR-NETO; COSTA; FECHINE, 2008) e 2.18 (MAMMONE; ZHANG; RAMACHANDRAN, 1996; FECHINE, 2000; COSTA, 2008) descrevem as duas formas de obtenção dos coeficientes cepstrais.

$$c(n) = \frac{1}{N} \sum_{k=0}^{N-1} log[X(k)] e^{j2\pi kn/N} \quad n = 0, 1, ..., N-1. \quad (2.17)$$

em que X(k) é o espectro do sinal.

$$\begin{cases} c(1) = -c_k(1) \\ c(i) = -c_k(i) - \sum_{k=1}^{i-1}(1 - \frac{k}{i}) c_k(k) c(i-k) \qquad , 1 < i < p. \end{cases} \quad (2.18)$$

A Equação 2.18 é recursiva, levando a uma computação eficiente dos coeficientes Cepstrais, evitando assim uma fatoração polinomial. Neste trabalho, os coeficientes Cepstrais são obtidos a partir dos coeficientes LPC.

2.3.5 Coeficientes Delta-Cepstrais

Os coeficientes cepstrais fornecem uma boa representação das propriedades locais do sinal para um bloco de amostras de voz. Pode-se estender a análise cepstral para caracterizar a informação temporal com a introdução da derivada cepstral no espaço de características. A derivada cepstral objetiva capturar a informação de transição de voz. Neste trabalho, é utilizada apenas a primeira derivada do cepstrum, que é definida como (MAMMONE; ZHANG; RAMACHANDRAN, 1996; FECHINE, 2000):

$$\frac{\triangle c(n,t)}{\triangle t} = \triangle c_i(n) \approx \phi \sum_{q=-Q}^{Q} qc(n,t+q), \tag{2.19}$$

em que $c(n,t)$ é o n-ésimo coeficiente Cepstral no tempo t, ϕ é uma constante de normalização e $2Q+1$ é o número de blocos de amostras sobre os quais o cálculo é realizado

Neste trabalho, é utilizada uma versão simplificada da Equação 2.19 (MAMMONE; ZHANG; RAMACHANDRAN, 1996):

$$\triangle(c_i(n)) = [\sum_{q=-Q}^{Q} qc_{i-q}(n)]G \quad , 1 \leq n \leq K. \tag{2.20}$$

em que G é o termo de ganho ($= 0,375$), K é o número de coeficientes Delta-cepstrais, $Q = 2$, *n* o índice do coeficiente e *i* o índice do bloco de amostras.

2.4 Técnicas de Classificação

Depois de obtidos os vetores de características dos sinais de voz, é formado um conjunto de treinamento, que é aplicado a um classificador, que permitirá o aprendizado e, consequente, diferenciação entre as diversas classes. No contexto deste trabalho, objetiva-se diferenciar, mediante o classificador, voz normal, voz afetada por edema de Reinke, e voz afetada por outra patologia.

Alguns classificadores utilizados na literatura utilizam Quantização Vetorial (COSTA, 2008; AGUIAR-NETO; COSTA; FECHINE, 2008); Modelos de Misturas de Gaussianas (GODINO-LLORENTE; GÓMEZ-VILDA; BLANCO-VELASCO, 2006); e Redes Neurais MLP (GODINO-LLORENTE; GÓMEZ-VILDA, 2004; MARTINEZ; RUFINER, 2000). Esses foram os classificadores utilizados neste trabalho, e a descrição de cada um deles é dada a seguir.

2.4.1 Quantização Vetorial

Quantização Vetorial (*Vector Quantization* - VQ) é um método de compressão baseado no princípio da codificação por blocos. Linde, Buzo e Gray (LBG) (LINDE; BUZO; GRAY, 1980) propuseram um algoritmo para projeto e construção do dicionário baseado em uma seqüência de treinamento. Uma VQ que seja projetada utilizando este algoritmo é referida na literatura como sendo uma VQ-LBG. A idéia principal da Quantização Vetorial consiste em obter um conjunto ótimo de vetores que represente os vetores de características tal que a distorção obtida pela substituição dos vetores de treinamento pelos vetores do dicionário seja mínima (RABINER; LEVINSON; SONDHI, 1983).

O primeiro passo do algoritmo LBG é determinar o número N de vetores código, ou seja, o tamanho do dicionário. Depois disso, são escolhidos aleatoriamente N vetores código dentre os vetores de características, e eles são definidos como o dicionário inicial. O próximo passo consiste em usar a medida de distância Euclidiana para agrupar os vetores em torno de cada vetor código. Em seguida, o novo conjunto de vetores código é calculado a partir da média de cada grupo. As duas últimas etapas são repetidas até que as mudanças dos vetores código sejam menores que um limiar, formando assim o dicionário final (LINDE; BUZO; GRAY, 1980; FECHINE, 2000).

2.4.2 Redes Neurais

As principais células que formam o Sistema Nervoso são os neurônios. Neurônios são células formadas por três elementos com funções específicas e complementares: corpo, dentritos e axônio. Os dentritos são responsáveis por captar estímulos recebidos em um determinado período de tempo e os transmitir ao corpo do neurônio, onde serão processados. Quando esses estímulos atingem um determinado limite, o corpo da célula envia um novo impulso que se propaga pelo axônio, que o transmite para os dentritos das células vizinhas por meio de processos chamados de sinapses (GUYTON; HALL, 1997).

Uma Rede Neural Artificial (RNA) é uma técnica computacional projetada para imitar a maneira a partir da qual o cérebro desempenha uma tarefa em particular. As RNA caracterizam-se por possuírem elementos de processamento de estrutura simples, com conexões entre eles. Cada conexão tem um peso associado. Este peso representa a intensidade

de interação ou acoplamento entre os elementos interligados e sua natureza é excitatória ou inibitória (HAYKIN, 2001).

Um dos primeiros modelos de RNA desenvolvido foi o Perceptron. O Perceptron consiste basicamente de um único neurônio com pesos sinápticos ajustáveis e uma polarização (*bias*) (Figura 2.13). O Perceptron possui a capacidade de convergir e separar duas classes atrás de um hiperplano, podendo assim separar duas classes que sejam linearmente separáveis (ROSENBLATT, 1958). O Perceptron de um único neurônio é limitado a classificar apenas entre duas classes. Para mais de duas classes, é necessário realizar uma expansão da capacidade computacional de saída do Perceptron, incluindo mais que um neurônio. Entretanto, essas classes têm que ser linearmente separáveis para que o Perceptron tenha desempenho adequado.

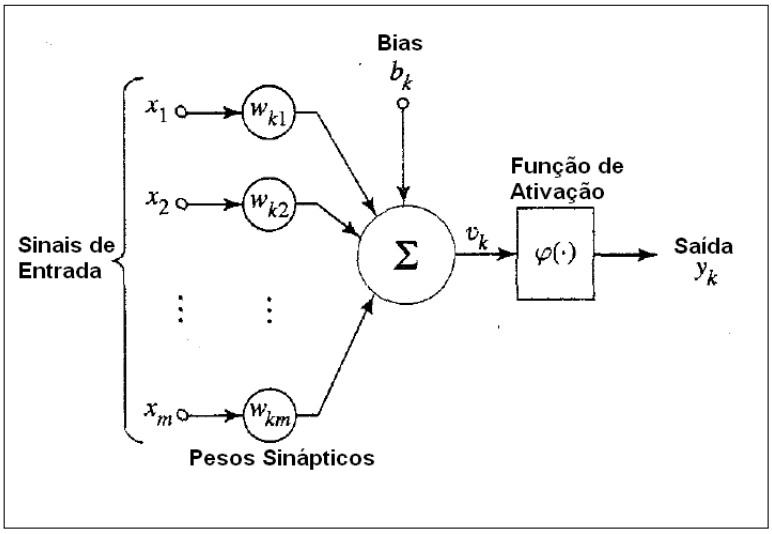

Figura 2.13: Perceptron.

Apaptado de (HAYKIN, 2001).

O Perceptron possui aprendizado supervisionado, ou seja, é apresentado a ele pares de entradas e saídas desejadas, e para cada entrada, ele irá produzir uma saída que será comparada com a saída desejada, ajustando assim os pesos conforme o erro obtido. O neurônio computa uma combinação linear das entradas aplicadas as suas sinapses com os pesos associados, e também incorpora o *bias*. A soma resultante é aplicada a uma função de ativação, que irá produzir uma saída igual a +1 ou a -1.

As Redes Neurais Multilayer Perceptron (MLP) são redes compostas de vários neurônios

do tipo perceptron com pesos sinápticos ajustáveis e uma polarização (bias). Esses neurônios estão dispostos em N camadas (N > 2), em que uma camada é a camada de entrada, $N - 2$ camadas são as camadas escondidas e a ultima camada é a camada de saída (Figura 2.14). As camadas escondidas permitem que as redes MLP possam solucionar problemas não-linearmente separáveis, sendo assim mais robustas que as redes Perceptron simples.

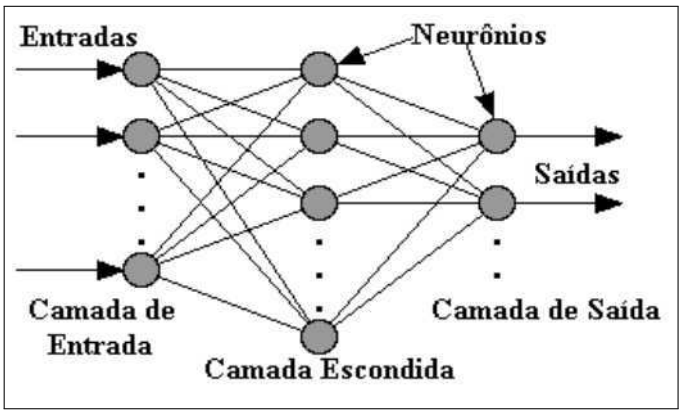

Figura 2.14: Rede MLP com 3 camadas.

Em se tratando do vetor de características, esse é dado para a camada de entrada, que o propaga até a camada de saída, então a saída do sistema é comparada com a saída desejada, caracterizando assim um treinamento supervisionado, e o erro é propagado de volta ajustando-se os pesos a partir do algoritmo *backpropagation*.

Cada neurônio de uma Rede Neural MLP é conectado a todos os neurônios da camada anterior, se houver, e a todos da camada posterior, se houver. Para cada conexão, é atribuído um peso aleatório inicial. A saída de um neurônio obedece a Equação 2.21.

$$y = f(\sum_{j=0}^{n} w_j x_j). \tag{2.21}$$

em que w_j é o peso da conexão com o neurônio j, e x_j é a saída do neurônio j. A função $f(x)$ é a função de ativação do neurônio. Entre as diversas funções de ativação utilizadas, pode-se citar a função sigmóide (Equação 2.22).

$$f(x) = \frac{1}{(1 + e^{-ZX})}. \tag{2.22}$$

A saída do neurônio é propagada como uma das entradas do próximo neurônio, ou uma parte da saída final da rede, caso esteja na ultima camada. Por fim, é aplicado o algoritmo

backpropagation para o ajuste dos pesos das conexões. Os pesos são adaptados conforme a Equação 2.23.

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj}, \tag{2.23}$$

em que $w_{ij}(t)$ é o peso atual entre os neurônios i e j, η é a taxa de aprendizagem, o_{pj} é a saída do neurônio, e δ_{pj} é o termo de erro, que é calculado pela Equação 2.24.

$$\delta_{pj} = \begin{cases} o_{pj}(1 - o_{pj})(t_{pj} - o_{pj}) & \text{, para neurônios da camada de saída;} \\ o_{pj}(1 - o_{pj}) \sum_k \delta_{pk} w_{jk} & \text{, para neurônios das camadas escondidas.} \end{cases} \tag{2.24}$$

t_{pj} é a saída esperada e k são os neurônios da camada posterior.

O treinamento se repete até que a condição de parada seja satisfeita.

2.4.3 Modelo de Misturas de Gaussianas

Modelo de Misturas de Gaussianas (*Gaussian Mixture Models* - GMM) é uma combinação linear de funções de densidade de probabilidade gaussianas (REYNOLDS, 1995). O GMM pode ser contínuo ou discreto. Neste trabalho foi utilizado o GMM discreto.

Um modelo contém G componentes, cada um representado por uma função de densidade de probabilidade (Equação 2.25).

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{C/2}|\sigma_i|^{1/2}} exp\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^T \sigma_i^{-1}(\vec{x} - \vec{\mu}_i)\} \quad \text{, para i = 1, 2, ..., G,} \tag{2.25}$$

em que C é a dimensão da função de densidade de probabilidade, que é a mesma do vetor de características $\vec{x}$, $\vec{\mu}_i$ o vetor de média e σ_i a matriz de covariância.

O modelo é representado por $\lambda = \{\rho_i, \vec{\mu}_i, \sigma_i\}, i = 1, 2, ..., G$, em que ρ_i é o peso de cada componente no modelo. Cada modelo é obtido a partir de um número elevado de segmentos, fazendo com que o modelo não represente nenhum segmento em particular, mas sim características comuns a todos eles, podendo assim considerar que um modelo pode representar uma classe de patologia, como edema, ou uma classe de voz normal.

Ao se inserir um vetor de características $\vec{x}$ em um GMM, obtém-se a probabilidade $p(\vec{x}/\lambda)$, que é a probabilidade de ocorrência do vetor $\vec{x}$ considerando o modelo λ. Cada

componente contribui na obtenção dessa probabilidade a partir de seu peso ρ_i. O cálculo da probabilidade de um vetor de características $\vec{x}$ ter sido gerado por um determinado modelo λ é feito a partir da Equação 2.26.

$$p(\vec{x}/\lambda) = \sum_{i=1}^{G} \rho_i b_i(\vec{x}). \tag{2.26}$$

Para treinamento do modelo, obtém-se o conjunto de vetores de características $X = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_T\}$, que contém as propriedades intrísecas da classe, nesse caso, de uma patologia da voz. Considerando, para simplificação matemática, que os vetores de características $\vec{x}_i$ são independentes, a probabilidade do conjunto X pertencer ao modelo λ é:

$$p(X/\lambda) = p(\vec{x}_1, \vec{x}_2, ..., \vec{x}_T/\lambda) = p(\vec{x}_1/\lambda)p(\vec{x}_2/\lambda)...p(\vec{x}_T/\lambda) = \prod_{i=1}^{T} p(\vec{x}_i/\lambda). \tag{2.27}$$

O treinamento objetiva maximizar a probabilidade de observação do conjunto de treinamento X. A princípio, a resolução da equação $\frac{\partial p(X/\lambda)}{\partial \lambda} = 0$ forneceria os subsídios para a determinação do modelo de máxima verossimilhança $p(X/\lambda)$. Entretanto, a resolução direta não é realizável devido a não-linearidade dos parâmetros do modelo λ (CARDOSO, 2009). Para tanto, é utilizado o conhecido algoritmo *expectation-maximization* (EM). A idéia do EM é inicializar o modelo λ, e reestimar os parâmetros para cada iteração até convergir para um valor máximo. A reestimação é realizada utilizando as equações a seguir:

$$\rho_i = \frac{1}{T} \sum_{t=1}^{T} p(i/\vec{x}_t, \lambda) \quad \text{, para i = 1, 2, ..., G.} \tag{2.28}$$

$$\vec{\mu}_i = \frac{\sum_{t=1}^{T} p(i/\vec{x}_t, \lambda)\vec{x}_t}{\sum_{t=1}^{T} p(i/\vec{x}_t, \lambda)} \quad \text{, para i = 1, 2, ..., G.} \tag{2.29}$$

$$\sigma_i = \frac{\sum_{t=1}^{T} p(i/\vec{x}_t, \lambda)(\vec{x}_t - \vec{\mu}_i)(\vec{x}_t - \vec{\mu}_i)^T}{\sum_{t=1}^{T} p(i/\vec{x}_t, \lambda)} \quad \text{, para i = 1, 2, ..., G.} \tag{2.30}$$

$$p(i/\vec{x}_t, \lambda) = \frac{\rho_i p_i(\vec{x}_t)}{\sum_{k=1}^{T} \rho_k p_k(\vec{x}_t)} \quad \text{, para i = 1, 2, ..., G.} \tag{2.31}$$

Os parâmetros iniciais do modelo para o treinamento pelo algoritmo EM são obtidos a partir do dicionário gerado pela Quantização Vetorial. Esse dicionário é utilizado como sendo as médias $\vec{\mu}$ iniciais dos G componentes do modelo. As matrizes de covariância $\vec{\sigma^2}$

iniciais foram obtidas utilizando-se as médias $\vec{\mu}$ e os vetores $\vec{x}$ agrupados para cada componente, em que V_i é o número de vetores pertencentes ao conjunto i:

$$\vec{\sigma_i^2} = \frac{1}{V_i - 1} \sum_{n=1}^{V_i} (\vec{x}_{(i,n)} - \vec{\mu_i})^2 \quad \text{, para i = 1, 2, ..., G.} \tag{2.32}$$

Os pesos $\vec{p_i}$ de cada componente são obtidos calculando a razão do conjunto de vetores pertencentes a cada conjunto pelo número total de vetores. Após inicializado, o modelo é treinado utilizando o algoritmo EM. Por fim, a classificação é efetuada calculando-se a probabilidade de um vetor de características pertencer a classe modelada por meio da Equação 2.26. Caso a probabilidade obtida esteja acima de um limiar pré-estabelecido, o vetor de características é classificado como sendo da classe modelada. Caso a probabilidade esteja abaixo do limiar, o vetor de características é classificado como não pertencente a classe modelada.

Capítulo 3

Trabalhos Relacionados

Neste capítulo, são descritos trabalhos relacionados ao diagnóstico de patologias da laringe por meio do processamento digital de voz. O escopo da revisão consiste no estudo de técnicas de extração de características e classificação de vozes afetadas por patologias, não sendo abordados trabalhos no âmbito do processamento digital de voz afetada por patologia com outros enfoques, a exemplo do cálculo da energia do ruído de vozes afetadas por patologias (MANFREDI, 2000) e avaliações de qualidade da voz afetada por patologia (RITCHINGS; MCGILLION; MOORE, 2002; SIMM; ROBERTS; JOYCE, 2005). Também não foram descritos trabalhos que usam base de dados não convencionais para esse tipo de pesquisa, como por exemplo, base de dados obtida por vídeo (VOIGT et al., 2009) e obtida via telefone (MORAN et al., 2006). Por fim, não serão descritos trabalhos que tratam de patologias que afetam a voz, mas que não são da Laringe, como por exemplo refluxo gástrico (DIBAZAR; BERGER; NARAYANAN, 2006).

Godino-Llorente e Gómez-Vilda (2004) realizaram um estudo comparativo entre dois tipos diferentes de Redes Neurais, as redes MLP e aprendizagem por Quantização Vetorial (LVQ - *Learning Vector Quantization*), para a classificação de voz afetada por patologia e voz normal. Na fase de pré-processamento, não é utilizado pré-ênfase. O vetor de características utilizado contém coeficientes mel-cepstrais, suas primeira e segunda derivadas, e a energia. É utilizado um vetor de características para cada janela do sinal. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994), a mesma base de dados utilizada neste trabalho. Ao todo, foram utilizados 53 arquivos de voz normal e 82 de voz afetada por patologia, todos com a vogal sustentada /a/. Foi obtido

como resultado uma taxa de acerto de 96% para as redes LVQ e 94% para redes MLP.

Fredouille et al. (2005) objetivaram classificar entre voz normal e voz afetada por patologia, e também avaliar a qualidade da voz afetada por patologia. Para diferenciar entre voz normal e voz afetada por patologia, foram utilizados coeficientes mel-cepstrais e sua primeira derivada. O classificador utilizado foi o GMM. A base de dados possui 20 arquivos de voz normal e 60 arquivos de voz afetada por patologia, todos com a vogal sustentada /a/. Para treinamento e classificação, foi utilizado o método *Leave-x-out*, o classificador foi treinado com todos os casos, exceto x casos que foram utilizados na classificação. Esse processo é repetido até que todos os casos sejam testados. Foi escolhido $x = 2$ para o conjunto de voz normal, e $x = 42$ para o conjunto de vozes afetadas por patologia. A taxa de acerto foi de 85%.

Godino-Llorente, Gómez-Vilda e Blanco-Velasco (2006) utilizaram GMM para diferenciação de voz normal e voz afetada por patologia. Novamente não foi utilizado pré-ênfase, e o vetor de características foi o mesmo, formado por coeficientes mel-cepstrais, suas primeira e segunda derivadas e a energia. A classificação foi realizada por janela, como no trabalho anterior. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Ao todo, foram utilizados 53 arquivos de voz normal e 173 de voz afetada por patologia, todos da vogal sustentada /a/. A taxa de acerto obtida foi de 94%, na mesma faixa obtida pelas redes MLP anteriores, e abaixo da obtida pelas redes LVQ, mas com a vantagem de que o tempo de treinamento é menor.

