UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

Coordenação de Pós-Graduação em Ciência da Computação

APLICAÇÃO DO PPM AO RECONHECIMENTO DE PADRÕES VOCAIS PATOLÓGICOS

HILDEGARD PAULINO BARBOSA

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

ÁREA DE CONCENTRAÇÃO: CIÊNCIA DA COMPUTAÇÃO

LINHA DE PESQUISA: PROCESSAMENTO DIGITAL DE SINAIS

JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO

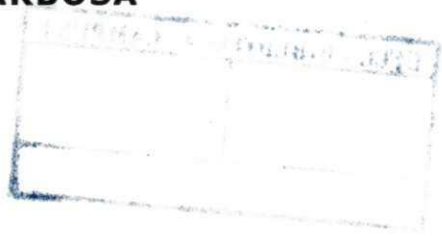
JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ

CAMPINA GRANDE

AGOSTO – 2013

MODELAGEM DE SINAIS DE VOZ, VIA PPM, APLICADA AO RECONHECIMENTO DE PADRÕES VOCAIS PATOLÓGICOS

HILDEGARD PAULINO BARBOSA



JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO - DRA
ORIENTADORA

JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ - DR
ORIENTADOR

SILVANA LUCIENE DO NASCIMENTO CUNHA COSTA - DRA
COMPONENTE DA BANCA

HERMAN MARTINS GOMES - DR
COMPONENTE DA BANCA

ELMAR MERCHER - DR
COMPONENTE DA BANCA

CAMPINA GRANDE
AGOSTO – 2013

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

B238m Barbosa, Hildegard Paulino.

Modelagem de sinais de voz via PPM, aplicada ao reconhecimento de padrões vocais patológicos / Hildegard Paulino Barbosa. – Campina Grande, 2013.

156 f.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

"Orientação: Profª. Drª. Joseana Macêdo Fechine Régis de Araújo, Prof. Dr. José Eustáquio Rangel de Queiroz".

Referências.

1. Detecção e Discriminação de Patologias da Fala. 2. Predição por Casamento Parcial (PPM). 3. Características Acústicas. 4. Características Temporais. 5. Características Estatísticas I. Araújo, Joseana Macêdo Fechine Régis. II. Queiroz, José Eustáquio Rangel de. III. Título.

CDU 004.4:81'234(043)

Resumo

A voz é o meio de comunicação mais utilizado pelo ser humano. Porém, o sistema fonador humano é suscetível a diversos tipos de patologias que podem prejudicar a produção da voz e, consequentemente, a comunicação.

Alguns tipos de exames têm sido utilizados para detectar estas patologias. Porém, eles apresentam desvantagens referentes à acurácia e ao conforto do paciente durante a aplicação, que podem desestimular a busca por tratamento. Por essa razão, técnicas computacionais têm sido empregadas com o intuito de detectar de modo confortável e preciso a presença e o tipo de patologia apresentada pelo sistema fonador. No entanto, os resultados obtidos ainda não possibilitam sua aplicação nas clínicas, principalmente pelo fato de ainda ser considerado um número reduzido de patologias.

Visando a contornar esse problema, esta pesquisa propõe uma abordagem fundamentada em um método ainda não utilizado neste contexto: a Predição por Casamento Parcial (*Prediction by Partial Matching* - PPM), concebida originalmente com fins à compressão de dados. O modelo criado e mantido a partir deste método é alimentado com características acústicas, temporais e estatísticas extraídas dos sinais de voz e permite sua classificação no que se refere à identificação da presença e do tipo de patologia a um baixo custo computacional (velocidade e recursos de armazenamento). Foram obtidos resultados satisfatórios no tocante à presença de patologias. Quanto à discriminação de patologias, os resultados sugerem um potencial do método, embora a sua aplicação ainda necessite de investigações mais aprofundadas.

Palavras-chave: Detecção e Discriminação de Patologias da Fala; Predição por Casamento Parcial (PPM); Características Acústicas; Características Temporais; Características Estatísticas.

Abstract

Voice is the most widely used means of communication of mankind. However, speech organs are susceptible to several sort of pathologies, which may harm voice production and, therefore, communication. Several techniques have been used to detect these pathologies. However, they present drawbacks related to accuracy and comfort of patients during the application, which may discourage search for treatment. Thence, computational techniques have been used in order to detect the presence and type of speech pathology comfortably and accurately. But, results are still not good enough for its application in clinics, due to the fact it is considered a small number of distinct pathologies.