Little et al. (2006) utilizaram características não-lineares para classificação entre voz normal e voz afetada por patologia. O vetor de características é formado por um valor de entropia obtido pelo algoritmo *Return Period Density Entropy* (RPDE) e por um valor α obtido pelo algoritmo *Detrended fluctuation analysis* (DFA). O classificador utilizado é um discriminador linear gaussiano. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Foram utilizados 53 arquivos de voz normal e 654 de voz afetada por patologia, todos da vogal sustentada /a/. A taxa de acerto foi de 91,4%.

Kukharchik et al. (2007) pesquisaram o uso da transformada wavelet para extração de características. Eles utilizaram o algoritmo da transformada wavelet contínua (MALLAT, 1998) para obtenção dos coeficientes e máquinas de suporte vetorial (*Support Vector Machine* -

SVM) para a classificação entre voz afetada por patologia e voz normal. Não foi utilizada pré-ênfase. A base de dados utilizada foi obtida do *Republican Center of Speech, Voice and Hearing Pathologies*. Não foi utilizada a vogal sustentada /a/, mas sim textos gravados por pacientes. Ao todo, são 70h de voz normal e 20h de voz afetada por patologia, de 118 sujeitos. O sistema conta ainda com um detector de vogal. Apenas segmentos com vogais foram utilizados para treinamento e teste. Os resultados obtidos foram excelentes, tendo uma taxa de acerto de 99%.

Falcão et al. (2008) empregaram uma abordagem não-linear para classificar entre voz normal e voz afetada por patologia. Na extração de características, foram utilizadas as entropias de Shannon, Relatica e de Tsallis. Para classificação, foi utilizada a Quantização Vetorial. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Foram utilizados 53 arquivos de voz normal e 63 de voz afetada por patologia, todos da vogal sustentada /a/. O melhor resultado obtido foi 80% para entropia de Tsallis.

Salhi, Talbi e Cherif (2008) utilizaram coeficientes baseados na transformação wavelet como características, e Redes Neurais MLP para classificação. Foram utilizados coeficientes obtidos a partir da transformada wavelet discreta, coeficientes de energia wavelet, coeficientes de entropia wavelet, e coeficientes obtidos a partir da transformada wavelet contínua. A base de dados foi obtida a partir do Laboratório G.E. da Universidade de Los Angeles e do RABTA Hospital de Tunis. Ao todo, são 50 arquivos de voz normal e 50 arquivos de voz afetada por patologia, cada um de uma palavra pronunciada por um locutor diferente. A melhor taxa de acerto foi de 95%, utilizando coeficientes de entropia wavelet e coeficientes obtidos a partir da transformada wavelet contínua.

Paulra et al. (2010) utilizaram um classificador Fuzzy simples para detectar e retirar o silêncio dos sinais de voz. Na parte restante, são obtidos 12 parâmetros a partir do algoritmo *Tri Mean Relative average pertubation*, que formam o vetor de características. O objetivo do trabalho é classificar entre voz normal e voz afetada por patologia, para isso, foi utilizada uma Rede Neural MLP. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Foram utilizados 53 arquivos de voz normal e 257 de voz afetada por patologia, todos da vogal sustentada /a/. A base de dados foi aumentada para 176 arquivos de voz normal e 470 de voz afetada por patologia ao

inserir ruído gaussiano gerado aleatoriamente. A melhor taxa de acerto foi de 92,76%.

Scholotthauer, Torres e Rufiner (2010) propuseram duas modificações para o algoritmo de Decomposição de modo empírico. A primeira permite a extração robusta da frequência fundamental em sinais de vogal sustentada, e a segunda modificação é aplicada para extração de características para classificação entre voz normal e voz afetada por patologia. O critério de classificação foi o K-vizinhos mais próximos. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Foram utilizados 53 arquivos de voz normal e 53 de voz afetada por patologia, todos da vogal sustentada /a/. Foi criada também uma base de dados artificial, por meios de modelos de fonação de voz normal e de voz afetada por patologia. A base de dados artifical contém 200 arquivos de voz normal e 200 arquivos de voz afetada por patologia. A taxa de acerto foi de 93,4% para a base de dados real e 99% para a base de dados artificial.

Esses trabalhos têm como ponto em comum o fato de classificarem os sinais de voz em voz normal ou voz afetada por patologia, sem fazer qualquer distinção sobre qual patologia o paciente possui. Alguns trabalhos foram feitos no sentido de detectar doenças específicas nos sinais de voz. Martinez e Rufiner (2000) propuseram um sistema para detecção de patologias da laringe utilizando coeficientes cepstrais, mel-cepstrais, delta-cepstrais, delta mel-cepstrais e coeficientes obtidos a partir da Transformada Rápida de Fourier (*Fourier Fast Transform* - FFT) como características extraídas e Redes Neurais MLP. Dois tipos de Redes Neurais foram utilizados. Uma das redes foi treinada para diferenciar voz afetada por patologia de voz normal, e outra foi treinada para distinguir entre voz normal, voz rouca e voz com modulação bicíclica. Os coeficientes delta-cepstrais e delta-mel-cepstrais só foram utilizados para classificação entre diferentes patologias. A base de dados para voz normal foi obtida por meio do TIMIT *continuous speech corpus* (GAMFOLO et al., 1993). A base de dados de voz afetada por patologia foi obtida por meio do *Speech Processing and Auditory Perception Laboratory* (ALWAN et al., 1995). Ao todo, foram utilizados 8 arquivos de voz normal, 13 arquivos de voz rouca e 9 arquivos de voz com modulação bicíclica, utilizando a vogal sustentada /a/. Esse sistema obteve uma taxa de acerto de 91,3% usando coeficientes cepstrais para diferenciar voz afetada por patologia de voz normal, e 87,7% com os coeficientes delta cepstrais quando usado para classificar entre as 3 classes diferentes.

Crovato (2004) propôs um método de classificação em 6 grupos de voz, em que 5 são

grupos patológicos (laringite crônica, degenerativo, mobilidade incorreta, alterações orgânicas e crescimento orgânico) e 1 grupo é formado por vozes normais. São utilizados coeficientes extraídos de pacotes wavelets. Para classificação, foram criadas 6 redes MLP, uma para cada grupo, cujo objetivo de cada rede era diferenciar entre duas classes: o grupo em específico da rede, ou outros grupos. A base de dados foi obtida pelo Hospital da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), com 13 arquivos do grupo laringite crônica, 7 do degenerativo, 10 de mobilidade incorreta, 5 de alterações orgânicas, 7 de crescimentos orgânicos e 13 de voz normal, todos da vogal sustentada /a/. Os resultados variaram entre 87,5% para o grupo de laringite crônica até 100% para o grupo de alterações orgânicas, grupo este em que se encontra a patologia estudada neste trabalho. Por causa do pequeno tamanho da base de dados, não foi utilizado grupo de teste, às taxas de acerto são referentes as taxas de acerto dos grupos de treinamento, sendo o objetivo do trabalho apenas validar o método.

O trabalho de Scholotthauer e Torres (2006) teve como objetivo diferenciar entre voz afetada por disfonia espasmódica, que possui causa neurológica, de voz afetada por disfonia de tensão muscular, que possui causa psicológica, e voz normal. Essas duas patologias possuem sintomas similares, fazendo com que possam ser facilmente confundidas uma com a outra, e, pelo fato de possuir causas diferentes, possuem tratamentos diferentes. O vetor de características utilizado possui 8 dimensões, sendo formado por: grau de paradas da voz, *jitter*, perturbação média relativa, quociente de perturbação do periodo de 5 pontos, *shimmer*, quociente de perturbação da amplitude de 3 pontos, quociente de perturbação da amplitude de 11 pontos, taxa de ruído. Para classificação, é utilizada Redes Neurais MLP. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Foram utilizados 53 arquivos de voz normal, 21 de voz afetada por disfonia espasmódica, e 15 de voz afetada por disfonia de tensão muscular, todos da vogal sustentada /a/. Foi aplicado o método *Leave-one-out*, sendo a Rede Neural treinada com todos os casos, exceto o caso que será classificado. O melhor resultado obtido foi 93,26% de acerto, sendo 100% para voz normal, 80,95% para voz afetada por disfonia espasmódica e 86,67% para voz afetada por disfonia de tensão muscular.

O trabalho de Fonseca (2008) objetivou diferenciar entre voz normal e voz afetada por patologia, voz normal e voz afetada por nódulo vocal, voz normal e voz afetada por edema de

Reinke, e voz afetada por nódulo e voz afetada por edema de Reinke. Para isso, na extração de características foi utilizada a Transformada Wavelet Discreta. Para diferenciar entre voz normal e voz afetada por alguma patologia (nódulo ou edema ou ambas), foi aplicado um filtro de predição linear inverso nos componentes wavelet, e para diferenciar entre voz afetada por edema e voz afetada por nódulo foi calculado o valor de *jitter* dos componentes wavelet. O classificador utilizado foi a Máquina de Suporte Vetorial (*Support Vector Machine* - SVM). A base de dados foi obtida do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP), com 30 arquivos de voz normal, 30 de voz afetada por nódulo e 16 de voz afetada por edema, todos da vogal sustentada /a/. As melhores taxas de acerto foram 90,1% para voz normal *versus* voz afetada por nódulo, 85,3% para voz normal *versus* voz afetada por edema, 88,2% para voz normal *versus* voz afetada por patologia e 82,4% para voz afetada por edema *versus* voz afetada por nódulo.

Aguiar-Neto, Costa e Fechine (2008) utilizaram coeficientes LPC, cepstrais e mel-cepstrais para três tipos diferentes de diferenciação de classes: voz afetada por edema de Reinke *versus* voz normal, voz afetada por edema de Reinke *versus* voz afetada por outra patologia e voz afetada por patologia *versus* voz normal. Para classificação, foi utilizada a Quantização Vetorial. A base de dados utilizada foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Ao todo, foram utilizados 53 arquivos de voz normal, 44 de voz afetada por edema e 23 de voz afetada por outra patologia, todos da vogal sustentada /a/. Os resultados foram muito bons para separar voz normal de voz afetada por patologia e de voz afetada por edema, obtendo uma taxa de acerto de cerca de 95% para diferenciar entre voz normal e voz afetada por patologia utilizando coeficientes LPC e coeficientes mel-cepstrais, e 99% para diferenciar entre voz normal e voz afetada por edema, utilizando coeficientes LPC. Para diferenciar entre voz afetada por edema e voz afetada por outra patologia, o melhor resultado obtido foi de 83%, utilizando coeficientes LPC.

Costa (2008) estendeu o trabalho anterior (AGUIAR-NETO; COSTA; FECHINE, 2008), refinando o processo de classificação com o uso de Modelos de Markov Escondidos (*Hidden Markov Models* - HMM) como uma segunda etapa de classificação após a Quantização Vetorial. Além dos coeficientes já usados anteriormente (LPC, cepstrais e mel-cepstrais), foram utilizados também coeficientes cepstrais ponderados, delta-cepstrais e delta-cepstrais

ponderados. Após a classificação utilizando Quantização Vetorial, os casos que não foram classificados corretamente foram submetidos ao HMM. As melhores taxas após a etapa de refinamento foram de 99% para diferenciar entre voz normal e voz afetada por patologia, utilizando coeficientes LPC, delta-cepstrais e cepstrais ponderados, 100% para diferenciar entre voz normal e voz afetada por edema, utilizando coeficientes LPC, e 96% para diferenciar entre voz afetada por edema e voz afetada por outra patologia, também utilizando coeficientes cepstrais.

Markaki e Stylianou (2009) pesquisaram o uso do espectro de modulação do sinal de voz como característica extraída. Cada espectro obtido de cada janela consiste em 257 frequências acústicas e 257 frequências modulares, tendo portanto dimensionalidade 257x257. Foi utilizado o algoritmo *Higher Order Singular Value Decomposition* para reduzir a dimensionalidade para 34x34 $= 1156$ parâmetros. O trabalho objetiva classificar voz normal *versus* voz afetada por patologia, voz afetada por pólipo *versus* voz afetada por nódulo, voz afetada por pólipo *versus* voz afetada por leucoplasia e voz afetada por pólipo *versus* voz afetada por disfonia espasmódica. Para classificação, foi utilizado SVM. A base de dados foi obtida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMETRICS, 1994). Para classificar entre voz normal e voz afetada por patologia, foram utilizados 53 arquivos de voz normal e 173 de voz afetada por patologia. Para classificar entre as patologias, foram utilizados 88 arquivos de voz afetada por alguma das patologias em questão (pólipo, nódulo, leucoplasia e disfonia espasmódica). A taxa de acerto foi de 94,07% para voz normal *versus* voz afetada por patologia, 82,5% para voz afetada por pólipo *versus* voz afetada por disfonia espasmódica, 81,8% para voz afetada por pólipo *versus* voz afetada por leucoplasia e 87,5% para voz afetada por pólipo *versus* voz afetada por nódulo.

Na Tabela 3.1 é apresentado um resumo das principais características dos trabalhos relacionados.

Tabela 3.1: Resumo das características dos trabalhos relacionados.

Autores	Classificação	Base de Dados	Características	Classificadores	Taxa de Acerto
Godino-Llorente e Gómez-Vilda (2004)	Normal x Patologia	53x82 MEEI	MCep + $\Delta MCep$ + $\Delta\Delta MCep$ + E	MLP, LVQ	96%
Fredouille et al. (2005)	Normal x Patologia	20x60 Prop.	MCep + $\Delta MCep$	GMM	85%
Godino-Llorente, Gómez-Vilda e Blanco-Velasco (2006)	Normal x Patologia	53x82 MEEI	MCep + $\Delta MCep$ + $\Delta\Delta MCep$ + E	GMM	94%
Little et al. (2006)	Normal x Patologia	53x654 MEEI	RPDE + DFA	DLG	91,4%
Kukharchik et al. (2007)	Normal x Patologia	70hx20h RCSVP	TWC	SVM	99%
Falcão et al. (2008)	Normal x Patologia	53x63 MEEI	eS, eR, eT	QV	80%
Continua na página seguinte					

Tabela 3.1 – Continuação

Autores	Classificação	Base de Dados	Características	Classificadores	Taxa de Acerto
Salhi, Talbi e Cherif (2008)	Normal x Patologia	50x50 GERABTA	TWD, ceW, cenW, TWC	MLP	95%
Paulra et al. (2010)	Normal x Patologia	53x257 MEEI	TMRAP	MLP	92,76%
Scholotthauer, Torres e Rufiner (2010)	Normal x Patologia	53x53 MEEI	DME	KVP	93,4%
Scholotthauer, Torres e Rufiner (2010)	Normal x Patologia	200x200 BDA	DME	KVP	99%
Martinez e Rufiner (2000)	Normal x Patologia	8 TIMITx22 SPAPL	Cep, MCep, FFT	MLP	91,3%
Martinez e Rufiner (2000)	Normal x Rouca x bicíclica	8 TIMITx13x9 SPAPL	Cep, MCep, ΔCep, $\Delta MCep$, FFT	MLP	87,7%
Crovato (2004)	Normal x Patologia	13x42 PUC	PW	MLP	89,06%

Continua na página seguinte

Tabela 3.1 – Continuação

Autores	Classificação	Base de Dados	Características	Classificadores	Taxa de Acerto
Crovato (2004)	Laringite Crônica x Outros	13x42 PUC	PW	MLP	87,5%
Crovato (2004)	Degenerativo x Outros	7x48 PUC	PW	MLP	95,31%
Crovato (2004)	Mobilidade Incorreta x Outros	10x45 PUC	PW	MLP	87,5%
Crovato (2004)	Alterações Orgânicas x Outros	5x50 PUC	PW	MLP	100%
Crovato (2004)	Crescimento Orgânico x Outros	7x48 PUC	PW	MLP	97,87%
Scholotthauer e Torres (2006)	normal x espamódica x tensão muscular	53x21x15 MEEI	GPV + jitter + PMR + QPP5pt + shimmer + QPA3pt + QPA11pt + TR	MLP	93,26%
Fonseca (2008)	Normal x Patologia	30x46 FMRP	TWD	SVM	88,2%

Continua na página seguinte

Tabela 3.1 – Continuação

Autores	Classificação	Base de Dados	Características	Classificadores	Taxa de Acerto
Fonseca (2008)	Normal x nodulo	30x30 FMRP	TWD	SVM	90,1%
Fonseca (2008)	Normal x edema	30x16FMRP	TWD	SVM	85,3%
Fonseca (2008)	edema x nodulo	16x30 FMRP	TWD	SVM	82,4%
Aguiar-Neto, Costa e Fechine (2008)	Normal x Patologia	53x67 MEEI	LPC, Cep, MCep	QV	95%
Aguiar-Neto, Costa e Fechine (2008)	Normal x edema	53x44 MEEI	LPC, Cep, MCep	QV	99%
Aguiar-Neto, Costa e Fechine (2008)	edema x outras Patologias	44x23 MEEI	LPC, Cep, MCep	QV	83%
Costa (2008)	Normal x Patologia	53x67 MEEI	LPC, Cep, CepP, ΔCep, $\Delta CepP$, MCep	QV + HMM	99%

Continua na página seguinte

Tabela 3.1 – Continuação

Autores	Classificação	Base de Dados	Características	Classificadores	Taxa de Acerto
Costa (2008)	Normal x edema	53x44 MEEI	LPC, Cep, CepP, ΔCep, $\Delta CepP$, MCep	QV + HMM	100%
Costa (2008)	edema x outras Patologias	44x23 MEEI	LPC, Cep, CepP, ΔCep, $\Delta CepP$, MCep	QV + HMM	96%
Markaki e Stylianou (2009)	Normal x Patologia	53 x 173 MEEI	EM	SVM	94,07%
Markaki e Stylianou (2009)	polipo x Nodulo	88 MEEI	EM	SVM	87,05%
Markaki e Stylianou (2009)	polipo x Espasmódica	88 MEEI	EM	SVM	82,5%
Markaki e Stylianou (2009)	polipo x Leucoplasia	88 MEEI	EM	SVM	81,8%

Tabela 3.2: Bases de dados utilizadas.

Base de Dados	Sinal	Sigla
Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab	Vogal Sustentada /a/	MEEI
Proprietária	Vogal Sustentada /a/	Prop.
Republican Center of Speech, Voice and Hearing Pathologies	Texto	RCSVHP
Laboratório G.E. da Universidade de Los Angeles e RABTA Hospital de Tunis	Palavra	GERABTA
TIMIT *continuous speech corpus*	Vogal Sustentada /a/	TIMIT
Speech Processing and Auditory Perception Laboratory	Vogal Sustentada /a/	SPAPL
Hospital da Pontifícia Universidade Católica do Rio Grande do Sul	Vogal Sustentada /a/	PUC
Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto	Vogal Sustentada /a/	FMRP
Base de dados Artificial	Vogal Sustentada /a/	BDA

Tabela 3.3: Tipos de Características utilizadas.

Característica	Sigla
Coeficientes Mel-cepstrais	MCep
Coeficientes Delta-Mel-cepstrais	$\Delta MCep$
Coeficientes Delta-Delta-Mel-cepstrais	$\Delta\Delta MCep$
Energia	E
Return Period Density Entropy	RPDE
Detrended fluctuation analysis	DFA
Transformada Wavelet Contínua	TWC
entropia de Shannon	eS

Continua na página seguinte

Tabela 3.3 – Continuação

Característica	Sigla
entropia Relativa	eR
entropia de Tsallis	eT
Transformada Wavelet Discreta	TWD
coeficientes de energia Wavelet	ceW
coeficientes de entropia Wavelet	cenW
Tri Mean Relative average pertubation	TMRAP
Decomposição de modo empírico	DME
Coeficientes Cepstrais	Cep
Coeficientes Delta-Cepstrais	ΔCep
Transformada Rápida de Fourier	FFT
Pacotes Wavelets	PW
Grau de Paradas de Voz	GPV
Jitter	jitter
Perturbação Média Relativa	PMR
quociente de perturbação do periodo de 5 pontos	QPP5pt
shimmer	shimmer
quociente de perturbação da amplitude de 3 pontos	QPA3pt
quociente de perturbação da amplitude de 11 pontos	QPA11pt
Taxa de ruído	TR
coeficientes LPC	LPC
coeficientes Cepstrais ponderados	CepP
coeficientes Delta-cepstrais ponderados	$\Delta CepP$
espectro de modulação	EM

Tabela 3.4: Tipos de Classificadores utilizados.

Classificador	Sigla
Redes Neurais MLP	MLP
Aprendizagem por Quantização Vetorial	LVQ
Modelo de Misturas de Gaussianas	GMM
discriminador linear Gaussiano	DLG
Support Vector Machine	SVM
Quantização Vetorial	QV
K-vizinhos mais próximos	KVP
Modelos de Markov Escondidos	HMM

Capítulo 4

Abordagem Proposta

Neste capítulo, é apresentada a abordagem proposta para a investigação experimental realizada neste trabalho. Na primeira seção, é descrita a base de dados. A metodologia do trabalho é apresentada na seção seguinte. As técnicas utilizadas na investigação experimental realizada foram discutidas no capítulo 2, juntamente com as patologias da laringe presentes na base de dados.

4.1 Base de Dados

A base de dados utilizada foi desenvolvida pelo Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab (KAY-ELEMETRICS, 1994). A base de dados (Disordered Voice Database, Model 4337) é composta por de 1400 amostras de vozes desordenadas de aproximadamente 700 sujeitos, com uma locução da vogal sustentada /ah/ e uma locução dos 12 primeiros segundos da "*Raibow Passage*"[1] para cada sujeito. Esta base de dados inclui amostras de pacientes de variadas desordens vocais, que possuem causas orgânicas, neurológicas, traumáticas, psicogênicas, entre outras,e possui como objetivo servir como auxílio para aplicações clínicas ou de pesquisa. Entre as desordens vocais tem-se edema de Reinke, Nódulos vocais, Cistos e Paralisia discutidos no Capítulo 2. Todas as amostras foram obtidas em um ambiente controlado com baixo nível de ruído, distância constante do microfone, taxa de amostragem de 25 kamostras/s para sinais de voz afetada por patologia e 50 kamostras/s

[1]Texto de domínio público que pode ser encontrado na página 127 de (FAIRBANKS, 1960)

para sinais de voz normal, e 16 bits por amostra [2]. Os nomes dos sinais utilizados da base de dados e algumas características dos locutores, como gênero e faixa-etária são apresentados em quadro no Apêndice 1.

Foram utilizados os seguintes casos da base de dados:

- Vozes afetadas por patologias:
  - Edemas nas dobras vocais: 43 vozes - 32 mulheres, na faixa de 17 a 85 anos e 11 homens, na faixa de 23 a 63 anos, a maioria com edema bilateral (31 casos);
  - Outras patologias nas dobras vocais: 21 casos contendo vozes de pessoas afetadas por cistos, nódulos e paralisia na faixa etária de 18 a 80 anos, sendo 8 homens entre 43 e 75 anos e 13 mulheres entre 18 e 80 anos.
- Vozes normais: 53 casos de vozes normais, sendo 32 mulheres entre 26 e 59 anos e 21 homens entre 22 e 52 anos.

4.2 Metodologia

O sistema proposto nesta dissertação para auxílio ao diagnóstico de patologias da laringe segue uma sequência típica de etapas (Figura 4.1) encontrada em vários outros trabalhos da literatura (MARTINEZ; RUFINER, 2000; GODINO-LLORENTE; GÓMEZ-VILDA; BLANCO-VELASCO, 2006; COSTA, 2008; SALHI; TALBI; CHERIF, 2008), a fim de se obter o diagnóstico final. As etapas são: Pré-processamento do sinal de voz, extração de características, classificação entre voz normal e voz afetada por patologia, e caso o sinal seja classificado como voz afetada por patologia, classificação entre voz afetada por edema e voz afetada por outra patologia.

A etapa do pré-processamento é dividida em 3 subetapas: aquisição, pré-ênfase e janelamento. A subetapa de aquisição foi realizada conforme descrito na seção anterior, em função da base de dados. De acordo com Zwetsch et al. (2006), as patologias do trato vocal afetam de maneira diferente a glote, sendo assim este mais um fator que ajudaria na diferenciação entre patologias. Naquele mesmo trabalho, não se recomenda usar pré-ênfase no problema de discriminação entre patologias, por justamente suavizar a excitação glotal. Por conta disso,

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

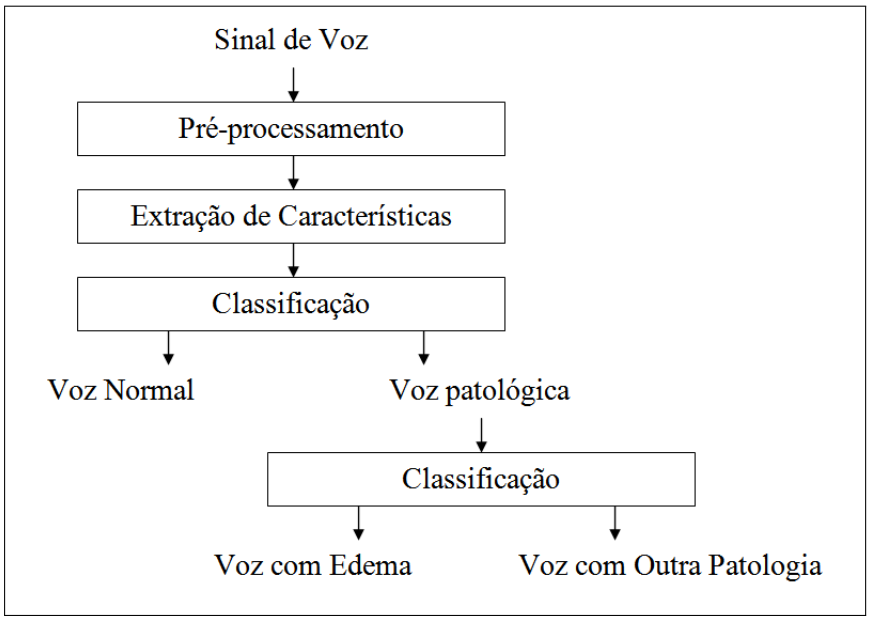

Figura 4.1: Sequência de etapas do sistema de auxílio ao diagnóstico de patologias.

neste trabalho é feita uma análise sem a subetapa de pré-ênfase. Para observar o ganho que a não utilização da pré-ênfase proporciona ao problema de discriminação entre patologias, é também realizada uma análise com a subetapa de pré-ênfase, onde é aplicado um filtro de resposta ao impulso finita de primeira ordem, para a suavização dos efeitos da radiação dos lábios e da variação da área da glote. O fator de pré-ênfase utilizado foi $a_p = 0,95$, por ser um valor típico de uso (FECHINE, 2000). Na subetapa de janelamento, foi utilizada janela de Hamming, por ser uma janela cossenoidal, que suaviza as extremidades e enfatiza o centro, evitando assim transições abruptas, com tamanho de 20 ms e sobreposição de 50% para não haver perda de informação nas extremidades.

Na etapa de extração de características, foram testados diferentes tipos de características extraídas dos sinais de voz em cada um dos classificadores. Os tipos de características escolhidas foram coeficientes LPC, coeficientes Cepstrais, coeficientes Delta-cepstrais e uma combinação de coeficientes LPC e Cepstrais. Para cada tipo de característica, foram extraídos 12 coeficientes de cada janela, e repetido o mesmo processo sem aplicação da pré-ênfase.

Por fim, nas etapas de classificação, foram utilizados diferentes classificadores: Redes Neurais, Quantização Vetorial e Modelos de Misturas de Gaussianas. Todos os classificadores foram testados com todos os tipos de características extraídas, apresentados no parágrafo anterior, com e sem a utilização de pré-ênfase, de forma a obter a combinação caracterís-

tica/classificador mais adequada ao problema.

A metodologia utilizada na especificação dos classificadores está descrita nas próximas subseções.

4.2.1 Redes Neurais

Neste trabalho foi utilizado o modelo de Redes Neurais Perceptron com múltiplas camadas (ou redes Multilayer Perceptron - MLP). Essas redes foram utilizadas com sucesso em outros trabalhos de discriminação de vozes afetadas por patologias (GODINO-LLORENTE; GÓMEZ-VILDA; BLANCO-VELASCO, 2006; SALHI; TALBI; CHERIF, 2008; PAULRA et al., 2010; SCHOLOTTHAUER; TORRES, 2006). Esta dissertação complementa todos aqueles trabalhos ao realizar um estudo comparativo sistemático envolvendo diferentes características e técnicas de classificação, algo não presente nos trabalhos citados, que focaram em uma configuração específica de técnicas. O algoritmo usado para treinamento das Redes Neurais foi o Backpropagation (HECHT-NIELSEN, 1989).

A taxa de aprendizado utilizada em todas as redes foi $\eta = 0,05$, e o critério de parada do treinamento das redes foi 2000 iterações. O sistema de treinamento e classificação utilizando Redes Neurais está descrito na Figura 4.2.

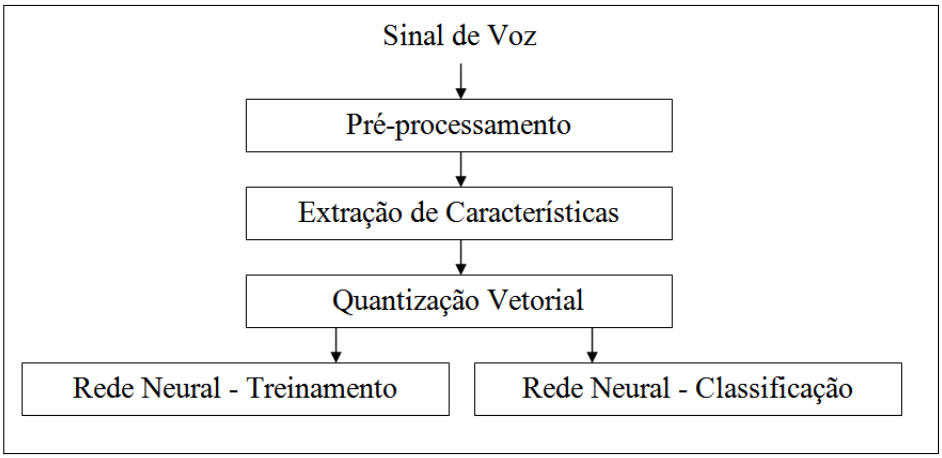

Figura 4.2: Sistema de treinamento e classificação utilizando Redes Neurais MLP.

Nas fases de treinamento, testes e utilização de uma rede neural MLP, requer-se que os padrões de entrada possuam o mesmo tamanho, o que não é o caso dos sinais de voz, que possuem tamanhos diferentes de locutor para locutor, e de locução para locução. Em particu-

lar, sinais de voz normal possuem tamanho maior do que sinais de voz afetada por patologia, por causa da dificuldade dos pacientes com patologias sustentarem a vogal /ah/ por muito tempo. Essa diferença no tamanho dos sinais de voz gera diferentes quantidades de janelas, e, por conseguinte, diferentes tamanhos de vetores de características. Para compressão e normalização dos vetores de características, foi utilizada Quantização Vetorial LBG. O número de níveis do quantizador utilizado foi 64, quantidade que representa bem um sinal de voz sem ter um grande volume de dados (FECHINE; AGUIAR-NETO, 1993).

As Redes Neurais utilizadas neste trabalho possuem 3 camadas, a saber:

- Camada de entrada - possui 768 neurônios (64 níveis do quantizador x 12 coeficientes) ou 1536 neurônios (64 níveis do quantizador x (12 coeficientes LPC + 12 coeficientes Cepstrais));
- Camada escondida ou oculta - possui 2, 4, 6, 8, ..., 18 ou 20 neurônios. Essa variação foi realizada visando testar diferentes tamanhos de Redes Neurais;
- Camada de saída - possui dois neurônios, para diferenciar entre voz normal e voz afetada por patologia, ou voz afetada por edema e voz afetada por outra patologia, em caso de a voz ser classificada previamente como patológica.

Para cada tipo de característica extraída do sinal de voz (coeficientes LPC, coeficientes Cepstrais, coeficientes Delta-cepstrais e a combinação entre coeficientes Cepstrais e LPC), com e sem a utilização da pré-ênfase, e para cada combinação M:N:2, em que M é o número de neurônios na camada de entrada, N o número de neurônios na camada escondida, e 2 o número de neurônios na camada de saída, foram criadas 10 Redes Neurais com inicializações aleatórias, e obtida média das taxas de acerto da fase de teste, a fim de obter a combinação tipo de característica/rede com maior precisão nas duas classificações desejadas.

Para cada rede, foram criados aleatoriamente um conjunto de treinamento e um conjunto de teste. Os conjuntos de treinamento utilizados na discriminação entre voz normal e voz afetada por patologia foram formados escolhendo-se aleatoriamente 27 sinais de voz normal dos 53 disponíveis, 22 sinais com edema de 43 disponíveis e 11 com outras patologias de 21 disponíveis. O restante foi utilizado no conjunto de teste. Os sinais de voz afetada por edema e afetada por outra patologia foram considerados como sendo da mesma classe, a classe

"Patologia", perfazendo assim 27 sinais de voz normal para treinamento e 26 para teste, além de 33 sinais de voz afetada por patologia para treinamento e 31 para teste (Tabela 4.1). Para diferenciar entre voz afetada por edema e voz afetada por outra patologia, foram escolhidos aleatoriamente para cada conjunto de treinamento 11 sinais de voz afetada por edema dos 43 disponíveis e 11 sinais de voz afetada por outra patologia dos 21 disponíveis. O restante dos sinais foi utilizado no conjunto de testes (Tabela 4.2). Os sinais de voz normal não foram utilizados nos experimentos de discriminação entre voz afetada por edema e voz afetada por outra patologia. A pequena quantidade de sinais de voz afetada por edema utilizada no conjunto de treinamento foi para se equiparar à quantidade de sinais com outra patologia, não desbalanceando a rede com padrões de edema em detrimento a outras patologias, para assim ter um maior equilíbrio no aprendizado dos padrões.

Tabela 4.1: Quantidade de sinais de cada classe para treinamento e teste para classificação entre voz normal e voz afetada por patologia utilizando MLP.

Classe	Treinamento	Teste
Normal	27	26
Patologia	33	31

Tabela 4.2: Quantidade de sinais de cada classe para treinamento e teste para classificação entre voz afetada por Edema e voz afetada por outra patologia utilizando MLP.

Classe	Treinamento	Teste
Normal	0	0
Edema	11	32
Outras Patologias	11	10

4.2.2 Quantização Vetorial

Devido à baixa quantidade de sinais de voz afetada por outra patologia, optou-se por utilizar classificadores que realizam classificação de uma classe (*One-class Classification*), a

exemplo da Quantização Vetorial. Esses classificadores são modelados com exemplos de uma única classe, criando um padrão de referência para aquela classe. Dado um padrão de entrada desconhecido, este tipo de classificador retorna como resultado o quão distante esse padrão está do padrão de referência, podendo assim estabelecer limiares, que separam elementos dessa classe dos não pertencentes à classe. Métodos de classificação de uma classe obtêm melhores resultados do que os métodos de classificação multiclasse quando a base de dados é pequena, ou existe uma diferença grande na quantidade de padrões de exemplo entre uma classe e outra, ou em ambas as situações (JOHANNES, 2001), como ocorre na base de dados utilizada neste trabalho, em que a quantidade de sinais de voz afetada por outra patologia é pequena, sendo menos da metade da quantidade de sinais de voz afetada por edema (21 contra 43). A Figura 4.3 mostra o sistema de treinamento e teste utilizando Quantização Vetorial.

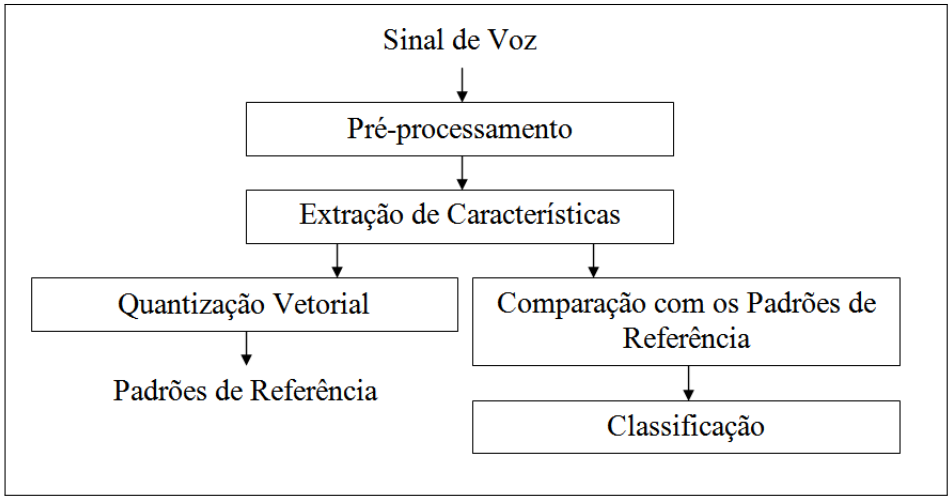

Figura 4.3: Sistema de treinamento e classificação utilizando Quantização Vetorial.

Para o uso da Quantização Vetorial LBG, foram escolhidos 25 sinais de voz afetada por edema entre os 43 para treinamento, e o restante dos sinais (18 sinais com edema, 21 com outra patologia e 53 de voz normal) foi utilizado para teste (Tabela 4.3). Para cada tipo de característica utilizada, coeficientes LPC, coeficientes Cepstrais, coeficientes Delta-cepstrais, e a combinação entre coeficientes LPC e Cepstrais, foi aplicada à Quantização Vetorial com 64 níveis e dimensão 12 (ou 24, para o caso da combinação de coeficientes) no conjunto de treinamento, gerando um dicionário para cada tipo de característica, obtendo-se assim os padrões de referência.

Tabela 4.3: Quantidade de sinais de cada classe para treinamento e teste para classificação utilizando Quantização Vetorial.

	Normal *versus* Patologia		Edema *versus* Outras Patologias	
Classe	Treinamento	Teste	Treinamento	Teste
Normal	0	53	0	0
Edema	25	18	25	18
Outras Patologias	0	21	0	21

Na fase de teste, os padrões de testes são comparados com os padrões de referência utilizando a medida de distância do erro quadrático médio mínimo. Quanto menor é o valor obtido, mais próximo da classe modelada, neste caso a classe edema. Obtendo-se todas as medidas de distorção, é possível traçar um limiar ótimo abaixo do qual considera-se voz afetada por patologia, e acima dele, considera-se voz normal. Para isso, supõe-se que sinais de voz afetada por outra patologia obtiveram medidas de distorção menores que as de sinais de voz normal, por terem características semelhantes aos sinais de voz afetada por edema. Para separar as classes edema e outras patologias, outro limiar é obtido, levando-se em consideração apenas as distorções dos sinais dessas duas classes. Um padrão que apresente uma medida abaixo desse limiar é considerado com sendo da classe edema, e acima do limiar, da classe outras patologias.

4.2.3 Modelo de Misturas de Gaussianas - GMM

GMM foi utilizado como um classificador de uma classe devido aos poucos sinais de voz afetada por outra patologia. Foi criado um GMM apenas para uma única classe, e para definir se um padrão de entrada qualquer pertence a essa classe, é calculada a probabilidade *a posteriori* de ele pertencer ao modelo. Quanto mais alta a probabilidade, maior são as chances de o padrão de entrada pertencer à classe modelada. O sistema de treinamento e classificação utilizando GMM é descrito na Figura 4.4.