Aiming to solve this problem, this research proposes using a method not previously employed in classification of vocal tract diseases: Prediction by Partial Matching (PPM), originally conceived for data compression purposes. The PPM model is fed with acoustical, temporal, and statistical features, all of them extracted from voice signals. This method allowed a satisfactory classification, concerning presence and type of pathology while requiring a low computational cost (speed and storage resources). It were obtained satisfactory results regarding presence of speech pathologies. With regard to pathologies discrimination, the results suggest that this is a highly promising technique, although its application still needs deeper investigations.

Keywords: Detection and Discrimination of Speech Pathologies; Prediction by Partial Matching; Temporal Features; Acoustical Features; Statistical Features

Lista de Figuras

Lista de Quadros

Lista de Abreviações e Siglas

DBLP	***DataBase systems and Logic Programming***
FPGA	***Field-Programmable Gate Array***
GMM	***Gaussian Mixture Models***
HMM	***Hidden Markov Models***
JDBC	***Java DataBase Connectivity***
LDA	***Linear Discriminant Analysis***
LPC	***Linear Predictive Coding***
LSI	***Latent Semantic Indexing***
MQR	**Medidas de Quantificação de Recorrência**
NLD	***Non-Linear Dynamics***
PPAM	***Prediction by Partial Approximate Matching***
PPM	***Prediction by Partial Matching***
RC	**Razão de Compressão**
SGBD	**Sistemas de Gerenciamento de Bancos de Dados**
SVM	***Support Vector Machines***
TEO	***Teager Energy Operator***
TMT	***Text Mining Toolkit***

Sumário

Capítulo 1

Considerações Iniciais

Nas subseções seguintes, será delineado ao leitor o escopo do trabalho, a partir dos seguintes elementos: (i) problemática envolvida; (ii) motivação para a execução da pesquisa; e (iii) abordagem utilizada para a resolução do problema.

1.1 Contextualização

A voz é o meio de comunicação mais importante e mais natural do ser humano, a partir da qual são expressos vontades, pensamentos, ordens e informações. Entretanto, para que a comunicação seja efetiva, é necessário o entendimento correto da voz enunciada por parte do interlocutor do processo, principalmente quando este é um dispositivo de reconhecimento ou de interpretação vocal que não dispõe das capacidades humanas para sua compreensão. Se isso não ocorrer, haverá maior propensão a equívocos, o que desestimulará a comunicação por ambas as partes causando, até mesmo, o constrangimento do locutor.

Esse tipo de problema, denominado *disfonia*, é causado muitas vezes por patologias da fala, às quais a voz humana é muito suscetível. Há a estimativa de que entre 3 e 10% da população mundial tenha o sistema de produção da fala comprometido por alguma patologia (STEMPLE; GLASE; KLABEN, 2010 apud COSTA et al., 2012). É comum um mesmo indivíduo ser acometido por até 8 patologias (KAY ELEMETRICS, 1994), as quais podem ser causadas por alterações psicoemocionais (FUKUDA, 2003), doenças neurodegenerativas (DAVIS, 1979; QUEK et al., 2002), mau uso da voz ou hábitos sociais não saudáveis, tais como o tabagismo e a ingestão de álcool (BEHLAU, 2001; STEMPLE; GLASE; KLABEN, 2010 apud COSTA et al., 2012). Algumas

destas causas explicam a ocorrência mais frequente de patologias da fala em fumantes e em categorias de profissionais que utilizam a voz como seu principal instrumento de trabalho, e.g., professores, cantores, radialistas, jornalistas (HAMMARBERG, 1998 apud MARINUS, 2010). Em um estudo com professores, 32% se auto-identificaram como portadores de alguma patologia da fala, contra 1% das demais ocupações investigadas (STEMPLE; GLASE; KLABEN, 2010 apud COSTA et al., 2012). Dentre as patologias da fala mais conhecidas estão o *Nódulo*, o *Edema*, a *Paralisia* e o *Pólipo*.