Para definir o conjunto de treinamento utilizado no sistema para diferenciar voz normal de voz afetada por patologia, foram utilizadas duas abordagens: a primeira, com o mesmo conjunto de treinamento utilizado na Quantização Vetorial, 25 sinais de edema entre os 43

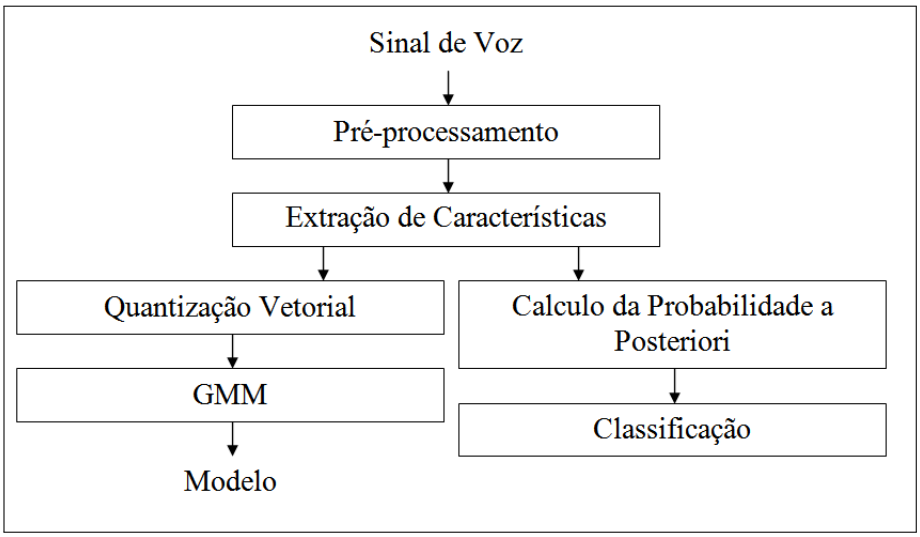

Figura 4.4: Sistema de treinamento e classificação utilizando Modelo de Misturas de Gaussianas.

disponíveis, o restante (18 sinais de edema, os 21 de outras patologias e os 53 normais) como conjunto de teste (Tabela 4.4). A segunda abordagem consistiu em modelar a classe Normal em vez de modelar a classe edema, escolhendo 25 sinais de voz normal para o conjunto de treinamento, sendo o restante para o conjunto de testes (Tabela 4.5). Nos dois casos, os sinais com edema e os sinais com outra patologia foram considerados como sendo da mesma classe "Patologia". Na fase de testes, foram obtidas as probabilidades *a posteriori* de cada sinal de teste, e traçado um limiar ótimo para separar elementos pertencentes a classe modelada ou não.

Na primeira abordagem, supõe-se que os sinais com outras patologias gerarão probabilidades de pertencer à classe edema maiores que os sinais normais, por terem mais características em comum, ficando assim acima do limiar estabelecido. Na segunda abordagem, os sinais com edema e com outras patologias terão probabilidades próximas, ficando abaixo do limiar estabelecido. Para diferenciação entre as classes edema e outra patologia, foram utilizados os 25 sinais de voz afetada por edema para modelar a classe edema, e o restante foi utilizado como conjunto de testes. Os sinais de voz normal foram desconsiderados.

Tabela 4.4: Quantidade de sinais de cada classe para treinamento e teste para classificação utilizando GMM e voz afetada por edema como conjunto de treinamento.

	Normal *versus* Patologia		Edema *versus* Outras Patologias	
Classe	Treinamento	Teste	Treinamento	Teste
Normal	0	53	0	0
Edema	25	18	25	18
Outras Patologias	0	21	0	21

Tabela 4.5: Quantidade de sinais de cada classe para treinamento e teste para classificação utilizando GMM e voz normal como conjunto de treinamento.

	Normal *versus* Patologia	
Classe	Treinamento	Teste
Normal	25	28
Edema	0	43
Outras Patologias	0	21

Foram utilizados coeficientes LPC, coeficientes Cepstrais, coeficientes Delta-cepstrais e a combinação entre coeficientes LPC e Cepstrais como tipos de características de entrada. Para cada tipo de característica, foi criado um modelo de uma classe de 2, 4, 8, 16, 24, 32, 48 ou 64 componentes. Previamente, foi criado um dicionário de C níveis (C é a quantidade de componentes que o modelo terá) e dimensão 12 (ou 24) a partir do conjunto de treinamento utilizando Quantização Vetorial LBG.

4.3 Considerações Finais

Neste capítulo, foi descrita a abordagem proporta para esta dissertação. Inicialmente é realizado o pré-processamento do sinal, onde pode ser realizada ou não a pré-ênfase, e depois o sinal é janelado. Após, são obtidos os vetores de características, que podem ser coeficientes LPC, Cesptrais, Delta-cepstrais, ou uma combinação de coeficientes LPC e Cepstrais. Por fim, é realizada a etapa de treino/classificação, podendo ser utilizado Redes Neurais MLP,

Quantização Vetorial e GMM. Para classificação com as Redes Neurais, o sinal passa por uma etapa intermediária, onde é utilizada a Quantização Vetorial para compressão e normalização dos vetores de características. Para o GMM, o padrão de referência criado pela Quantização Vetorial é utilizado como parâmetro de entrada do modelo.

No próximo capítulo, serão mostrados os resultados obtidos, e uma análise sobre os mesmos.

Capítulo 5

Avaliação Experimental e Análise de Resultados

Neste capítulo, são apresentados e discutidos os resultados obtidos a partir das diferentes abordagens utilizadas nesta dissertação. Ao todo, foram utilizados 4 tipos diferentes de características: coeficientes LPC, coeficientes Cepstrais, coeficientes Delta-cepstrais, e uma combinação entre coeficientes LPC e Cepstrais. Foram utilizados também 3 classificadores: Redes Neurais MLP, Quantização Vetorial e GMM. Cada tipo de característica foi utilizada como entrada de cada um dos classificadores selecionados, sendo obtida a taxa de acerto para cada combinação característica-classificador.

No trabalho de (ZWETSCH et al., 2006), que utiliza análise cepstral para distinção entre sinais de voz afetados por diferentes patologias, não foi utilizada pré-ênfase do sinal, porque altera o sinal de excitação da glote. Por esse motivo, cada abordagem foi investigada com e sem a presença da presença da pré-ênfase na fase de pré-processamento do sinal.

Conforme discutido no capítulo anterior, para o método GMM, foi testada a utilização tanto da classe edema quanto da classe normal para criação do modelo de referência para diferenciação entre voz normal e voz afetada por patologia. Para diferenciação entre voz afetada por edema e voz afetada por outra patologia, foi utilizada apenas a classe edema. Para este tipo de classificador, foram realizadas apenas análises preliminares.

Ao todo, foram investigadas 3 abordagens diferentes para diferenciação entre voz normal e voz afetada por patologia. As 3 abordagens estão descritas na Tabela 5.1, em que a coluna "Característica" indica a característica utilizado, a coluna "Pré-ênfase" indica a presença

ou não de pré-ênfase, a coluna “Classificador” informa o classificador utilizado, e a coluna “Conjunto de Treinamento” indica qual classe foi utilizada para treinamento do classificador.

Tabela 5.1: Descrição das abordagens para diferenciação entre voz normal e voz afetada por patologia.

Abordagem	Característica	Pré-ênfase	Classificador	Conjunto de Treinamento
1	LPC	SIM	MLP	Normal+Patologia
	LPC	NÃO	MLP	Normal+Patologia
	Cepstral	SIM	MLP	Normal+Patologia
	Cepstral	NÃO	MLP	Normal+Patologia
	Delta-cepstral	SIM	MLP	Normal+Patologia
	Delta-cepstral	NÃO	MLP	Normal+Patologia
	LPC+Cepstral	SIM	MLP	Normal+Patologia
	LPC+Cepstral	NÃO	MLP	Normal+Patologia
2	LPC	SIM	QV	Edema
	LPC	NÃO	QV	Edema
	Cepstral	SIM	QV	Edema
	Cepstral	NÃO	QV	Edema
	Delta-cepstral	SIM	QV	Edema
	Delta-cepstral	NÃO	QV	Edema
	LPC+Cepstral	SIM	QV	Edema
	LPC+Cepstral	NÃO	QV	Edema
3	LPC	SIM	GMM	Edema
	LPC	NÃO	GMM	Edema
	Cepstral	SIM	GMM	Edema
	Cepstral	NÃO	GMM	Edema
	Delta-cepstral	SIM	GMM	Edema
	Delta-cepstral	NÃO	GMM	Edema
	LPC+Cepstral	SIM	GMM	Edema
	LPC+Cepstral	NÃO	GMM	Edema

Continua na página seguinte

Tabela 5.1 – Continuação.

Abordagem	Característica	Pré-ênfase	Classificador	Conjunto de Treinamento
	LPC	SIM	GMM	Normal
	LPC	NÃO	GMM	Normal
	Cepstral	SIM	GMM	Normal
	Cepstral	NÃO	GMM	Normal
	Delta-cepstral	SIM	GMM	Normal
	Delta-cepstral	NÃO	GMM	Normal
	LPC+Cepstral	SIM	GMM	Normal
	LPC+Cepstral	NÃO	GMM	Normal

Para diferenciação entre voz afetada por edema e voz afetada por outra patologia, foram investigadas 3 abordagens, descritas na Tabela 5.2.

Tabela 5.2: Descrição das abordagens para diferenciação entre voz afetada por edema e voz afetada por outra patologia.

Abordagem	Característica	Pré-ênfase	Classificador	Conjunto de Treinamento
1	LPC	SIM	MLP	Patologia
	LPC	NÃO	MLP	Patologia
	Cepstral	SIM	MLP	Patologia
	Cepstral	NÃO	MLP	Patologia
	Delta-cepstral	SIM	MLP	Patologia
	Delta-cepstral	NÃO	MLP	Patologia
	LPC+Cepstral	SIM	MLP	Patologia
	LPC+Cepstral	NÃO	MLP	Patologia
2	LPC	SIM	QV	Edema
	LPC	NÃO	QV	Edema
	Cepstral	SIM	QV	Edema
	Cepstral	NÃO	QV	Edema
	Delta-cepstral	SIM	QV	Edema

Continua na página seguinte

Tabela 5.2 – Continuação.

Abordagem	Característica	Pré-ênfase	Classificador	Conjunto de Treinamento
	Delta-cepstral	NÃO	QV	Edema
	LPC+Cepstral	SIM	QV	Edema
	LPC+Cepstral	NÃO	QV	Edema
3	LPC	SIM	GMM	Edema
	LPC	NÃO	GMM	Edema
	Cepstral	SIM	GMM	Edema
	Cepstral	NÃO	GMM	Edema
	Delta-cepstral	SIM	GMM	Edema
	Delta-cepstral	NÃO	GMM	Edema
	LPC+Cepstral	SIM	GMM	Edema
	LPC+Cepstral	NÃO	GMM	Edema

5.1 Classificação entre Voz Normal e Voz Afetada por Patologia

Nas subseções a seguir, são apresentados os resultados de cada uma das abordagens para diferenciar voz normal de voz afetada por patologia. Ao final, tem-se a discussão dos resultados.

5.1.1 Abordagem 1 - Redes Neurais MLP

Conforme discutido no capítulo 4, cada arquitetura de rede neural foi treinada 10 vezes, e foi obtida a média da taxa de acerto de cada uma. Os resultados obtidos ao se utilizar Redes Neurais MLP estão nas tabelas abaixo, onde "N" indica a quantidade de neurônios na camada escondida, "Normal (%)" a média de acerto dos 10 treinamentos para voz normal, "Patologia (%)" a média de acerto dos 10 treinamentos para voz afetada por patologia, "Total (%)" a média de acerto total, e "σ_N" e "σ_P", os desvios padrão para voz normal e voz afetada por patologia, respectivamente.

Tabela 5.3: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes LPC.

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	87,69	96,13	92,28	6,49	4,31
4	88,46	95,16	92,11	6,83	4,86
6	86,92	95,16	91,40	7,02	5,93
8	86,15	94,84	90,88	6,23	5,05
10	86,54	95,81	91,58	5,57	5,52
12	86,92	95,16	91,40	5,44	5,09
14	86,15	96,45	91,75	5,68	5,65
16	86,92	95,81	91,75	6,54	5,09
18	86,54	96,45	91,93	6,40	5,22
20	86,15	96,45	91,75	6,49	5,65

Verifica-se, a partir da análise da Tabela 5.3, que as taxas de classificação globais ficaram num patamar muito próximo (entre 90,88 e 92,28%) ao se variar a quantidade de neurônios na camada escondida. Os resultados com mais neurônios foram piores que os resultados com 2 e 4 neurônios. Houve um maior acerto para os sinais de voz afetada por patologia do que para os sinais de voz normal. A maior taxa de acerto foi alcançada com 2 neurônios na camada escondida, 92,28%.

Tabela 5.4: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes LPC (sem pré-ênfase).

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	83,46	93,87	89,12	8,78	8,74
4	82,31	94,84	89,12	8,85	4,31
6	81,15	94,84	88,60	9,64	5,70
8	81,54	95,16	88,95	7,94	5,09
10	81,54	94,52	88,60	9,73	5,28
Continua na página seguinte					

Tabela 5.4 – Continuação

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
12	82,69	94,19	88,95	8,00	4,67
14	82,69	95,16	89,47	7,35	4,86
16	81,54	94,52	88,60	7,94	5,00
18	81,92	94,84	88,95	7,26	4,82
20	82,31	95,16	89,30	6,98	5,31

Verifica-se, a partir da análise da Tabela 5.4, que as taxas de classificação globais ficaram num patamar muito próximo (entre 88,60 e 89,47%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por patologia do que para os sinais de voz normal. A maior taxa de acerto foi alcançada com 20 neurônios na camada escondida, 89,30%.

Tabela 5.5: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes Cepstrais.

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	92,31	92,58	92,46	4,87	9,52
4	92,31	93,23	92,81	4,51	11,26
6	91,92	93,55	92,81	3,65	9,67
8	93,08	94,19	93,68	3,72	10,36
10	91,54	94,19	92,98	3,53	7,26
12	91,54	94,19	92,98	3,53	9,91
14	92,31	94,84	93,68	3,24	9,00
16	93,08	94,52	93,86	4,51	9,83
18	91,92	93,87	92,98	3,17	10,20
20	90,77	94,19	92,63	4,05	9,18

Verifica-se, a partir da análise da Tabela 5.5, que as taxas de classificação globais ficaram num patamar muito próximo (entre 92,46 e 93,68%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por patologia do

que para os sinais de voz normal. A maior taxa de acerto foi alcançada com 16 neurônios na camada escondida, 93,86%.

Tabela 5.6: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes Cepstrais (sem pré-ênfase).

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	95,38	91,94	93,51	5,38	5,31
4	94,62	93,23	93,86	4,05	7,57
6	95,00	93,55	94,21	4,60	6,17
8	95,00	92,90	93,86	4,60	6,81
10	94,23	95,16	94,74	4,15	4,62
12	93,85	94,84	94,39	4,37	5,28
14	94,23	93,87	94,04	4,23	4,81
16	93,85	94,19	94,04	4,37	5,78
18	94,62	94,84	94,74	4,44	4,82
20	93,85	95,16	94,56	4,37	5,52

Verifica-se, a partir da análise da Tabela 5.4, que as taxas de classificação globais ficaram num patamar muito próximo (entre 93,51 e 94,74%), havendo uma melhora pouco significativa ao se aumentar a quantidade de neurônios na camada escondida. A partir de 10 neurônios as taxas de acerto ficaram acima de 94%, sendo que a maior taxa de acerto foi alcançada com 10 e com 18 neurônios na camada escondida, 94,74%.

Tabela 5.7: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes Delta-cepstrais.

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	96,54	76,77	85,79	6,60	12,07
4	97,31	74,52	84,91	5,75	14,51
6	96,15	77,74	86,14	5,50	11,48
Continua na página seguinte					

Tabela 5.7 – Continuação

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
8	96,92	76,77	85,96	6,33	11,08
10	95,38	76,13	84,91	5,06	12,56
12	93,46	77,74	84,91	5,14	10,43
14	96,15	78,39	86,49	6,07	11,15
16	95,38	78,39	86,14	6,49	12,05
18	95,38	78,06	85,96	5,96	10,80
20	95,38	77,10	85,44	5,68	11,36

Verifica-se, a partir da análise da Tabela 5.7, que as taxas de classificação globais ficaram num patamar muito próximo (entre 84,91 e 86,49%) ao se variar a quantidade de neurônios na camada escondida. As taxas de acerto dos sinais de voz normal foram bastante superiores as taxas de acerto dos sinais de voz afetada por patologia. A maior taxa de acerto foi obtida com 14 neurônios na camada escondida, 86,49%.

Tabela 5.8: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes Delta-cepstrais (sem pré-ênfase).

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	96,54	77,10	85,96	6,60	12,05
4	95,77	76,77	85,44	5,75	11,88
6	96,54	78,06	86,49	5,51	12,58
8	96,54	76,77	85,79	5,80	12,26
10	96,54	77,42	86,14	5,57	12,26
12	95,00	77,42	85,44	5,86	12,02
14	96,15	78,71	86,67	6,59	10,97
16	95,77	79,68	87,02	6,03	11,50
18	95,77	78,39	86,32	6,29	12,15
20	94,23	80,00	86,49	6,89	10,97

Verifica-se, a partir da análise da Tabela 5.8, que as taxas de classificação globais ficaram num patamar muito próximo (entre 85,44 e 87,02%) ao se variar a quantidade de neurônios na camada escondida. As taxas de acerto dos sinais de voz normal foram bastante superiores as taxas de acerto dos sinais de voz afetada por patologia. A maior taxa de acerto foi obtida com 16 neurônios na camada escondida, 87,02%.

Tabela 5.9: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes LPC+Cepstrais.

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	87,31	96,45	92,28	7,35	8,30
4	88,08	96,45	92,63	6,03	7,57
6	89,23	95,48	92,63	6,07	7,17
8	89,62	95,48	92,81	4,81	7,79
10	90,00	96,77	93,68	6,83	7,13
12	88,46	96,13	92,63	5,79	7,76
14	89,23	95,81	92,81	4,51	7,48
16	88,85	96,77	93,16	3,74	7,52
18	89,62	97,42	93,86	4,07	7,67
20	90,00	97,42	94,04	3,53	8,25

Verifica-se, a partir da análise da Tabela 5.9, que as taxas de classificação globais ficaram num patamar muito próximo (entre 92,28 e 94,04%), havendo uma melhora pouco significativa ao se aumentar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por patologia do que para os sinais de voz normal. A maior taxa de acerto foi alcançada com 20 neurônios na camada escondida, 94,04%.

Tabela 5.10: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Redes Neurais MLP e Coeficientes LPC+Cepstrais (sem pré-ênfase).

N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
2	91,54	92,58	92,11	5,06	6,66
4	92,69	92,26	92,46	4,53	7,94
6	90,77	92,90	91,93	6,78	7,91
8	91,54	93,55	92,63	5,68	7,36
10	90,77	93,23	92,11	7,73	7,21
12	91,54	93,23	92,46	6,49	7,26
14	90,77	94,52	92,81	7,25	7,72
16	91,15	92,90	92,11	5,86	7,14
18	90,38	94,19	92,46	8,20	7,52
20	91,54	94,19	92,98	5,38	8,39

Verifica-se, a partir da análise da Tabela 5.10, que as taxas de classificação globais ficaram num patamar muito próximo (entre 91,93 e 92,98%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por patologia do que para os sinais de voz normal. A maior taxa de acerto foi alcançada com 20 neurônios na camada escondida, 92,98%.

5.1.2 Abordagem 2 - Quantização Vetorial

A Figura 5.1 contém uma representação gráfica da distorção obtida para coeficientes Cepstrais, com utilização de pré-ênfase, e o limiar utilizado para diferenciar voz normal de voz afetada por patologia. Os valores de distorção foram normalizados entre 0 e 1000 para facilitar o cálculo do limiar ótimo.

Pode-se observar a partir da figura, que a separação entre voz normal e voz afetada por patologia é bem clara, enquanto que a separação entre voz afetada por edema e voz afetada por outra patologia não é bem clara, sendo as duas classes de difícil separação. Isso mostra que as 2 classes possuem similaridades entre si, sendo essa similaridade devido ao fato de as patologias escolhidas afetarem o trato vocal, especialmente quanto a cistos, nódulos e edemas (COSTA, 2008).

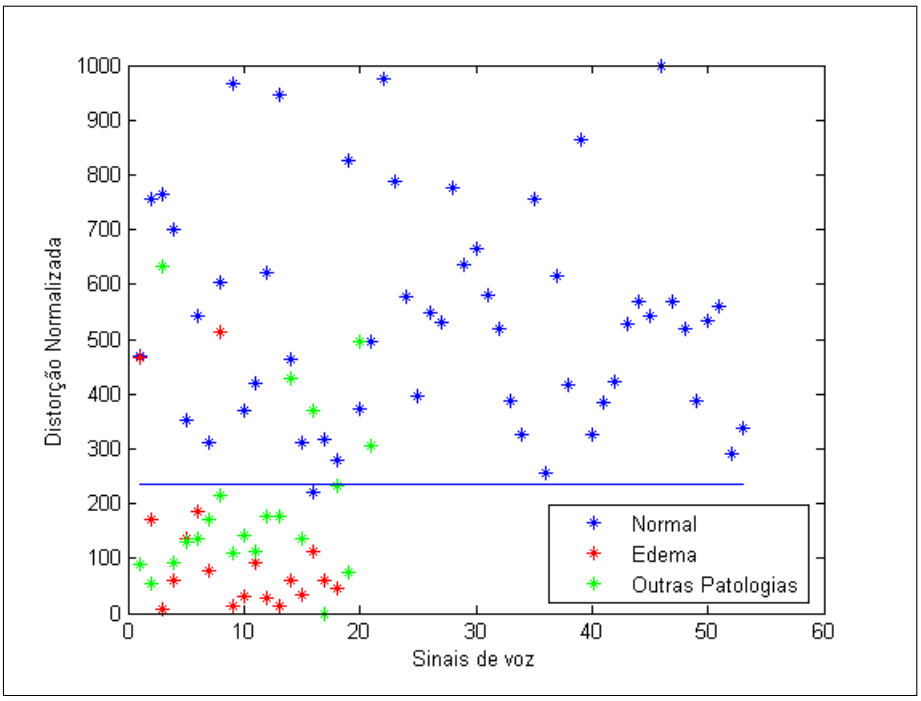

Figura 5.1: Distorção entre os sinais de voz de teste e o dicionário, utilizando coeficientes Cepstrais e pré-ênfase.

A Quantização Vetorial foi aplicada com $k = 64$ níveis. Na Tabela 5.11 é apresentada a taxa de acerto para cada uma das características para classificar entre voz normal e voz afetada por patologia. A coluna "Característica" indica a característica utilizada, a utilização de pré-ênfase é indicada na coluna "Pré-ênfase", as colunas "Normal", "Patologia" e "Total" indicam a quantidade de sinais de voz normal, afetada por patologia, e total, respectivamente, corretamente classificados. As colunas "Normal(%)", "Patologia(%)", e "Total(%)" indicam a porcentagem de acerto para voz normal, voz afetada por patologia, e total, respectivamente.

Tabela 5.11: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando Quantização Vetorial LBG e sinais com Edema como conjunto de treinamento.

Característica	Pré-ênfase	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
LPC	SIM	51	34	85	96,23	87,18	92,39
LPC	NÃO	49	35	84	92,45	89,74	91,30
Cepstral	SIM	52	32	84	98,11	82,05	91,30
Cepstral	NÃO	52	32	84	98,11	82,05	91,30
Delta-cepstral	SIM	51	32	83	96,23	82,05	90,22

Continua na página seguinte

Tabela 5.11 – Continuação

Característica	Pré-ênfase	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
Delta-cepstral	NÃO	51	31	82	96,23	79,49	89,13
LPC+Cepstral	SIM	50	34	84	94,34	87,18	91,30
LPC+Cepstral	NÃO	45	37	82	84,91	94,87	89,13

As taxas de acerto ao diferenciar voz normal de voz afetada por patologia alcançaram bons resultados, acima dos 90% para grande parte das características, sendo as melhores taxas de acerto para coeficientes LPC utilizando a pré-ênfase, com 92,39%. A não utilização da pré-ênfase não representou ganho em nenhuma característica, inclusive diminuiu a taxa de acerto para coeficientes LPC e LPC+Cepstrais.

5.1.3 Abordagem 3 - Modelo de Misturas de Gaussianas

Os resultados ao se utilizar GMM estão nas tabelas abaixo, onde a coluna "Componentes" mostra a quantidade de componentes do modelo, as colunas "Normal", "Patologia" e "Total" a quantidade de sinais corretamente classificados e "Normal(%)", "Patologia(%)" e "Total(%)" as taxas de acerto.

Tabela 5.12: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes LPC e sinais com Edema como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	49	7	56	92,45	17,95	60,87
4	38	15	53	71,70	38,46	57,61
8	48	11	59	90,57	28,21	64,13
16	17	34	51	32,08	87,18	55,43
24	47	13	60	88,68	33,33	65,22
32	26	30	56	49,06	76,92	60,87
48	50	11	61	94,34	28,21	66,30
64	49	9	58	92,45	23,08	63,04

Verifica-se, a partir da análise da Tabela 5.12, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de

componentes não influenciou na taxa de acerto. A maior taxa de acerto, obtida com modelo de 48 componentes, foi 66,30%.

Tabela 5.13: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes LPC (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	52	3	55	98,11	7,69	59,78
4	52	1	53	98,11	2,56	57,61
8	48	6	54	90,57	15,38	58,70
16	37	13	50	69,81	33,33	54,35
24	40	14	54	75,47	35,90	58,70
32	45	10	55	84,91	25,64	59,78
48	42	15	57	79,25	38,46	61,96
64	52	2	54	98,11	5,13	58,70

Verifica-se, a partir da análise da Tabela 5.13, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de componentes não influenciou na taxa de acerto. A maior taxa de acerto, obtida com modelo de 48 componentes, foi 61,96%.

Tabela 5.14: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes Cepstrais e sinais com Edema como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	49	5	54	92,45	12,82	58,70
4	49	5	54	92,45	12,82	58,70
8	24	27	51	45,28	69,23	55,43
16	40	21	61	75,47	53,85	66,30
Continua na página seguinte						

Tabela 5.14 – Continuação

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
24	42	16	58	79,25	41,03	63,04
32	40	18	58	75,47	46,15	63,04
48	13	36	49	24,53	92,31	53,26
64	45	10	55	84,91	25,64	59,78

Verifica-se, a partir da análise da Tabela 5.14, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de componentes não influenciou na taxa de acerto. A maior taxa de acerto, obtida com modelo de 16 componentes, foi 66,30%.

Tabela 5.15: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes Cepstrais (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	40	11	51	75,47	28,21	55,43
4	48	6	54	90,57	15,38	58,70
8	47	13	60	88,68	33,33	65,22
16	45	12	57	84,91	30,77	61,96
24	45	12	57	84,91	30,77	61,96
32	49	9	58	92,45	23,08	63,04
48	24	28	52	45,28	71,79	56,52
64	49	9	58	92,45	23,08	63,04

Verifica-se, a partir da análise da Tabela 5.15, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de componentes não influenciou na taxa de acerto. A maior taxa de acerto, obtida com modelo de 8 componentes, foi 65,22%.

Tabela 5.16: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes Delta-cepstrais e sinais com Edema como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	45	23	68	84,91	58,97	73,91
4	45	23	68	84,91	58,97	73,91
8	48	22	70	90,57	56,41	76,09
16	47	23	70	88,68	58,97	76,09
24	48	23	71	90,57	58,97	77,17
32	40	28	68	75,47	71,79	73,91
48	38	31	69	71,70	79,49	75,00
64	39	31	70	73,58	79,49	76,09

Verifica-se, a partir da análise da Tabela 5.16, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de componentes não influenciou na taxa de acerto. A maior taxa de acerto, obtida com modelo de 24 componentes, foi 77,17%.

Tabela 5.17: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes Delta-cepstrais (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	48	25	73	90,57	64,10	79,35
4	48	26	74	90,57	66,67	80,43
8	48	26	74	90,57	66,67	80,43
16	49	26	75	92,45	66,67	81,52
24	47	28	75	88,68	71,79	81,52
32	47	27	74	88,68	69,23	80,43
48	45	31	76	84,91	79,49	82,61
Continua na página seguinte						

Tabela 5.17 – Continuação

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
64	44	31	75	83,02	79,49	81,52

Verifica-se, a partir da análise da Tabela 5.17, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de componentes aumentou discretamente a taxa de acerto, de 79,35% no modelo com 2 componentes a 82,61%, no modelo com 48 componentes, que proporcionou a maior taxa de acerto.

Tabela 5.18: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes Cepstrais e LPC e sinais com Edema como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	13	34	47	24,53	87,18	51,09
4	51	5	56	96,23	12,82	60,87
8	31	23	54	58,49	58,97	58,70
16	39	20	59	73,58	51,28	64,13
24	38	21	59	71,70	53,85	64,13
32	43	16	59	81,13	41,03	64,13
48	38	19	57	71,70	48,72	61,96
64	38	19	57	71,70	48,72	61,96

Verifica-se, a partir da análise da Tabela 5.18, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de componentes não influenciou na taxa de acerto. A maior taxa de acerto, obtida com os modelos de 16, 24 e 32 componentes, foi 64,13%.

Tabela 5.19: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, Coeficientes Cepstrais e LPC (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	40	11	51	75,47	28,21	55,43
4	48	6	54	90,57	15,38	58,70
8	47	13	60	88,68	33,33	65,22
16	45	12	57	84,91	30,77	61,96
24	45	12	57	84,91	30,77	61,96
32	49	9	58	92,45	23,08	63,04
48	24	28	52	45,28	71,79	56,52
64	49	9	58	92,45	23,08	63,04

Verifica-se, a partir da análise da Tabela 5.19, que as taxas de classificação globais ficaram das apresentadas nas abordagens anteriores e na literatura. O aumento do número de componentes não influenciou na taxa de acerto. A maior taxa de acerto, obtida com o modelo de 8 componentes foi 65,22%.

Tabela 5.20: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes LPC, e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	19	60	79	67,86	93,75	85,87
4	21	61	82	75,00	95,31	89,13
8	25	53	78	89,29	82,81	84,78
16	24	57	81	85,71	89,06	88,04
24	22	57	79	78,57	89,06	85,87
32	22	58	80	78,57	90,63	86,96
48	23	56	79	82,14	87,50	85,87
64	22	55	77	78,57	85,94	83,70

As taxas de acerto alcançaram valores próximos a 90%, sendo que o melhor resultado foi obtido com o modelo de 4 componentes, com 89,13% de acerto. As taxas de acerto de sinais de voz afetada por patologias foram superiores às taxas dos sinais de voz normal. E o crescimento da quantidade de componentes do modelo não influenciou no resultado.

Tabela 5.21: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes LPC (sem Pré-ênfase), e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	25	59	84	89,29	92,19	91,30
4	22	61	83	78,57	95,31	90,22
8	25	56	81	89,29	87,50	88,04
16	26	56	82	92,86	87,50	89,13
24	26	58	84	92,86	90,63	91,30
32	26	58	84	92,86	90,63	91,30
48	26	56	82	92,86	87,50	89,13
64	26	56	82	92,86	87,50	89,13

As taxas de acerto alcançaram valores superiores a 90% em alguns casos, sendo que o melhor resultado foi obtido com os modelos de 2, 24 e 32 componentes, com 91,30% de acerto. O crescimento da quantidade de componentes do modelo não influenciou no resultado.

Tabela 5.22: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes Cepstrais, e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	25	55	80	89,29	85,94	86,96
4	26	54	80	92,86	84,38	86,96
8	24	56	80	85,71	87,50	86,96
16	22	57	79	78,57	89,06	85,87

Continua na página seguinte

Tabela 5.22 – Continuação

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
24	27	52	79	96,43	81,25	85,87
32	28	48	76	100,00	75,00	82,61
48	26	50	76	92,86	78,13	82,61
64	25	56	81	89,29	87,50	88,04

As taxas de acerto alcançaram valores próximos a 90%, sendo que o melhor resultado foi obtido com o modelo de 64 componentes, com 88,04% de acerto. O crescimento da quantidade de componentes do modelo influenciou negativamente no resultado, cujas taxas de acerto diminuíram até o modelo com 48 componentes, aumentando no modelo de 64 componentes.

Tabela 5.23: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes Cepstrais (sem Pré-ênfase), e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	24	62	86	85,71	96,88	93,48
4	24	62	86	85,71	96,88	93,48
8	23	56	79	82,14	87,50	85,87
16	23	57	80	82,14	89,06	86,96
24	22	60	82	78,57	93,75	89,13
32	22	63	85	78,57	98,44	92,39
48	21	61	82	75,00	95,31	89,13
64	21	62	83	75,00	96,88	90,22

As taxas de acerto alcançaram valores superiores a 90% em alguns casos, sendo que o melhor resultado foi obtido com os modelos de 2 e 4 componentes, com 93,48% de acerto. As taxas de acerto de sinais de voz afetada por patologias foram superiores às taxas dos sinais de voz normal. O crescimento da quantidade de componentes do modelo não influenciou no resultado.

Tabela 5.24: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes Delta-cepstrais, e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	24	57	81	85,71	89,06	88,04
4	24	57	81	85,71	89,06	88,04
8	26	54	80	92,86	84,38	86,96
16	26	53	79	92,86	82,81	85,87
24	26	55	81	92,86	85,94	88,04
32	26	54	80	92,86	84,38	86,96
48	26	55	81	92,86	85,94	88,04
64	26	55	81	92,86	85,94	88,04

As taxas de acerto alcançaram valores próximos a 90%, sendo que o melhor resultado foi obtido com os modelos de 2, 4, 24, 48 e 64 componentes, com 88,04% de acerto. O crescimento da quantidade de componentes do modelo não influenciou no resultado.

Tabela 5.25: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes Delta-cepstrais (sem Pré-ênfase), e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	25	54	79	89,29	84,38	85,87
4	26	50	76	92,86	78,13	82,61
8	26	52	78	92,86	81,25	84,78
16	26	50	76	92,86	78,13	82,61
24	27	48	75	96,43	75,00	81,52
32	26	50	76	92,86	78,13	82,61
48	26	51	77	92,86	79,69	83,70
64	26	51	77	92,86	79,69	83,70

As taxas de acerto alcançaram valores um pouco acima de 80%, sendo que o melhor resultado foi obtido com o modelo de 2 componentes, com 85,87% de acerto. As taxas de acerto de sinais de voz normal foram superiores às taxas dos sinais de voz afetada por patologia. O crescimento da quantidade de componentes do modelo não influenciou no resultado.

Tabela 5.26: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes Cepstrais+LPC, e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	21	61	82	75,00	95,31	89,13
4	24	55	79	85,71	85,94	85,87
8	22	58	80	78,57	90,63	86,96
16	23	55	78	82,14	85,94	84,78
24	24	54	78	85,71	84,38	84,78
32	23	55	78	82,14	85,94	84,78
48	24	51	75	85,71	79,69	81,52
64	23	55	78	82,14	85,94	84,78

As taxas de acerto alcançaram valores próximos a 90%, sendo que o melhor resultado foi obtido com o modelo de 2 componentes, com 89,13% de acerto. O crescimento da quantidade de componentes do modelo não influenciou no resultado.

Tabela 5.27: Classificação dos Sinais de Voz entre duas classes (Normal *versus* Patologia), utilizando GMM, coeficientes Cepstrais+LPC (sem Pré-ênfase), e sinais Normais como conjunto de treinamento.

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
2	27	56	83	96,43	87,50	90,22
4	24	60	84	85,71	93,75	91,30
8	27	52	79	96,43	81,25	85,87
Continua na página seguinte						

Tabela 5.27 – Continuação

Componentes	Normal	Patologia	Total	Normal (%)	Patologia (%)	Total (%)
16	24	59	83	85,71	92,19	90,22
24	23	60	83	82,14	93,75	90,22
32	24	58	82	85,71	90,63	89,13
48	25	56	81	89,29	87,50	88,04
64	25	56	81	89,29	87,50	88,04

As taxas de acerto alcançaram valores um pouco superiores a 90%. O melhor resultado foi obtido com o modelo de 4 componentes, com 91,30% de acerto. O crescimento da quantidade de componentes do modelo não influenciou no resultado.

5.1.4 Análise dos Resultados

As taxas de acerto obtidas ao se utilizar a classe edema como conjunto de treinamento do GMM (tabelas 5.12 a 5.19), assim como é usada na Quantização Vetorial, são extremamente baixas quando comparadas às taxas obtidas com Redes Neurais e Quantização Vetorial. Uma possível causa pode ser o baixo grau de correlação entre os sinais de voz afetada por edema utilizados, fazendo com que o modelo não seja representativo para a classe. Na Tabela 5.28 é apresentada uma matriz de covariância do modelo de 8 componentes utilizando coeficientes Cepstrais e pré-ênfase.

Tabela 5.28: Matriz de covariância de um dos componentes de um modelo de 8 componentes treinado com coeficientes Cepstrais, utilizando pré-ênfase, extraídos de sinais de voz afetada por Edema.

0,0040	0,0003	-0,0018	0,0006	0,0003	0,0019	0,0011	-0,0019	0,0010	-0,0004	-0,0002	0,0008
0,0003	0,0022	0,0007	0,0012	-0,0012	-0,0009	0,0011	7,11E-05	-0,0004	-0,0001	-5,33E-05	0,0001
-0,0018	0,0007	0,0055	0,0012	-0,0010	-0,0042	0,0003	0,0020	-0,0016	8,43E-05	0,0005	-0,0004
0,0006	0,0013	0,0013	0,0027	-0,0009	-0,0010	0,0013	-0,0003	2,60E-05	-0,0002	-0,0003	0,0002
0,0003	-0,0012	-0,0010	-0,0009	0,0025	0,0014	-0,0011	-0,0003	0,0009	-0,0002	-0,0001	-4,61E-05
0,0019	-0,0009	-0,0042	-0,0010	0,0014	0,0045	-0,0002	-0,0021	0,0018	-0,0003	-0,0005	0,0004
0,0011	0,0011	0,0003	0,0013	-0,0011	-0,0002	0,0017	-0,0004	-0,0001	-4,25E-05	-3,59E-05	0,0004
-0,0019	7,11E-05	0,0020	-0,0003	-0,0003	-0,0021	-0,0004	0,0019	-0,0011	0,0002	0,0005	-0,0004
0,0010	-0,0004	-0,0016	2,60E-05	0,0009	0,0018	-0,0001	-0,0011	0,0014	-0,0003	-0,0005	0,0003
-0,0004	-0,0001	8,43E-05	-0,0002	-0,0002	-0,0003	-4,25E-05	0,0002	-0,0003	0,0006	1,93E-05	-0,0003
-0,0002	-5,33E-05	0,0005	-0,0003	-0,0001	-0,0005	-3,59E-05	0,0005	-0,0005	1,93E-05	0,0007	-0,0001
0,0008	0,0002	-0,0004	0,0002	-4,61E-05	0,0004	0,0004	-0,0004	0,0003	-0,0003	-0,0001	0,0005

Pode-se observar na Tabela 5.28 que os valores de covariância do modelo são baixos, proximos de zero, mostrando pouca correlação entre os elementos da classe. Devido a esse problema, foram realizados experimentos utilizando sinais de voz normal como conjunto de treinamento para o GMM (tabelas 5.20 a 5.27), que são sinais com maior grau de correlação. As taxas de acerto obtidas foram aceitáveis, em muitos casos acima dos 90%, comparáveis com as obtidas nas outras abordagens. Deste ponto em diante, dentre as abordagens que usam GMM, o texto irá tratar apenas das que usam sinais de voz normal como conjunto de treinamento.

Ao se utilizar pré-ênfase, a combinação dos coeficientes LPC com os coeficientes Cepstrais proporcionaram os melhores resultados nas Redes Neurais (Tabela 5.9), com uma taxa de acerto de 94,04%, quando utilizados 20 neurônios na camada escondida, observando-se assim um ganho em relação ao uso de coeficientes LPC e Cepstrais isoladamente (tabelas 5.3 e 5.5), que obtiveram taxas de acerto de 92,28% e 93,86%, respectivamente. Na abordagem que usa Quantização Vetorial (Tabela 5.11), os coeficientes LPC isoladamente se mostraram as melhores características, com taxa de acerto de 92,39%. No GMM, o uso de coeficientes LPC, tanto isoladamente (Tabela 5.20) quanto combinado com coeficientes Cepstrais (Tabela 5.26), proporcionaram as melhores taxas de acerto, 89,13% com 4 e 2 componentes, respectivamente.

Os piores resultados para os três classificadores foram obtidos com a utilização de coeficientes Delta-cepstrais (tabelas 5.7, 5.11 e 5.24). Este resultado pode ser explicado pelo fato de os coeficientes Delta-cepstrais utilizarem informações de transição de voz (FECHINE, 2000), enquanto que a vogal sustentada /a/ utilizada neste trabalho possui pouca variação ao longo do sinal.

Nos experimentos que não utilizaram pré-ênfase, os melhores resultados nos três classificadores foram obtidos com o uso dos coeficientes Cepstrais, obtendo-se uma taxa de acerto de 94,74% para as Redes Neurais (Tabela 5.6) para 10 e 18 neurônios na camada escondida, 91,30% para Quantização Vetorial (Tabela 5.11) e para o GMM (Tabela 5.23), 93,48% para modelos com 2 e 4 componentes. Houve um aumento da taxa de acerto para coeficientes Cepstrais em Redes Neurais e GMM (tabelas 5.6 e 5.23) em relação as que utilizaram pré-ênfase (tabelas 5.5 e 5.22). Para Quantização Vetorial (Tabela 5.11), a taxa de acerto permaneceu a mesma ao se retirar a pré-ênfase. Pode-se observar também que as taxas de

acerto da combinação de coeficientes LPC e Cepstrais (tabelas 5.10, 5.11 e 5.27) seguiram as taxas de acerto dos experimentos que utilizaram apenas coeficientes LPC (tabelas 5.4, 5.11 e 5.21), diminuíndo as taxas de acerto nas Redes Neurais e Quantização Vetorial comparando-se com a utilização da pré-ênfase, e aumentando as taxas de acerto para o GMM quando comparado com a utilização da pré-ênfase. As observações acima evidenciam que os coeficientes LPC tem uma maior influência do que os coeficientes Cepstrais quando os dois são utilizados juntos no mesmo vetor de características. Os piores resultados para cada classificador novamente foram obtidos ao se usar coeficientes Delta-cesptrais (tabelas 5.8, 5.11 e 5.25).