1.2 Motivação

Na detecção de patologias da fala, são usados, tradicionalmente, dois tipos de mecanismos. O primeiro, consiste na escuta da elocução vocal do paciente por um profissional (normalmente, um fonoaudiólogo ou um otorrinolaringologista), visando a diagnosticar a presença ou ausência de uma patologia. Até há poucos anos, este era o método mais usado (HU; LOIZOU, 2008; SÁENZ-LECHÓN et al. 2006 apud LONDOÑO, 2010). Entretanto, não é difícil perceber seu caráter subjetivo e propenso à indução de erros, principalmente nos casos em que a patologia se encontra em estágios iniciais, devido à forte dependência da experiência, da acurácia, do nível de fadiga e da sensibilidade do sistema auditivo do profissional[1] (LOPES et al., 2008; OATES, 2009 apud LONDOÑO, 2010). Diante do exposto, este tipo de exame deveria ser realizado apenas na inexistência de outras alternativas.

O segundo mecanismo consiste em procedimentos clínicos a partir dos quais a voz do paciente é avaliada por meio de recursos visuais. Dentre os exames mais comuns desta natureza estão (i) a *videolaringoscopia*, que consiste na visualização e no estudo da laringe e das dobras vocais do paciente, por meio de uma fibra óptica (luz contínua); e (ii) a *videoestroboscopia*, a qual lança mão de luz

[1] Diferentes diagnósticos podem ser dados por diferentes profissionais ou, até mesmo, pelo mesmo profissional, em ocasiões diferentes.

estroboscópica[2] (descontínua) para tal visualização e estudo (MARTINEZ; RUFINER, 2000). Estes exames, embora precisos, são considerados invasivos e desconfortáveis para o paciente, causando, em alguns casos, a ação de reflexo durante a aplicação, em função de sua sensibilidade laríngea, o que pode causar distorções nos dados obtidos e, com isso, acarretar falsos diagnósticos (ADNENE; LAMIA, 2003; ALONSO et al., 2001). Além disto, comprometem financeiramente ambas as partes, já que os equipamentos requeridos para executá-los são caros e sofisticados, obrigando o repasse dos custos ao paciente[3] e restringindo o seu acesso a grande parte da população.

Lieberman (1963) foi o primeiro a estudar as perturbações causadas por patologias na voz usando medidas acústicas e, desde então, devido às desvantagens dos métodos tradicionais apresentados anteriormente, inúmeras pesquisas sobre a detecção de patologias por computador têm sido desenvolvidas. A ideia é processar o sinal de voz digitalizado a partir de uma técnica computacional que apresente o máximo de precisão possível na detecção de patologias, com o intuito de auxiliar o clínico, dando a ele mais uma fonte de informação confiável para a tomada de decisão, e reduzir significativamente a necessidade e a frequência de exames visuais. Pode-se perceber que tal abordagem combina as vantagens dos dois tipos de exames supradescritos e elimina muitas de suas desvantagens, de modo que não se afigura irreal crer que poderá vir a ser a abordagem amplamente adotada em um futuro não muito distante.

Dando continuidade à pesquisa de Lieberman (1963), outros autores investigaram diversas técnicas computacionais no contexto de classificação de patologias da fala, tais como Redes Neurais, Máquinas de Suporte Vetorial, Análise Cepstral, dentre outras, quase sempre utilizando

---

[2] A argumentação de muitos profissionais é que se trata do único tipo de iluminação que permite visualizar a vibração das dobras vocais e emitir um diagnóstico acurado da patologia vocal investigada.

[3] Em entrevista, a Dra. Lavínia Brandão, fonoaudióloga na cidade de Campina Grande, afirmou que os preços desses exames giram em torno de R$ 140,00 (BRANDÃO, 2012).

sinais de vozes na elocução da vogal */ah/* sustentada[4]. Porém, na literatura revisada não foi encontrada nenhuma investigação associada à discriminação de patologias da fala, por meio da qual, dado um sinal de voz, seja diagnosticada a patologia apresentada por seu sistema fonador. Na maior parte da documentação existente, mesmo em registros recentes, obtém-se como principal resultado a detecção precisa (chegando ao percentual 100%) da presença ou ausência de patologias. Quando se trata de patologias específicas, no máximo três patologias são consideradas, embora atualmente se tenha conhecimento de mais de 120 (BRANDT, 2012; ARIAS-LONDOÑO et al., 2011; TAVARES et al., 2011; LIMA et al. 2012; PATIL; BALJEKAR, 2012; COSTA et al., 2012; OROZCO et al. 2012; KAY ELEMETRICS, 1994).