Na Tabela 5.29, tem-se as melhores médias de taxas de acerto obtidas para Redes Neurais. A coluna "Característica" indica a característica utilizada, "Pré-ênfase" indica a utilização ou não de pré-ênfase, "N" é a quantidade de neurônios na camada escondida, "Normal (%)", "Patologia (%)" e "Total (%)" mostram as taxas de acerto para voz normal, voz afetada por patologia e total, respectivamente, e os desvios padrão para voz normal e afetada por patologia estão nas colunas "σ_N" e "σ_P", respectivamente.

Tabela 5.29: Melhores taxas de acerto entre duas classes (Normal *versus* Patologia) para Redes Neurais MLP.

Característica	Pré-ênfase	N	Normal (%)	Patologia (%)	Total (%)	σ_N	σ_P
LPC	SIM	2	87,69	96,13	92,28	6,49	4,31
LPC	NÃO	14	82,69	95,16	89,47	7,35	4,86
Cepstral	SIM	16	93,08	94,52	93,86	4,51	9,83
Cepstral	NÃO	10	94,23	95,16	94,74	4,15	4,62
Delta-cepstral	SIM	14	96,15	78,39	86,49	6,07	11,15
Delta-cepstral	NÃO	16	95,77	79,68	87,02	6,03	11,50
LPC+Cepstral	SIM	20	90,00	97,42	94,04	3,53	8,25
LPC+Cepstral	NÃO	20	91,54	94,19	92,98	5,38	8,39

O melhor resultado foi obtido com a utilização de coeficientes Cepstrais, sem pré-ênfase, alcançando uma média de acerto de 94,74% para 10 neurônios na camada escondida. Apesar de a diferença ser pequena, pode-se notar que houve um ganho nas taxas de acerto ao se

aumentar a quantidade de neurônios na camada escondida, já que a maior parte dos melhores resultados foram obtidos com números elevados de neurônios.

Na Tabela 5.30, são apresentadas as melhores médias de taxas de acerto obtidas para GMM. A coluna "Característica" indica a característica utilizada, "Pré-ênfase" indica a utilização ou não de pré-ênfase, "Componentes" é a quantidade de componentes do modelo, "Normal (%)", "Patologia (%)" e "Total (%)" mostram as taxas de acerto para voz normal, voz afetada por patologia e total, respectivamente.

Tabela 5.30: Melhores taxas de acerto entre duas classes (Normal *versus* Patologia) para GMM.

Característica	Pré-ênfase	Componentes	Normal (%)	Patologia (%)	Total (%)
LPC	SIM	4	75,00	95,31	89,13
LPC	NÃO	2	89,29	92,19	91,30
Cepstral	SIM	64	89,29	87,50	88,04
Cepstral	NÃO	2	85,71	96,88	93,48
Delta-cepstral	SIM	2	85,71	89,06	88,04
Delta-cepstral	NÃO	2	89,29	84,38	85,87
LPC+Cepstral	SIM	2	75,00	95,31	89,13
LPC+Cepstral	NÃO	4	85,71	93,75	91,30

O melhor resultado foi obtido utilizando-se coeficientes Cepstrais sem pré-ênfase, assim como nas Redes Neurais, com 93,48%. Nota-se que grande parte das características obtiveram melhores resultados quando são utilizados apenas 2 componentes no modelo, mostrando que poucos componentes podem representar bem a classe de voz normal. Para algumas características, a taxa de acerto para uma quantidade maior de componentes foi igual àquela utilizando 2 componentes. Neste caso, o critério utilizado na escolha foi o modelo com o menor número de componentes, por demandar menos esforço computacional no treinamento. Algumas características apresentaram resultados abaixo daqueles obtidos com a Quantização Vetorial (Tabela 5.11). Nesses casos, é indicada a utilização apenas de Quantização Vetorial para evitar esforço computacional desnecessário.

5.2 Classificação entre Voz Afetada por Edema e Voz Afetada por Outra Patologia

Nas próximas subseções, são apresentados os resultados de cada uma das abordagens para diferenciar voz afetada por edema de voz afetada por outra patologia. Ao final, tem-se a discussão dos resultados.

5.2.1 Abordagem 1 - Redes Neurais MLP

Os resultados do uso de Redes Neurais MLP para diferenciar entre voz afetada por edema e voz afetada por outra patologia estão nas tabelas abaixo, onde "N" indica a quantidade de neurônios na camada escondida, "Edema(%)" a média de acerto para voz afetada por edema, "Outra Patologia (%)" a média de acerto para voz afetada por outra patologia, "Total (%)" a média de acerto total, e "σ_E e "σ_{OP}", os desvios padrão para voz afetada por edema e voz afetada por outra patologia, respectivamente.

Tabela 5.31: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes LPC.

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	55,63	60,00	56,67	12,66	14,14
4	57,81	62,00	58,81	13,20	13,17
6	56,88	65,00	58,81	12,99	16,50
8	59,69	65,00	60,95	12,80	17,16
10	58,13	64,00	59,52	13,91	17,76
12	58,44	62,00	59,29	13,34	15,49
14	57,81	61,00	58,57	12,52	15,95
16	58,44	61,00	59,05	12,59	15,95
18	58,75	62,00	59,52	13,24	15,49
20	58,44	63,00	59,52	12,15	16,36

Verifica-se, a partir da análise da Tabela 5.31, que as taxas de classificação globais ficaram num patamar muito próximo (entre 56,67 e 60,95%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por outra patologia do que para os sinais de voz afetada por edema. As taxas de acerto ficaram muito baixas, próximos a 60%. A maior taxa de acerto, alcançada com 8 neurônios na camada escondida, foi 60,95%.

Tabela 5.32: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes LPC (sem pré-ênfase).

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	53,44	61,00	55,24	12,01	22,83
4	57,19	69,00	60,00	13,51	15,24
6	56,88	66,00	59,05	13,57	14,30
8	56,56	66,00	58,81	13,30	14,30
10	56,88	64,00	58,57	12,91	13,50
12	57,81	64,00	59,29	11,34	14,30
14	59,06	65,00	60,48	11,64	17,80
16	57,81	62,00	58,81	11,24	14,76
18	58,13	62,00	59,05	11,71	14,76
20	59,06	64,00	60,24	11,83	15,78

Verifica-se, a partir da análise da Tabela 5.32, que as taxas de classificação globais ficaram num patamar muito próximo (entre 55,24 e 60,48%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por outra patologia do que para os sinais de voz afetada por edema. As taxas de acerto ficaram muito baixas, próximos a 60%. A maior taxa de acerto, alcançada com 14 neurônios na camada escondida, foi 60,48%.

Tabela 5.33: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes Cepstrais.

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	46,88	52,00	48,10	18,34	26,16
4	47,19	53,00	48,57	17,02	22,63
6	47,50	53,00	48,81	18,45	25,84
8	46,56	52,00	47,86	17,89	25,73
10	46,25	53,00	47,86	17,97	21,63
12	47,50	52,00	48,57	17,48	21,50
14	46,56	51,00	47,62	17,71	25,58
16	47,81	52,00	48,81	17,37	24,40
18	45,94	53,00	47,62	17,80	24,52
20	47,19	53,00	48,57	18,19	22,63

Verifica-se, a partir da análise da Tabela 5.33, que as taxas de classificação globais ficaram num patamar muito próximo (entre 47,62 e 48,81%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por outra patologia do que para os sinais de voz afetada por edema. As taxas de acerto foram extremamente baixas, abaixo de 50%. A maior taxa de acerto, alcançada com 6 e 16 neurônios na camada escondida, foi 48,81%.

Tabela 5.34: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes Cepstrais (sem pré-ênfase).

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	54,38	58,00	55,24	11,14	16,19
4	53,75	55,00	54,05	9,86	20,14
6	53,75	56,00	54,29	8,44	16,47
8	53,44	54,00	53,57	8,77	16,47
10	53,44	56,00	54,05	10,46	14,30
Continua na página seguinte					

Tabela 5.34 – Continuação

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
12	54,06	58,00	55,00	10,63	15,49
14	54,06	56,00	54,52	10,93	15,78
16	54,69	57,00	55,24	10,85	14,94
18	54,06	55,00	54,29	11,51	16,50
20	54,38	55,00	54,52	9,68	14,34

Verifica-se, a partir da análise da Tabela 5.34, que as taxas de classificação globais ficaram num patamar muito próximo (entre 53,57 e 55,24%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por outra patologia do que para os sinais de voz afetada por edema. As taxas de acerto ficaram muito baixas, próximas a 55%. A maior taxa de acerto, alcançada com 2 e com 16 neurônios na camada escondida, foi 55,24%.

Tabela 5.35: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes Delta-cepstrais.

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	63,44	37,00	57,14	7,66	22,14
4	67,81	36,00	60,24	7,80	20,11
6	65,63	39,00	59,29	9,08	20,79
8	65,63	35,00	58,33	10,62	15,09
10	65,63	37,00	58,81	11,51	17,03
12	65,31	41,00	59,52	9,82	18,53
14	67,50	40,00	60,95	9,68	17,64
16	67,19	36,00	59,76	9,58	12,65
18	65,31	39,00	59,05	11,07	15,95
20	64,06	38,00	57,86	11,15	18,74

Verifica-se, a partir da análise da Tabela 5.35, que as taxas de classificação globais ficaram num patamar muito próximo (entre 57,14 e 60,95%) ao se variar a quantidade de

neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por edema do que para os sinais de voz afetada por outra patologia. As taxas de acerto ficaram muito baixas, em torno dos 60%. A maior taxa de acerto, alcançada com 14 neurônios na camada escondida, foi 60,95%.

Tabela 5.36: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes Delta-cepstrais (sem pré-ênfase).

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	68,13	48,00	63,33	9,06	18,74
4	63,13	55,00	61,19	12,83	18,41
6	65,00	47,00	60,71	12,66	18,29
8	66,88	49,00	62,62	14,15	19,69
10	63,75	47,00	59,76	11,33	20,03
12	64,69	48,00	60,71	11,51	20,98
14	65,00	49,00	61,19	12,13	23,78
16	65,31	49,00	61,43	11,07	20,79
18	66,25	45,00	61,19	12,40	21,21
20	63,44	49,00	60,00	10,00	19,69

Verifica-se, a partir da análise da Tabela 5.36, que as taxas de classificação globais ficaram num patamar muito próximo (entre 59,76 e 63,33%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por edema do que para os sinais de voz afetada por outra patologia. As taxas de acerto ficaram muito baixas, por volta dos 60%. A maior taxa de acerto, alcançada com 2 neurônios na camada escondida, foi 63,33%.

Tabela 5.37: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes LPC+Cepstrais.

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	53,44	55,00	53,81	14,16	22,24
4	55,00	52,00	54,29	14,15	22,01
6	55,63	53,00	55,00	14,11	20,03
8	55,31	53,00	54,76	14,44	21,11
10	55,31	46,00	53,10	14,13	20,66
12	55,00	47,00	53,10	14,60	18,89
14	55,31	51,00	54,29	14,29	21,83
16	55,00	50,00	53,81	14,22	20,55
18	55,31	50,00	54,05	14,13	20,00
20	55,00	51,00	54,05	14,52	19,69

Verifica-se, a partir da análise da Tabela 5.37, que as taxas de classificação globais ficaram num patamar muito próximo (entre 53,10 e 55,00%) ao se variar a quantidade de neurônios na camada escondida. Houve um maior acerto para os sinais de voz afetada por edema do que para os sinais de voz afetada por outra patologia. As taxas de acerto ficaram muito baixas, entre 53% e 55%. A maior taxa de acerto, alcançada com 6 neurônios na camada escondida, foi 55,00%.

Tabela 5.38: Classificação dos Sinais de Voz entre duas classes (Edema *versus* Outra Patologia), utilizando Redes Neurais MLP e Coeficientes Cepstrais+LPC (sem pré-ênfase).

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
2	59,06	64,00	60,24	10,97	15,06
4	59,69	62,00	60,24	10,57	15,49
6	61,25	59,00	60,71	9,79	15,24
8	58,44	59,00	58,57	12,76	15,24
10	59,06	59,00	59,05	11,55	14,49
Continua na página seguinte					

Tabela 5.38 – Continuação

N	Edema (%)	Outra Patologia (%)	Total (%)	σ_E	σ_{OP}
12	59,06	58,00	58,81	12,01	16,19
14	59,38	59,00	59,29	12,50	15,24
16	59,06	57,00	58,57	11,17	15,67
18	58,75	60,00	59,05	12,57	14,91
20	59,06	60,00	59,29	12,19	14,91

Verifica-se, a partir da análise da Tabela 5.38, que as taxas de classificação globais ficaram num patamar muito próximo (entre 58,57 e 60,71%) ao se variar a quantidade de neurônios na camada escondida.. Na maioria dos casos, houve um maior acerto para os sinais de voz afetada por outra patologia do que para os sinais de voz afetada por edema. As taxas de acerto ficaram muito baixas, entre 58% e 61%. A maior taxa de acerto, alcançada com 6 neurônios na camada escondida, foi 60,71%.

5.2.2 Abordagem 2 - Quantização Vetorial

O quantizador foi utilizado com $k = 64$ níveis. Na Tabela 5.39 é apresentada a taxa de acerto para cada uma das características para classificar voz afetada por edema e voz afetada por outra patologia. A coluna "Característica" indica a característica utilizada, a utilização de pré-ênfase é indicada na coluna "Pré-ênfase", as colunas "Edema", "O. P." e "Total" indicam a quantidade de sinais de voz afetada por edema, voz afetada por outra patologia, e total, respectivamente, corretamente classificados. As colunas "Edema(%)", "O. P.(%)", e "Total(%)" indicam a porcentagem de acerto para voz afetada por edema, voz afetada por patologia, e total, respectivamente.

Tabela 5.39: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando Quantização Vetorial LBG e sinais com Edema como conjunto de treinamento.

Característica	Pré-ênfase	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
LPC	SIM	11	15	26	61,11	71,43	66,67
LPC	NÃO	10	16	26	55,56	76,19	66,67
Cepstral	SIM	12	17	29	66,67	80,95	74,36
Cepstral	NÃO	16	16	32	88,89	76,19	82,05
Delta-cepstral	SIM	9	21	30	50,00	100,00	76,92
Delta-cepstral	NÃO	14	14	28	77,78	66,67	71,79
Cepstral+LPC	SIM	11	17	28	61,11	80,95	71,79
Cepstral+LPC	NÃO	12	16	28	66,67	76,19	71,79

As melhores taxas de acerto foram obtidas para coeficientes Cepstrais sem utilizar pré-ênfase, 82,05%, corroborando com o trabalho de Zwetsch et al. (2006), o qual se afirma que a não utilização de pré-ênfase para coeficientes Cepstrais facilita a diferenciação entre patologias.

5.2.3 Abordagem 3 - Modelos de Misturas de Gaussianas

Os resultados para o uso do GMM estão nas tabelas abaixo, onde a coluna "Componentes" mostra a quantidade de componentes do modelo, as colunas "Edema", "O. P." e "Total" a quantidade de sinais corretamente classificados, e "Edema (%)", "O. P. (%)" e "Total(%)" as taxas de acerto.

Tabela 5.40: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes LPC e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	14	13	27	77,78	61,90	69,23
Continua na página seguinte						

Tabela 5.40 – Continuação

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
4	11	17	28	61,11	80,95	71,79
8	11	16	27	61,11	76,19	69,23
16	16	11	27	88,89	52,38	69,23
24	14	15	29	77,78	71,43	74,36
32	15	14	29	83,33	66,67	74,36
48	14	15	29	77,78	71,43	74,36
64	15	16	31	83,33	76,19	79,49

As taxas de acerto melhoraram conforme aumentou-se a quantidade de componentes do modelo, variando de 69,23% no modelo com 2 componentes para 79,43% no modelo com 64 componentes. Na maior parte dos casos, as taxas de acerto para voz afetada por edema foram superiores às taxas de acerto para voz afetada por outra patologia.

Tabela 5.41: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes LPC (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	15	11	26	83,33	52,38	66,67
4	13	16	29	72,22	76,19	74,36
8	11	18	29	61,11	85,71	74,36
16	10	18	28	55,56	85,71	71,79
24	16	11	27	88,89	52,38	69,23
32	9	19	28	50,00	90,48	71,79
48	10	17	27	55,56	80,95	69,23
64	15	12	27	83,33	57,14	69,23

O crescimento da quantidade de componentes do modelo não influenciou nas taxas de acerto. O melhor resultado obtido foi 74,26% para os modelos de 4 e 8 componentes. Na maior parte dos casos, as taxas de acerto para voz afetada por outra patologia foram superi-

ores às taxas de acerto para voz afetada por edema.

Tabela 5.42: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes Cepstrais e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	12	13	25	66,67	61,90	64,10
4	13	15	28	72,22	71,43	71,79
8	11	14	25	61,11	66,67	64,10
16	9	16	25	50,00	76,19	64,10
24	12	12	24	66,67	57,14	61,54
32	15	11	26	83,33	52,38	66,67
48	10	14	24	55,56	66,67	61,54
64	13	15	28	72,22	71,43	71,79

O crescimento da quantidade de componentes do modelo não influenciou nas taxas de acerto. O melhor resultado obtido foi 71,79% para os modelos de 4 e 64 componentes. Na maior parte dos casos, as taxas de acerto para voz afetada por edema foram superiores às taxas de acerto para voz afetada por outra patologia.

Tabela 5.43: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes Cepstrais (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	10	17	27	55,56	80,95	69,23
4	11	19	30	61,11	90,48	76,92
8	11	14	25	61,11	66,67	64,10
16	16	9	25	88,89	42,86	64,10
24	16	9	25	88,89	42,86	64,10
Continua na página seguinte						

Tabela 5.43 – Continuação

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
32	11	15	26	61,11	71,43	66,67
48	13	15	28	72,22	71,43	71,79
64	12	16	28	66,67	76,19	71,79

O crescimento da quantidade de componentes do modelo não influenciou nas taxas de acerto. O melhor resultado obtido foi 76,92% para o modelo de 4 componentes. Na maior parte dos casos, as taxas de acerto para voz afetada por outra patologia foram superiores às taxas de acerto para voz afetada por edema.

Tabela 5.44: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes Delta-cepstrais e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	10	20	30	55,56	95,24	76,92
4	10	20	30	55,56	95,24	76,92
8	10	19	29	55,56	90,48	74,36
16	10	18	28	55,56	85,71	71,79
24	10	20	30	55,56	95,24	76,92
32	10	19	29	55,56	90,48	74,36
48	10	18	28	55,56	85,71	71,79
64	8	21	29	44,44	100,00	74,36

O crescimento da quantidade de componentes do modelo não influenciou nas taxas de acerto. O melhor resultado obtido foi 76,92% para os modelos de 2, 4 e 24 componentes. As taxas de acerto para voz afetada por outra patologia foram superiores às taxas de acerto para voz afetada por edema.

Tabela 5.45: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes Delta-cepstrais (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	12	16	28	66,67	76,19	71,79
4	12	17	29	66,67	80,95	74,36
8	12	16	28	66,67	76,19	71,79
16	11	18	29	61,11	85,71	74,36
24	11	19	30	61,11	90,48	76,92
32	11	18	29	61,11	85,71	74,36
48	11	18	29	61,11	85,71	74,36
64	11	18	29	61,11	85,71	74,36

O crescimento da quantidade de componentes do modelo não influenciou nas taxas de acerto. O melhor resultado obtido foi 76,92% para o modelo de 24 componentes. As taxas de acerto para voz afetada por outra patologia foram superiores às taxas de acerto para voz afetada por edema.

Tabela 5.46: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes Cepstrais e LPC e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	12	14	26	66,67	66,67	66,67
4	13	16	29	72,22	76,19	74,36
8	10	18	28	55,56	85,71	71,79
16	13	16	29	72,22	76,19	74,36
24	12	16	28	66,67	76,19	71,79
32	10	17	27	55,56	80,95	69,23
48	13	14	27	72,22	66,67	69,23
Continua na página seguinte						

Tabela 5.46 – Continuação

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
64	16	12	28	88,89	57,14	71,79

O crescimento da quantidade de componentes do modelo não influenciou nas taxas de acerto. O melhor resultado obtido foi 74,36% para os modelos de 4 e 16 componentes. Na maior parte dos casos, as taxas de acerto para voz afetada por outra patologia foram superiores às taxas de acerto para voz afetada por edema.

Tabela 5.47: Classificação dos Sinais de Voz entre duas classes (Edema x Outras Patologias), utilizando GMM, Coeficientes Cepstrais e LPC (sem Pré-ênfase) e sinais com Edema como conjunto de treinamento.

Componentes	Edema	O. P.	Total	Edema (%)	O. P. (%)	Total (%)
2	9	19	28	50,00	90,48	71,79
4	16	14	30	88,89	66,67	76,92
8	15	12	27	83,33	57,14	69,23
16	14	15	29	77,78	71,43	74,36
24	14	14	28	77,78	66,67	71,79
32	14	14	28	77,78	66,67	71,79
48	12	15	27	66,67	71,43	69,23
64	10	18	28	55,56	85,71	71,79

O crescimento da quantidade de componentes do modelo não influenciou nas taxas de acerto. O melhor resultado obtido foi 76,92% para o modelo de 4 componentes. Na maior parte dos casos, as taxas de acerto para voz afetada por edema foram superiores às taxas de acerto para voz afetada por outra patologia.

5.2.4 Análise dos Resultados

Ao se utilizar pré-ênfase, os coeficientes LPC proporcionaram os melhores resultados nas Redes Neurais (Tabela 5.31), com 60,95% para 8 neurônios na camada escondida, e no GMM (Tabela 5.40), com 79,49% para o modelo com 64 componentes, e proporcionaram os piores

na Quantização Vetorial (Tabela 5.39), com 66,67%. A característica que obteve o melhor resultado para Quantização Vetorial foi a combinação de coeficientes LPC e Cepstrais, com 76,92% de acerto. Para a utilização de Redes Neurais (tabelas 5.31 a 5.38) e GMM (tabelas 5.40 a 5.47), os piores resultados foram obtidos ao se usar coeficientes Cepstrais (tabelas 5.33 e 5.42), com 48,81% e 71,79% de acerto, respectivamente.

Para a não utilização da pré-ênfase, o melhor resultado para Redes Neurais foi obtido a partir dos coeficientes Delta-cepstrais (Tabela 5.36), 63,33% para 2 neurônios na camada escondida. Para Quantização Vetorial, o melhor resultado obtido foi a partir dos coeficientes Cepstrais (Tabela 5.39), 82,05%. Para GMM, houve um empate entre coeficientes Cepstrais, coeficientes Delta-cepstrais e o combinado de coeficientes LPC e Cepstrais (tabelas 5.43, 5.45 e 5.47 respectivamente), com 76,92% de acerto para os modelos com 4, 24 e 4 componentes, respectivamente. O pior resultado para Redes Neurais foi novamente com coeficientes Cepstrais (Tabela 5.34), com 54,52% para 14 neurônios na camada escondida. Para Quantização Vetorial e para o GMM, os piores resultados foram obtidos com utilizam LPC (tabelas 5.39 e 5.41), com 66,67% e 74,36% para o modelo com 4 componentes, respectivamente.

Para os três classificadores, houve um aumento das taxas de acerto para coeficientes Cepstrais quando não foi utilizado pré-ênfase em comparação com a utilização de pré-ênfase, o que corrobora com (ZWETSCH et al., 2006), que recomenda não utilizar a pré-ênfase com coeficientes Cepstrais para diferenciar entre diferentes patologias, porque a pré-ênfase altera o sinal de excitação da glote, afetado de maneira diferente por cada patologia.

Na Tabela 5.48, estão apresentadas as melhores médias de taxas de acerto obtidas para Redes Neurais MLP. A coluna "Característica" indica a característica utilizada, "Pré-ênfase" indica a utilização ou não de pré-ênfase, "N" é a quantidade de neurônios na camada escondida, "Edema (%)", "O. P. (%)" e "Total (%)" mostram as taxas de acerto para voz afetada por edema, voz afetada por outra patologia e total, respectivamente, e os desvios padrão para voz afetada por edema e afetada por outra patologia estão nas colunas "σ_E" e "σ_{OP}", respectivamente.

Tabela 5.48: Melhores taxas de acerto entre duas classes (Edema *versus* Outra Patologia) para Redes Neurais MLP.

Característica	Pré-ênfase	N	Edema (%)	O. P. (%)	Total (%)	σ_E	σ_{OP}
LPC	SIM	8	59,69	65,00	60,95	12,80	17,16
LPC	NÃO	14	59,06	65,00	60,48	11,64	17,80
Cepstral	SIM	6	47,50	53,00	48,81	18,45	25,84
Cepstral	NÃO	14	54,06	56,00	54,52	10,93	15,78
Delta-cepstral	SIM	14	67,50	40,00	60,95	9,68	17,64
Delta-cepstral	NÃO	2	68,13	48,00	63,33	9,06	18,74
LPC+Cepstral	SIM	8	55,31	53,00	54,76	14,44	21,11
LPC+Cepstral	NÃO	6	61,25	59,00	60,71	9,79	15,24

As médias das taxas de acerto foram muito baixas, no caso de coeficientes Cepstrais utilizando pré-ênfase ficando abaixo de 50%. O melhor resultado foi obtido utilizando-se coeficientes Delta-cepstrais sem pré-ênfase, cerca de 63,33%. Apesar disso, os coeficientes Delta-cepstrais demonstraram baixo poder de discriminação para sinais da classe "outra patologia". A média total de acerto mais elevada em comparação à outras características se deve a uma melhor taxa de acerto entre os sinais de voz afetados por edema, que eram ampla maioria no conjunto de teste (32 sinais contra 10 sinais de outras patologias), e, portanto, teve um peso maior na média total.

Na Tabela 5.49, são apresentadas as melhores médias de taxas de acerto obtidas para GMM. A coluna "Característica" indica a característica utilizada, "Pré-ênfase" indica a utilização ou não de pré-ênfase, "Componentes" é a quantidade de componentes do modelo, "Edema (%)", "O. P. (%)" e "Total (%)" mostram as taxas de acerto para voz afetada por edema, voz afetada por outra patologia e total, respectivamente.

Tabela 5.49: Melhores taxas de acerto entre duas classes (Edema *versus* Outra Patologia) para GMM.

Característica	Pré-ênfase	Componentes	Edema (%)	O. P. (%)	Total (%)
LPC	SIM	64	83,33	76,19	79,49
LPC	NÃO	4	72,22	76,19	74,36
Cepstral	SIM	4	72,22	71,43	71,79
Cepstral	NÃO	4	61,11	90,48	76,92
Delta-cepstral	SIM	2	55,56	95,24	76,92
Delta-cepstral	NÃO	24	61,11	90,48	76,92
LPC+Cepstral	SIM	4	72,22	76,19	74,36
LPC+Cepstral	NÃO	4	88,89	66,67	76,92

Os coeficientes LPC proporcionaram a melhor taxa de acerto, 79,49%, superando consideravelmente a taxa de acerto obtida por esses na Quantização Vetorial (Tabela 5.39), 66,67%. As outras características, exceto coeficientes Cepstrais, também superaram as taxas de acerto da Quantização Vetorial. Entretanto, a maior taxa obtida com Quantização Vetorial, quando se utiliza coeficientes Cepstrais sem pré-ênfase (Tabela 5.39), 82,05%, é superior a maior taxa obtida pelo GMM. Grande parte das características proporcionou sua melhor taxa de acerto ao se utilizar 4 componentes no modelo, mostrando que a classe edema também pode ser modelada com um baixo número de componentes para se diferenciar de outras patologias.

5.3 Considerações Finais

As melhores taxas de acerto de cada abordagem para diferenciar voz normal de voz afetada por patologia, e para diferenciar voz afetada por edema de voz afetada por outra patologia estão nas Tabelas 5.50 e 5.51, respectivamente, em que a coluna "Característica" é o tipo de característica utilizada, "Pré-ênfase" indica a presença ou não de pré-ênfase, "Classificador" mostra o classificador utilizado na abordagem, e "Taxa de Acerto (%)" são as melhores taxas de acerto para cada abordagem.

Tabela 5.50: Melhores taxas de acerto entre duas classes (Normal *versus* Patologia) para cada uma das abordagens investigadas.

Abordagem	Característica	Pré-ênfase	Classificador	Taxa de Acerto (%)
1	LPC	SIM	MLP	92,28
	LPC	NÃO	MLP	89,47
	Cepstral	SIM	MLP	93,86
	Cepstral	NÃO	MLP	94,74
	Delta-cepstral	SIM	MLP	86,49
	Delta-cepstral	NÃO	MLP	87,02
	LPC+Cepstral	SIM	MLP	94,04
	LPC+Cepstral	NÃO	MLP	92,98
2	LPC	SIM	QV	92,39
	LPC	NÃO	QV	91,30
	Cepstral	SIM	QV	91,30
	Cepstral	NÃO	QV	91,30
	Delta-cepstral	SIM	QV	90,22
	Delta-cepstral	NÃO	QV	89,13
	LPC+Cepstral	SIM	QV	91,30
	LPC+Cepstral	NÃO	QV	89,13
3	LPC	SIM	GMM	89,13
	LPC	NÃO	GMM	91,30
	Cepstral	SIM	GMM	88,04
	Cepstral	NÃO	GMM	93,48
	Delta-cepstral	SIM	GMM	88,04
	Delta-cepstral	NÃO	GMM	85,87
	LPC+Cepstral	SIM	GMM	89,13
	LPC+Cepstral	NÃO	GMM	91,30

Tabela 5.51: Melhores taxas de acerto entre duas classes (Edema *versus* Outra Patologia) para cada uma das abordagens investigadas.

Abordagem	Característica	Pré-ênfase	Classificador	Taxa de Acerto (%)
1	LPC	SIM	MLP	60,95
	LPC	NÃO	MLP	60,48
	Cepstral	SIM	MLP	48,81
	Cepstral	NÃO	MLP	54,52
	Delta-cepstral	SIM	MLP	60,95
	Delta-cepstral	NÃO	MLP	63,33
	LPC+Cepstral	SIM	MLP	54,76
	LPC+Cepstral	NÃO	MLP	60,71
2	LPC	SIM	QV	66,67
	LPC	NÃO	QV	66,67
	Cepstral	SIM	QV	74,36
	Cepstral	NÃO	QV	82,05
	Delta-cepstral	SIM	QV	76,92
	Delta-cepstral	NÃO	QV	71,79
	Cepstral+LPC	SIM	QV	71,79
	Cepstral+LPC	NÃO	QV	71,79
3	LPC	SIM	GMM	79,49
	LPC	NÃO	GMM	74,36
	Cepstral	SIM	GMM	71,79
	Cepstral	NÃO	GMM	76,92
	Delta-cepstral	SIM	GMM	76,92
	Delta-cepstral	NÃO	GMM	76,92
	LPC+Cepstral	SIM	GMM	74,36
	LPC+Cepstral	NÃO	GMM	76,92

Para ambas as classificações, o melhor resultado foi obtido ao se utilizar coeficientes Cepstrais sem pré-ênfase. Para diferenciar voz normal de voz afetada por patologia, o melhor resultado foi 94,74%, obtido na abordagem que utiliza Redes Neurais, e para diferenciar

voz afetada por edema de voz afetada por outra patologia, o melhor resultado obtido foi de 82,05%, na abordagem que utiliza Quantização Vetorial.

O uso de Redes Neurais proporcionou um resultado muito abaixo dos outros classificadores para diferenciar voz afetada por edema de voz afetada por outra patologia. Isso se deve ao número pequeno de exemplos de treinamento. Quantização Vetorial e GMM foram melhores porque sistemas de classificação utilizando apenas uma classe se adequam melhor a problemas com conjunto de treinamento de tamanho reduzido (JOHANNES, 2001).

Por fim, faz-se necessário novos estudos para GMM por terem sido aplicados neste trabalho em experimento único, enquanto que cada arquitetura de rede neural foi testada várias vezes, de forma a obter-se a média das taxas de acerto. Um refinamento maior no GMM, com variação de parâmetros de entrada, pode melhorar o desempenho.

Capítulo 6

Considerações Finais e Sugestões para Trabalhos Futuros

Neste capítulo, é apresentado um resumo do trabalho com seus principais pontos, os objetivos alcançados, as contribuições, e sugestões para trabalhos futuros.

6.1 Resumo da Pesquisa

No Capítulo 1, foi apresentado o problema referente ao diagnóstico de patologias da laringe, e a motivação para trabalho, que residiu do fato de que as técnicas existentes para diagnóstico são técnicas ou subjetivas ou invasivas, causando desconforto ao paciente. Foram apresentados os objetivos do trabalho, sendo o principal o estudo de técnicas para classificação de vozes afetadas por patologias da laringe visando auxiliar o diagnóstico de um especialista.

No Capítulo 2, foi apresentada a fundamentação teórica do trabalho. Foi descrita a fisiologia e algumas patologias da fala. Sobre a fisiologia, foram apresentados maiores detalhes sobre a laringe e as dobras vocais. As seguintes patologias foram detalhadas naquele capítulo: (i) edema de Reinke, (ii) cistos vocais, (iii) nódulos vocais e (iv) paralisia. Por fim, foram apresentadas as etapas do processamento digital do sinal de voz, mais especificamente, as etapas de pré-processamento, extração de características e treinamento/classificação. Em se tratando da extração de características, foram descritos os algoritmos para obtenção dos coeficientes LPC, Cepstrais e Delta-cepstrais. Os classificadores abordados foram Redes Neurais Multilayer Perceptron, Quantização Vetorial e Modelos de Misturas de Gaussianas.

No Capítulo 3, foi apresentada uma revisão bibliográfica de trabalhos realizados nessa área de detecção de patologias da laringe.

No Capítulo 4, foi apresentada a abordagem proposta, com a descrição da base de dados, da metodologia para extração das características e para treinamento e classificação dos classificadores.

No Capítulo 5, foram apresentados e analisados os resultados obtidos com os testes dos classificadores. Foi indicado que a não utilização da pré-ênfase melhorou os resultados de classificação utilizando coeficientes Cepstrais para classificação entre voz afetada por edema e voz afetada por outra patologia, corroborando com (ZWETSCH et al., 2006). As Redes Neurais proporcionaram melhor desempenho para classificar entre voz normal e voz afetada por patologia, e a Quantização Vetorial foi melhor para classificar entre voz afetada por edema e voz afetada por outra patologia. Para ambos os casos, utilizando coeficientes Cepstrais sem da pré-ênfase.

Diante dos resultados obtidos ao longo da pesquisa, e fazendo uma comparação com os resultados obtidos em outros trabalhos apresentados na revisão de literatura, pode-se considerar que o objetivo da pesquisa foi alcançado com sucesso.

6.2 Contribuições

Foi realizado um estudo entre diferentes classificadores a fim de estabelecer uma comparação entre esses. Mostrou-se a eficácia de Redes Neurais em relação aos outros classificadores para distinção entre voz normal e voz afetada por patologia. E mostrou-se a eficiência da Quantização Vetorial, que é o classificador mais simples, para distinção entre voz afetada por edema e voz afetada por outra patologia.

Na literatura, poucos trabalhos realizam classificação entre patologias, normalmente é realizada apenas entre voz normal e voz afetada por patologia, sem distinguir qual patologia afeta o paciente. E os trabalhos que classificam entre patologias normalmente possuem resultados abaixo de 90% ((MARTINEZ; RUFINER, 2000; SCHOLOTTHAUER; TORRES, 2006; FONSECA, 2008; COSTA, 2008)). Este trabalho mostrou que Redes Neurais e GMM não são eficazes para classificação entre patologias, ficando abaixo dos resultados obtidos pela Quantização Vetorial, conforme verificado também em Costa (2008).

Observou-se, também, que a não utilização da pré-ênfase na fase de processamento do sinal melhora a precisão do sistema ao utilizar coeficientes Cepstrais em qualquer classificador, e piora em relação aos coeficientes LPC. A melhoria se dá porque a pré-ênfase altera o sinal de excitação da glote, e as patologias afetam a glote de maneira diferente. A piora em relação aos coeficientes LPC se dá devido a menor tolerância que os coeficientes LPC têm em relação ao ruído do sinal de voz afetada por patologia, que é atenuado na pré-ênfase.

Dois artigos foram publicados ao longo deste trabalho: o primeiro (MARINUS et al., 2009b) utiliza coeficientes LPC e Redes Neurais MLP para classificação entre voz normal e voz afetada por patologia; e o segundo (MARINUS et al., 2009a) utiliza coeficientes Cepstrais e Redes Neurais MLP para classificação entre voz normal, voz afetada por edema e voz afetada por outra patologia.

6.3 Sugestões de Trabalhos Futuros

Como sugestões para trabalhos futuros, tem-se o estudo de novos métodos para a extração de características, como *Wavelets*, o estudo de novos métodos para classificação, como *Support Vector Machine*, Mapas auto-organizáveis, dentre outros.

Outra sugestão seria a ampliação da base de dados de voz afetada por patologia, em especial voz afetada por alguma patologia que não seja edema, e o estudo de classificação entre mais de uma patologia além de edema, como nódulos e cistos.

Por fim, realizar testes estatísticos de significância e realizar novos experimentos com GMM, modificando parâmetros de inicialização do modelo, utilizando classificação cruzada, e outros métodos mais robustos de classificação.

Referências Bibliográficas

ABREU, M. H. L. de. *Edema de Reinke - Aspectos Gerais e Tratamento*. 1999. Monografia de conclusão do curso de especialização em Voz - Centro de Especialização em Fonoaudiologia Clínica.

ADNENE, C.; LAMIA, B. Analysis of pathological voices by speech processing. In: *Proceedings of the Seventh International Symposium on Signal Processing and Its Applications*. Paris: IEEE Publishers, 2003. v. 1, p. 365–367.

AGUIAR-NETO, B. G. *Signalaufbereitung in Digitalen Sprachubertragungssystemen*. Tese (Doutorado) — Technischen Universitat Berlin, 1987. Vom Fachbereich Elektrotechnik der Technischen Universitat Berlin zur Verleihung des akademischen Grades Doktor-Ingenieur genehmigte Dissertation.

AGUIAR-NETO, B. G.; COSTA, S. C.; FECHINE, J. M. Lpc modeling and cepstral analysis applied to vocal fold pathology detection. *International Journal of Functional Informatics and Personalised Medicine Issue*, v. 1, n. 2, p. 156–170, 2008.

ALBERNAZ, P. L. M.; GANANA, M. M.; FUKUDA, Y. *Otorrinolaringologia para o clínico geral*. São Paulo: Byk, 1997.

ALENCAR, V. F. S. *Atributos e Domínios de Interpolação Eficientes em Reconhecimento de Voz Distribuído*. Dissertação (Mestrado) — Pontifícia Universidade Católica, 2005. Dissertação de Mestrado.

ALONSO, J. B. et al. Automatic detection of pathologies in the voice by hos based parameters. *EURASIP Journal on Applied Signal Processing*, v. 4, p. 275–284, 2001.

ALWAN, A. et al. Time and frequency synthesis parameters for severe pathological voice qualities. In: *Proceedings of the International Congress of Phonetic Sciences*. Estolcomo: American Speech-Language-Hearing Association, 1995. p. 250–253.

ATAL, B. S.; HANAUER, S. L. Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acourticol Society of America*, p. 637–655, 1971.

BEHLAU, M. *Voz - O Livro do Especialista*. Rio de Janeiro: Revinter, 2001.

BEHLAU, M. et al. Disfonias funcionais. In: *Voz - O Livro do Especialista*. Rio de Janeiro: Revinter, 2001. v. 1, p. 248–270.

BEHLAU, M.; PONTES, P. *Avaliação e tratamento das disfonias*. São Paulo: Lovise, 1995.

BENJAMIN, B. *Cirurgia Endolaríngea*. Rio de Janeiro: Revinter, 2000.

BOUCHAYER, M. et al. Epidermoid cysts, sulci, and mucosal bridges of the vocal cord; a report of 157 cases. *Laringoscope*, v. 95, p. 1087–1094, 1985.

BOYANOV, B. et al. Robust hybrid pitch detector. *Electronic Letters*, v. 29, n. 22, p. 1924–1926, 1993.

BRAGA, P. L. *Reconhecimento de voz dependente de locutor utilizando Redes Neurais Artificiais*. 2006. Trabalho de Conclusão de Curso - Universidade de Pernambuco.