Por tal razão, a busca de técnicas computacionais que discriminem com precisão o máximo possível de patologias distintas se afigura um tema de pesquisa relevante, uma vez que a classificação robusta e acurada da patologia pode auxiliar o terapeuta a direcionar corretamente o tratamento do paciente. Vale ressaltar que cada patologia exige um tratamento diferente, dentre os quais se incluem a terapia vocal, a cirurgia e, até mesmo, a radioterapia (MARTINEZ; RUFINER, 2000), além de ser insuficiente, para fins de tratamento, o simples diagnóstico "o paciente apresenta uma patologia".

Um tipo de abordagem sobre a qual não se encontrou registro na revisão de literatura foi o uso de métodos estatísticos de compressão de dados visando ao diagnóstico de sinais de voz. Embora esses métodos tenham sido projetados inicialmente para comprimir dados, i.e., gerar um fluxo de dados menor a partir de outro de modo reversível, percebeu-se que o rico modelo estatístico gerado por alguns destes algoritmos a partir do fluxo original (contendo probabilidades de símbolos e sequências de

[4] O interesse nesta elocução advém do fato de que as dobras vocais vibram durante a toda a emissão vocal correspondente a esta vogal (permanecem sempre em movimento), facilitando a análise do comportamento do sistema fonador durante esse processo e a verificação da existência de patologias (MONTEIRO et al., 2011; GODINO-LLORENTE; GÓMEZ-VILDA; BLANCO-VELASCO, 2006).

símbolos referentes a este fluxo) pode ser empregado também em atividades de classificação, ao serem feitas consultas às probabilidades armazenadas nestes modelos durante a leitura e utilizando-as como base para a tomada de decisão. Por esta razão, isto é, pela crença na hipótese de que o emprego de um método estatístico de compressão com fins de classificação seria capaz de discriminar patologias, e também pelo fato de as investigações com outras técnicas não terem empregado esforços em discriminar o máximo de patologias distintas, se afigura importante o estudo da eficácia e eficiência de métodos estatísticos de compressão de dados na discriminação de patologias.

Segundo Medeiros et al. (2011), um dos métodos mais eficazes de compressão de dados atualmente denomina-se Predição por Casamento Parcial (*Prediction by Partial Matching* - PPM). Seu princípio de funcionamento será descrito em detalhes na Seção 2.4. Porém, deve-se considerar que bons resultados têm sido obtidos a partir do seu uso em atividades de compressão e classificação de arquivos binários, textos, sinais de eletrocardiograma e imagens, dentre outros tipos de sinais. Exemplos deste uso poderão ser encontrados nas Seções 3.2 e 3.3 e no Apêndice A. Em todos eles, fluxos de dados de teste foram comprimidos utilizando modelos PPM construídos durante a fase de treinamento. Se considerava que o modelo a partir do qual foi obtido o menor fluxo de dados comprimido em um determinado momento tinha sido construído com arquivos do mesmo tipo que o do fluxo de dados testado (original). Na verdade, a tomada de decisão era feita pelo uso do conceito de *Razão de Compressão* (RC), comum no campo da Compressão de Dados, que consiste na razão entre os tamanhos dos fluxos de dados original e comprimido. Sendo assim, o modelo a partir do qual foi gerado o menor fluxo de dados comprimido em um determinado momento era, na verdade, o modelo a partir do qual foi obtida a maior *Razão de Compressão* da compressão executada.

Portanto, a investigação da eficácia e eficiência do PPM na discriminação de patologias da fala a partir de sinais de voz se mostra

válida, sobretudo na expectativa de que se possa contribuir sob as seguintes perspectivas:

- Auxiliar no campo da Medicina Diagnóstica assistida por computador, no que se refere a patologias da voz;
- Evidenciar o uso do método PPM na discriminação de patologias da fala; e
- Oferecer uma alternativa não invasiva de auxílio ao diagnóstico de patologias da voz que tenha altos índices de acerto (que seja confiável ao clínico que utilizá-la), além de rápida execução e baixo consumo de recursos de memória com relação às técnicas usuais.