BRANCO, A. B.; ROMARIZ, M. S. Doenças das cordas vocais e sua relação com o trabalho. *Comunicação em Ciências da Saúde*, v. 17, n. 1, p. 37–45, 2006.

CALINON, S. *Robot Programming by Demonstration: A Probabilistic Approach*. Lausanne: EPFL/CRC Press, 2009.

CARDOSO, D. P. *Identificação de locutor usando modelos de mistura de gaussianas*. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, 2009. Dissertação de Mestrado em Engenharia.

CASE, J. L. *Clinical Management of Voice Disorders*. Austin, Texas: Pro-ed Inc, 1996.

COLTON, R.; CASPER, J. Conduta médica e cirúrgica dos distúrbios vocais. In: *Compreendendo os Problemas de Voz*. Porto Alegre: Artes Médicas, 1996.

COSTA, H. O. et al. Caracterização do profissional da voz para o laringologista. *Revista Brasileira de Otorrinolaringologia*, v. 66, n. 2, p. 129–134, 2000.

COSTA, S. C. *Análise Acústica, baseadan no Modelo Linear de produção da fala, para discriminação de vozes patológicas*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2008. Doutorado em Engenharia Elétrica.

COSTA, W. C. A. *Reconhecimento de Fala Utilizando Modelos de Markov Escondidos (HMM's) de Densidades Contínuas*. Dissertação (Mestrado) — Universidade Federal da Paraba, 1994.

CROVATO, C. D. P. *Classificação de Sinais de Voz Utilizando a Transformada Wavelet Packet e Redes Neurais Artificiais*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2004. Dissertação de Mestrado em Engenharia Elétrica.

DAJER, M. E. *Padrões Visuais de Sinais de Voz através de Técnica de Análise de Não-Linear*. Dissertação (Mestrado) — Escola de Engenharia de São Carlos, 2006. Dissertação de Mestrado em Bioengenharia.

DANIEL, R.; BOONE, S.; MCFARLANE, C. *A voz e a terapia vocal*. Porto Alegre: Artes Médicas, 1994.

DAVIS, S. B. Acoustic characteristics of normal and pathological voices. In: *Speech and language: advances in basic research and practice*. New York: Academic Publishers, 1979. p. 271–314.

DIBAZAR, A. A.; BERGER, T. W.; NARAYANAN, S. S. Pathological voice assessment. In: *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Nova York: IEEE Publishers, 2006. p. 1669–1673.

ESPINOSA, C. H.; FERNáNDEZ-REDONDO, M.; GOMEZ, V. P. Diagnosis of vocal and voice disorders by the speech signal. In: *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*. Como, Itália: International Neural Network Society, 2000. v. 4, p. 253–258.

FAIRBANKS, G. *Voice and articulation drillbook*. 2. ed. Nova York: Harper, 1960. 196 p.

FALCãO, H. H. et al. O uso da entropia na discriminação de vozes patológicas. In: *Anais do 21º Congresso Brasileiro de Engenharia Biomédica*. Salvador: Sociedade Brasileira de Engenharia Biomédica, 2008. p. 1599–1602.

FECHINE, J. M. *Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2000. Doutorado em Engenharia Elétrica.

FECHINE, J. M.; AGUIAR-NETO, B. G. Modelamento de identidade vocal utilizando modelos de markov escondidos. In: *anais do XVI Congresso Nacional de Matemática Aplicada e Computacional - CNMAC*. Uberlândia: Sociedade Brasileira de Matemática Aplicada e Computacional, 1993.

FONSECA, E. S. *Wavelets, Predição Linear e LS-SVM Aplicados na Análise e Classificação de Sinais de Vozes Patológicas*. Tese (Doutorado) — Escola de Engenharia de São Carlos - Universidade de São Paulo, 2008. Doutorado em Engenharia Elétrica.

FREDOUILLE, C. et al. Application of automatic speaker recognition techniques to pathological voice assessment (dysphonia). In: *Proceeding of 9th European Conference on Speech Communication and Technology, Interspeech*. Lisboa: International Speech Communication Association, 2005. p. 149–152.

FUKUDA, Y. *Otorrinolaringologia: Guias de Medicina Ambulatorial e Hospitalar*. So Paulo: Manole, 2003.

FURUI, S. *Digital Speech Processing, Synthesis, and Recognition*. 2. ed. Nova York: CRC Press, 2001.

GAMFOLO et al. Darpa timit acoustic-phonetic continuos speech corpus documentation. *National Institute of Standards and Technology*, 1993.

GODINO-LLORENTE, J. I.; GÓMEZ-VILDA, P. Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering*, v. 51, n. 2, p. 380–384, 2004.

GODINO-LLORENTE, J. I.; GÓMEZ-VILDA, P.; BLANCO-VELASCO, M. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, v. 53, n. 10, p. 1943–1953, 2006.

GONZáLES, J. N. *Fonacion y Alteraciones de la Laringe*. Buenos Aires: Panamericana, 1990.

GOTAAS, C.; STARR, C. D. Vocal fatigue among teachers. *Folia Fhoniatr*, v. 45, p. 120–129, 1993.

GRAY, S. D. Basement membrane zone injury in vocal nodules. In: *Vocal Fold Physiology*. Stockholm, Sweden: Singular, 1991. p. 21–27.

GREEN, G. Psyco-behavioral characteristics of children with vocal nodules: Wpbic ratings. *Journal of Speech and Hearing Research*, v. 54, p. 306–312, 1989.

GREENE, M. C. L. Distúrbios da voz. In: *Distúrbios da Voz*. São Paulo: Manole, 1989. p. 345–373.

GUYTON, A.; HALL, J. *Tratado de fisiologia médica*. 9. ed. Rio de Janeiro: Guanabara Koogan, 1997. 638 p.

HALL, M. et al. The weka data mining software: An update. *SIGKDD Explorations*, v. 11, n. 1, 2009.

HAMMARBERG, B. Perception and acoustics of voice disorders: a combined approach. In: *Proceedings of the VOICEDATA98, Symposium on databases in voice quality research and education*. Utrecht: Utrecht Institute of Linguistics OTS, 1998. p. 1–6.

HANSEN, J. H.; GAVDIDIA-CEBALLOS, L.; KAISER, R. J. F. A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment. *IEEE Transactions on Biomedical Engineering March*, v. 45, p. 937–940, 1998.

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

HAYKIN, S.; VEEN, B. V. *Signals and systems*. Nova York: Wiley, 2002.

HECHT-NIELSEN, R. Theory of the backpropagation neural network. In: *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. Washington, USA: International Neural Network Society, 1989. v. 1, p. 593–605.

HERSAN, R. C. G. P. Avaliação de voz em crianças. *Pró-Fono Revista de Atualização Científica*, v. 3, p. 3–9, 1991.

HIRANO, M. Structure of the vocal folds in normal and disease states: anatomical and physical studies. In: *Proceedings of the Conference on the Assessment of Vocal Pathology*. Ludlow: ASHA Report 17, 1981. p. 11–30.

HOCEVAR-BOLTEZAR, I.; RADSEL, Z.; ZARGI, M. The role of allergy in the etiopathogenisis of laryngeal mucosal lesions. In: *Acta-Otolaryngol-Suppl-Stockh*. [S.l.: s.n.], 1997. p. 134–137.

IMAMURA, R.; TSUJU, D. H.; SENNES, L. U. Fisiologia da laringe. In: *Tratado de Otorrinolaringologia*. São Paulo: Rocca, 2002. p. 743–750.

JOHANNES, D. M. *One-class classification: Concept-learning in the absence of counter-examples*. Tese (Doutorado) — Technische Universiteit Delft, 2001.

JR., J. R. D.; HANSEN, J. H. L.; PROAKIS, J. G. *Discrete-Time Processing of Speech Signals*. New York: IEEE Press, 2000.

KAY-ELEMETRICS. *Kay Elemetrics Corp. Disordered Voice Database, Model 4337*. 3. ed. [S.l.]: Massachussetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab, 1994.

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

KOUFMAN, J. A.; ISAACSON, E. G. *Functional Voice Disorders*. Nova York: Otolaryngologic Clinics of North America, 1991.

KUKHARCHIK, P. et al. Vocal fold pathology detection using modified wavelet-like features and support vector machines. In: *Proceedings of 15th European Signal Processing Conference*. Poznan, Polônia: EURASIP, 2007.

LINDE, Y.; BUZO, A.; GRAY, R. M. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, v. 28, n. 1, p. 84–95, 1980.

LITTLE, M. et al. Nonlinear, biophysically-informed speech pathology detection. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. Nova York: IEEE Publishers, 2006.

MAKHOUL, J. Linear prediction: A tutorial review. In: *Proceedings of the IEEE*. Cambridge: IEEE Press, 1975. p. 561–580.

MALLAT, S. *A wavelet tour of signal processing*. San Diego CA: Academic, 1998.

MAMMONE, R. J.; ZHANG, X.; RAMACHANDRAN, R. P. Robust speaker recognition - a feature-based approach. *IEEE Signal Processing Magazine*, v. 13, n. 5, p. 58–71, 1996.

MANFREDI, C. Adaptive noise energy estimation in pathological speech signals. *IEEE Transactions on Biomedical Engineering*, v. 47, n. 11, p. 1538–1543, 2000.

MANFREDI, C.; PIERAZZI, L.; BRUSCAGLIONI, P. Pitch estimation for noise retrieval in time and frequency domain. *Med. Biol. Eng. Comput.*, v. 37, n. 2, p. 532–533, 1999.

MARINUS, J. V. M. L. et al. On the use of cepstral coefficients and multilayer perceptron networks for vocal fold edema diagnosis. In: *Proceedings of 9th International Conference on Information Technology and Applications in Biomedicine*. Lanarca - Chipre: IEEE Press, 2009a.

MARINUS, J. V. M. L. et al. Detecção automática de patologias da laringe usando codificação por predição linear e redes neurais mlp. In: *Anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Computacional*. Ouro Preto: Sociedade Brasileira de Redes Neurais, 2009b.

MARKAKI, M.; STYLIANOU, Y. Using modulation spectra for voice pathology detection and classification. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Minneapolis: IEEE Press, 2009. p. 2514–2517.

MARTINEZ, C. E.; RUFINER, H. L. Acoustic analysis of speech for detection of laryngeal pathologies. In: *Proceedings of the 22th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Chicago: IEEE Press, 2000. p. 2369–2372.

MATLAB. *MATLAB User's Guide*. Palo Alto: The Mathworks, Inc., 1998.

MONDAY, L. A. et al. Epidermoid cysts of the vocal cords. In: *Annals of Otology Rhinology & Laryngology*. [S.l.]: American Broncho-Esophagological Association, 1983. v. 92, p. 124–127.

MONDAY, L. A. et al. Diagnosis and treatment of intracordal cysts. *The Journal of Otolaryngology*, v. 10, n. 5, 1981.

MORAN, R. J. et al. Telephony-based voice pathology assessment using automated speech analysis. *IEEE Transactions on Biomedical Engineering*, v. 53, n. 3, p. 468–477, 2006.

MYERS, C.; ALEKSANDER, I. Output functions for probabilistic logic nodes. In: *Proceedings of IEEE International Conference on Artificial Neural Networks*. UK: IEEE Press, 1989. p. 310–314.

NEGREIROS, B. C. P. *Cisto em Prega Vocal*. 1997. Monografia de conclusão do curso de especialização em Voz - Centro de Especialização em Fonoaudiologia Clínica.

PAPARELLA, M. M.; SHUMRICK, D. A. *Otorrinolaringologia - Cabeza y Cuello*. Buenos Ares: Panamericana, 1982. 2453 p.

PARRAGA, A. *Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2002. Dissertação de Mestrado em Engenharia Elétrica.

PAULRA, M. P. et al. Fuzzy voice segment classifier for voice pathology classification. In: *Proceedings of the 6th International Colloquium on Signal Processing & Its Applications*. Malaca, Malásia: IEEE Press, 2010. p. 190–195.

PINTO, J. A. Paralisias da laringe. In: *Câncer da Laringe*. Rio de Janeiro: Revinter, 1997. p. 251–258.

QUEK, F. et al. Speech pauses and gestural holds in parkinson's disease. In: *Proceedings of International Conference on Spoken Language Processing*. Denver: Speech Research Lab, 2002. p. 2485–2488.

RABINER, L. R.; HUANG, B. H. *Fundamentals of Speech Recognition*. New Jersey: Prentice Hall, 1993.

RABINER, L. R.; LEVINSON, S. E.; SONDHI, M. M. On the application of vector quantization and hidden markovmodels to speaker-independent, isolated word recognition. *The Bell System Technical Journal*, v. 62, n. 4, p. 1075–1105, 1983.

RABINER, L. R.; SCHAFER, R. W. *Digital Processing of Speech Signals*. Upper Saddle River, New Jersey: Prentice Hall, 1978.

REYNOLDS, D. A. Speaker identification and verification using gaussian mixture seaker models. *Speech Communication*, p. 91–108, 1995.

RIBEIRO, C. E. M. *Processamento Digital de Fala*. Lisboa: Instituto Superior de Engenharia de Lisboa, 2003.

RITCHINGS, R. T.; MCGILLION, M.; MOORE, C. J. Pathological voice quality assessment using artificial neural networks. *Medical Engineering & Physics*, v. 24, p. 561–564, 2002.

ROSA, M. O.; PEREIRA, J. C.; GRELLET, M. Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Trans. Biomedical Engineering*, v. 47, n. 1, p. 96–104, 2000.

ROSENBLATT. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, p. 386–408, 1958.

SALHI, L.; TALBI, M.; CHERIF, A. Voice disorders identification using hybrid approach: Wavelet analysis and multilayer neural networks. *World Academy of Science, Engineering and Technology*, v. 45, p. 330–339, 2008.

SAPIR, S.; KEIDAR, A.; MATHERS-SCHMIDT, B. Vocal attrition inteachers: Survey findings. *Eur J Disord Commun*, v. 28, p. 177–185, 1993.

SCALASSARA, P. R. *Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias da Laringe*. Tese (Doutorado) — Universidade de São Paulo, 2009. Doutorado em Engenharia Elétrica.

SCHOLOTTHAUER, G. N.; TORRES, M. E. Automatic diagnosis of pathological voices. In: *Proceedings of the 6th WSEAS International Conference on Signal, Speech and Image Processing*. Lisboa: WSEAS, 2006. p. 150–155.

SCHOLOTTHAUER, G. N.; TORRES, M. E.; RUFINER, H. L. Pathological voice analysis and classification based on empirical mode decomposition. *Computer Science: Development of Multimodal Interfaces: Active Listening and Synchrony*, v. 5967, p. 364–381, 2010.

SILVA, A. J. S. *Quantização Vetorial: Aplicações a um Vocoder LPC*. Dissertação (Mestrado) — Universidade Federal da Paraíba, 1992.

SIMM, W. A.; ROBERTS, P. E.; JOYCE, M. J. Signal processing for use in the assessment of dysarthric speech. In: *The 3rd IEEE International Seminar on Medical Applications of Signal Processing*. Londres: IEE Seminar Digests, 2005. p. 147–152.

SOTOMAYOR, C. A. M. *Realce de Voz Aplicado à Verificação Automática de Locutor*. Dissertação (Mestrado) — Instituto Militar de Engenharia, 2003. Dissertação de Mestrado.

STEFFEN, N.; MOSCHETTI, M. B.; ZAFFARI, R. J. Cisto em pregas vocais - análise de 96 casos. *Revista Brasileira de Otorrinolaringologista*, v. 61, n. 3, 1995.

TOLBA, H.; O'SHAUGHNESSY, D. Voiced-unvoiced classification using the first mel frequency cepstral coefficient. In: *Proceedings of International Conference on Speech Processing*. Seul: [s.n.], 1997. v. 1, p. 137–142.

TUMA, J. et al. Configuração das pregas vestibulares em laringes de pacientes com nódulo vocal. *Revista Brasileira de Otorrinolaringologia*, v. 71, n. 5, p. 576–581, 2005.

VIEIRA, M. N. *Módulo Frontal para um Sistema de Reconhecimento Automático de Voz*. Dissertação (Mestrado) — Universidade de Campinas, 1989.

VOIGT, D. et al. Voice pathology classification by using features from high-speed videos. In: *Proceedings of the 12th Conference on Artificial Intelligence in Medicine: Artificial Intelligence in Medicine*. Berlin, Heidelberg: Springer-Verlag, 2009. p. 315–324.

WALLEN, E. J.; HANSEN, J. H. A screening test for speech pathology assessment using bjective quality measures. *ICSLP 96. Proc.*, v. 2, p. 776–779, 1996.

WILSON, D. K. *Problemas de voz em crianças*. São Paulo: Manole, 1993.

ZWETSCH, I. C. et al. Processamento digital de sinais no diagnóstico diferencial de doenças laríngeas benignas. *Scientia Medica*, v. 16, n. 3, p. 109–114, 2006.

Apêndice A

Base de Dados

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

Nº	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 CORDAS 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	LXD000	LXD22AN.NSP	85	F	unilateral left		English	White- not Hispanic

Nº	PAT_ID	FILE VOWEL'AH'	AGE	SEX	LOCATION	SMOKE	NATLANG	ORIGIN
28	MCA000	MCA07AN.NSP	37	F		N	Portuguese	White- not Hispanic
29	MCW001	MCW21AN.NSP	39	F	bilateral	N	English	White- not Hispanic
30	NFG000	NFG08AN.NSP	49	F	bilateral		English	White- not Hispanic
31	NLC000	NLC08AN.NSP	48	F	bilateral		English	White- not Hispanic
32	OAB000	OAB28AN.NSP	43	M	periarytenoid area		English	White- not Hispanic
33	PAT000	PAT10AN.NSP	33	M	unilateral left	Y	English	White- not Hispanic
34	PMF000	PMF03AN.NSP	34	F	bilateral		English	White- not Hispanic
35	RCC000	RCC11AN.NSP	49	F	bilateral	N	English	White- not Hispanic
36	RJL000	RJL28AN.NSP	47	M		Y	English	White- not Hispanic
37	RTL000	RTL17AN.NSP	39	M	bilateral	N	English	White- not Hispanic
38	RXP000	RXP02AN.NSP	26	M	bilateral	N	Frech/Creole	Black- not Hispanic
39	SLC000	SLC23AN.NSP	28	F	bilateral		English	White- not Hispanic
40	SXG000	SXG23AN.NSP	70	F		N	English	White- not Hispanic
41	TLP000	TLP13AN.NSP	24	F	bilateral	N	English	White- not Hispanic
42	VAW000	VAW07AN.NSP	39	F		N	English	White- not Hispanic
43	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	PMD000	PMD25AN.NSP	45	F	cyst	unilateral right		English	White- not Hispanic
6	SWS000	SWS04AN.NSP	26	F	cyst	unilateral left	Y	English	White- not Hispanic
7	BSA000	BSA08AN.NSP	69	M	paralysis	unilateral left	Y	Arabic	White- not Hispanic
8	CAR000	CAR10AN.NSP	66	F	paralysis	unilateral left	Y	English	White- not Hispanic
9	CTY000	CTY09AN.NSP	75	M	paralysis	unilateral left	Y	English	White- not Hispanic
10	DJP000	DJP04AN.NSP	43	M	paralysis	unilateral right	N	English	White- not Hispanic
11	EDG000	EDG19AN.NSP	80	F	paralysis	unilateral right		English	White- not Hispanic
12	EJH000	EJH24AN.NSP	49	M	paralysis	unilateral left	N	English	White- not Hispanic
13	DAC000	DAC26AN.NSP	64	F	paralysis	unilateral left	Y	English	White- not Hispanic
14	DAG000	DAG01AN.NSP	75	M	paralysis	unilateral left	Y	English	White- not Hispanic
15	JCC000	JCC10AN.NSP	48	F	vocal nodules	bilateral		English	White- not Hispanic
16	KAS000	KAS09AN.NSP	18	F	vocal nodules	bilateral	N	English	White- not Hispanic
17	KCG000	KCG25AN.NSP	39	F	vocal nodules	unilateral left		English	White- not Hispanic
18	MRC000	MRC20AN.NSP	40	F	vocal nodules	bilateral	N	Portuguese	White- not Hispanic
19	MXN000	MXN24AN.NSP	21	F	vocal nodules		N	Japanese	Asian or Pacific Islander
20	NJS000	NJS06AN.NSP	21	F	vocal nodules	bilateral	N	English	White- not Hispanic
21	SEC000	SEC02AN.NSP	21	F	vocal nodules	bilateral	N	English	White- not Hispanic