1.3 Questões de Pesquisa e Hipóteses

A partir do delineamento do quadro atual do campo de diagnóstico de patologias da fala e da verificação da importância do estudo da eficácia e eficiência de um método estatístico de compressão de dados neste campo (sendo o PPM o método escolhido para a condução dos estudos), foram formuladas as seguintes questões de pesquisa, a partir das quais surgiu a motivação para a pesquisa ora documentada:

- P_1: O método PPM é capaz de detectar a presença de patologias da fala com eficácia (baixo índice de erros)?
- P_2: O método PPM é capaz de discriminar entre patologias da fala com eficácia?
- P_3: O método PPM é capaz de realizar estas tarefas de modo eficiente (baixos tempo de execução e consumo de memória)?

A partir destas questões, foram formuladas hipóteses, as quais nortearam a pesquisa:

- H_1: É possível, utilizando o método PPM, obter altos percentuais de acerto na classificação de um sinal de voz como Normal ou Patológico;
- H_2: É possível discriminar diversas patologias utilizando métodos computacionais;

- H_3: É possível, utilizando o método PPM, obter altos percentuais de acerto na discriminação entre patologias referida na hipótese H_2;
- H_4: O PPM é capaz de realizar estas tarefas de modo eficiente, isto é, de modo rápido e com baixa utilização dos recursos de memória disponíveis.

1.4 Objetivos

1.4.1 Objetivo Geral

Esta pesquisa objetivou, principalmente, analisar a aplicação de métodos estatísticos de compressão de dados (mais especificamente, o PPM) na detecção e discriminação de diferentes patologias da fala, considerando aspectos de eficácia (percentuais de acerto elevados) e eficiência (tempo de resposta rápido e pouca utilização de recursos de hardware, a exemplo de memória).

1.4.2 Objetivos Específicos

Considerando o objetivo geral exposto na Seção 1.4.1, esta pesquisa foi conduzida visando a alcançar os seguintes objetivos específicos:

- Seleção da melhor configuração do método PPM com base no processamento estatístico (projeto experimental e testes de hipóteses) de seus resultados;
- Obtenção de percentuais de acerto elevados na classificação entre voz normal e voz patológica;
- Obtenção de percentuais de acerto elevados na classificação de patologias distintas;
- Modelagem do sistema de modo que sua execução seja rápida e com uso de poucos recursos de memória.

1.5 Estrutura da Dissertação

O restante deste documento está estruturado como segue: (i) no **Capítulo 2** (**Fundamentação Teórica**) são explorados os diversos

conceitos relacionados à pesquisa; (ii) o **Capítulo 3** (**Trabalhos Relacionados**) contém uma revisão da literatura da área cujo foco é a aplicação de técnicas diversas destinadas à detecção de patologias e à aplicação do PPM em processos de classificação e compressão; (iii) no **Capítulo 4** (**Descrição da Modelagem Aplicada**) é detalhado o procedimento metodológico adotado nos experimentos conduzidos, com ou sem êxito, visando a alcançar resultados satisfatórios; (iv) o **Capítulo 5** (**Apresentação e Discussão de Resultados**) contém a apresentação e discussão dos resultados dos diversos experimentos de classificação conduzidos e do processamento estatístico associado, visando a encontrar a melhor configuração do classificador[5]; (vi) no **Capítulo 6** (**Considerações Finais**) são integradas conclusões advindas dos resultados obtidos, contribuições associadas à pesquisa e sugestões de pesquisas que poderão ser conduzidas de modo a se obterem resultados mais abrangentes e satisfatórios para a abordagem ora documentada; (vii) no **Apêndice A** são mostrados resultados obtidos ao serem comprimidos (experimentalmente) diferentes tipos de arquivos com o PPM, entre eles letras de músicas, livros e sinais de voz; (viii) no **Apêndice B** são listados os arquivos utilizados nos experimentos conduzidos para esta pesquisa; (ix) no **Anexo A** é apresentada uma explicação detalhada do conceito de Projeto Experimental (técnica de apuração de resultados utilizada neste trabalho, conforme mencionado no Capítulo 4; (x) no **Anexo B** é explicado em detalhes o conceito de Intervalos de Confiança, bastante utilizado em análises estatísticas e também na que foi empreendida nesta pesquisa; (xi) no **Anexo C** está contido o artigo aceito para publicação no IADIS 2013, realizado no Texas, EUA; (xii) e no **Anexo D** está contido o artigo aceito para publicação no BRICS-CBIC 2013, realizado em Porto de Galinhas, no município de Ipojuca - PE.

---

[5] Aquela que retorna os melhores resultados.

Capítulo 2

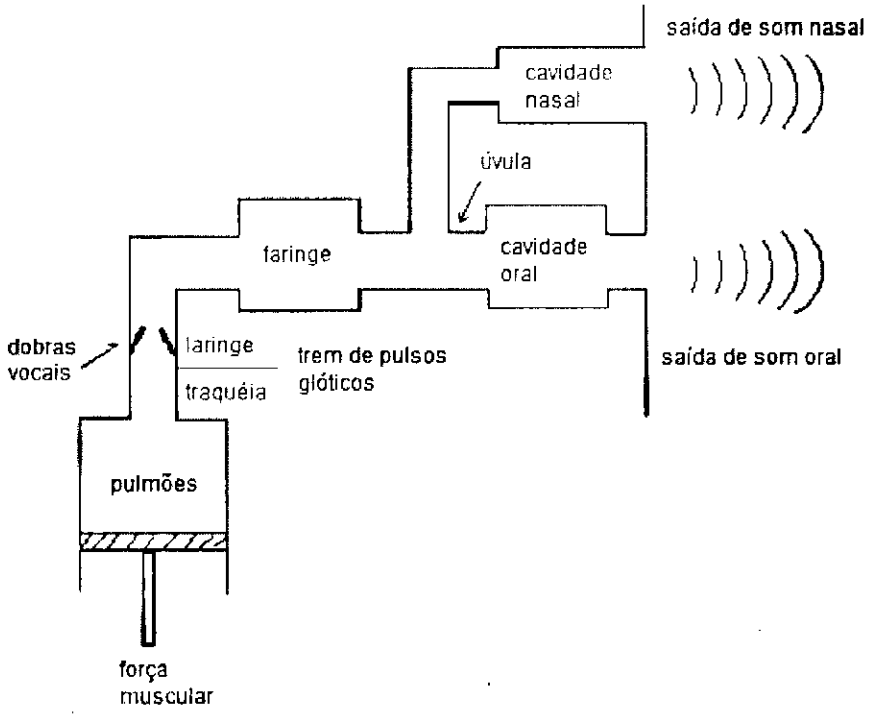
Fundamentação Teórica

Neste capítulo, são descritos tópicos relativos à produção da fala, às patologias que lhe afetam, às medidas que são comumente extraídas para caracterizá-la e ao classificador adotado para os experimentos.

2.1 Produção da Fala

Para se entender a importância do estudo das patologias da fala e seu impacto na comunicação vocal humana, é necessário compreender o funcionamento do sistema fonador humano. A Figura 1 contém um diagrama esquemático deste mecanismo vocal.

Figura 1 - Esquema simplificado de produção da fala



FONTE: Deller, Proakis & Hansen (1993)

A área da região vocal compreende a região que se estende da abertura das dobras vocais (que pode ter entre 0 - completamente

fechadas - e 20 cm^2 - completamente abertas - durante a produção da fala) aos lábios, sendo composta pela faringe e pela cavidade oral e determinada pelas posições da língua, lábios e maxilar. A área nasal se inicia na úvula e termina nas fossas nasais (cavidade nasal).

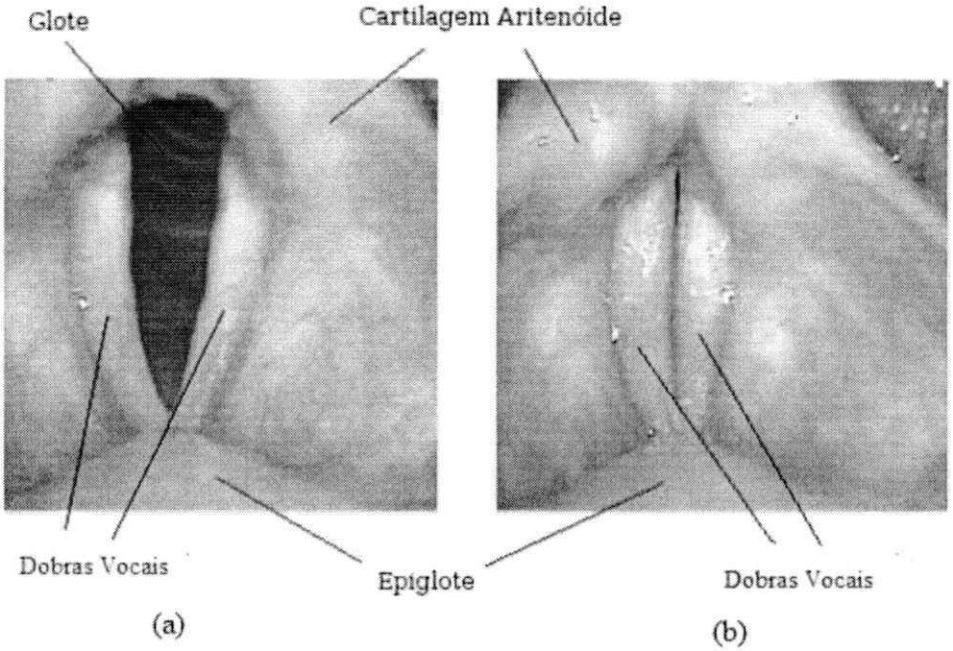
O processo de produção da fala se inicia com a expansão dos pulmões, permitindo a entrada de ar pelas narinas e pela boca por meio da inspiração. Esta etapa compreende o fornecimento de energia do processo de produção da voz e consiste na contração do diafragma, com posterior compressão da víscera e consequente expansão do volume de ar dos pulmões. Em seguida, o ar é processado pelos pulmões e o diafragma relaxa, voltando para sua posição de repouso, o que permite a liberação do ar por intermédio da traqueia. Neste estágio, dá-se início de fato à produção da voz, por meio da interferência dos diversos órgãos do sistema fonador.

O primeiro e também o principal órgão a causar interferência no ar liberado pelos pulmões é a laringe, que consiste em um tubo cartilaginoso que conecta o sistema respiratório (traqueia e pulmões) e o trato vocal e cavidade oral. Nela se localizam as dobras vocais, duas fibras elásticas ligadas às cartilagens aritenóides, que vibram durante a produção da fala. Essa vibração, na verdade, consiste em intervalos de completa abertura e completo fechamento do espaço entre as dobras, denominado glote. Esse movimento, porém, não ocorre ao acaso. Inicialmente em repouso (glote fechada), há o aumento da chamada pressão subglótica, fazendo com que as dobras vocais abram-se repentinamente, liberando o ar e diminuindo a pressão glótica. Essa diminuição relaxa as dobras vocais, ocasionando novo fechamento da glote. Esse ciclo dura enquanto durar a emissão vocal (COSTA, 2008; ANDRADE SOBRINHO, 2011; GODINO-LLORENTE, 2002 apud LONDOÑO, 2010; RABINER; SCHAFER, 1978; RUSSO; BEHLAU, 1993 apud FECHINE, 2000).

Em geral, para a passagem do ar, as dobras vocais estão em *abdução*, i.e., abertas e afastadas da linha média. Para que haja produção

da voz, é preciso que as dobras vocais estejam em *adução*, i.e, fechadas na linha média. Ambos os estados são mostrados na Figura 2.

Figura 2 - Dobras vocais em: abdução (a) e adução (b)



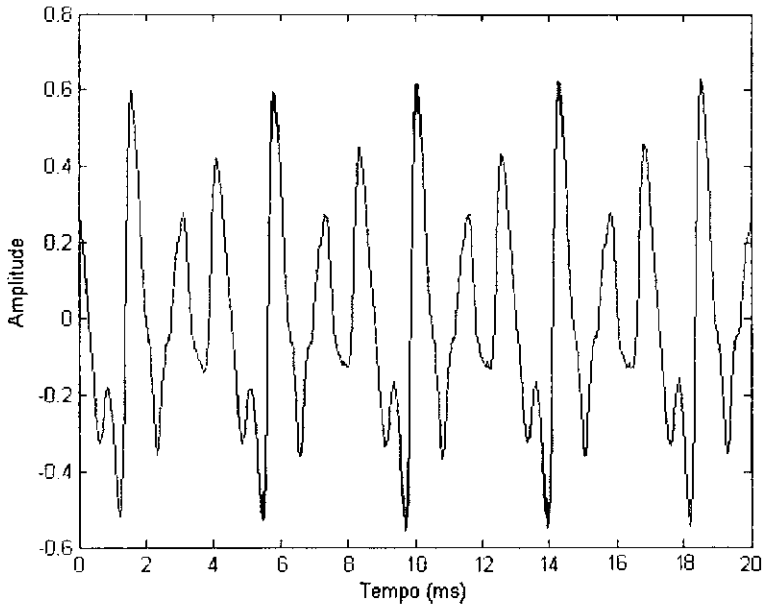
FONTE: Adaptada de Tortora e Grabowski (2002)

Sendo produzida como uma sequência de sons, a fala reflete o estado das dobras vocais, assim como as posições, a forma e o tamanho das várias articulações e as alterações que se processam ao longo do tempo da emissão vocal. Quando as dobras vocais formam uma abertura estreita, o fluxo de ar proveniente dos pulmões as faz vibrar, gerando pulsos aerodinâmicos periódicos, denominados *pulsos glotais*, responsáveis pela produção dos chamados sons *vozeados*. Por outro lado, quando a glote mantém-se levemente aberta, o fluxo de ar proveniente dos pulmões não é mais periódico, adquirindo características ruidosas e produzindo os sons denominados *não-vozeados*.

Na Figura 3, é ilustrada uma pequena parte de um sinal de voz saudável, na qual é possível perceber o movimento de vibração das dobras vocais ou ciclos. Elas vibram centenas de vezes por segundo e este movimento determina a Frequência Fundamental da voz, que interfere diretamente em sua tonalidade e varia bruscamente entre os gêneros

(detalhes desta medida são apresentados na Seção 2.3.2.1). Por exemplo, na Figura 3 são exibidos 20 ms da forma de onda de um sinal, no qual as dobras vocais vibram (abrem-se e fecham-se completamente) 5 vezes. Várias características das dobras vocais interferem na velocidade deste movimento e, consequentemente, alteram a Frequência Fundamental, tais como o comprimento, a massa, a elasticidade e a rigidez, dentre outros.

Figura 3 - Sinal de voz e os ciclos



Em seguida, o ar passa pelo trato vocal e, dependendo do som, também pelo trato nasal. O trato vocal é uma estrutura tubular que funciona como um ressonador, devido à excitação das moléculas de ar ao passarem por estruturas como faringe, cavidades oral e nasal, palato duro, língua e dentes, modulando os pulsos provenientes da glote. A ressonância é um fenômeno físico que ocorre quando uma estrutura é excitada por outra e passa a vibrar de modo similar. Neste caso, o pulso glótico passa a vibrar na mesma frequência do trato vocal, quando chega ali, conferindo à voz as características conhecidas, tais como a altura (ou intensidade) e o timbre. Tal frequência é denominada *formante* e depende das dimensões e da forma do trato vocal.

Posteriormente, o ar é conduzido à cavidade oral, na qual pode ser obstruído pela língua e/ou pelos lábios (que finalizam o trato vocal), ao

serem pronunciadas consoantes. Caso seja produzido um som nasal (e.g., palavras com til ou que terminam em m ou n), uma estrutura denominada *véu palatino* é abaixada, acoplando-se ao trato vocal via faringe e recebendo parte do ar por este expirado (COSTA, 2008; GODINO-LLORENTE, 2002 apud LONDOÑO, 2010; RABINER; SCHAFER, 1978 apud FECHINE, 2000; STEMPLE; GLASE; KLABEN, 2010).

Tendo sido descrito o mecanismo completo de produção da fala, é possível perceber que alguns dos componentes influenciam diretamente na qualidade vocal, a destacar: saúde laríngea (qualidade das dobras vocais), suporte respiratório e ressonância supraglótica (do trato vocal). Por exemplo, a baixa capacidade pulmonar pode limitar a vibração das dobras vocais, ao gerar pressão subglótica insuficiente para a produção da fala com altura e qualidade suficientes.

2.2 Patologias da Fala

Na seção anterior, foi descrito o funcionamento de um sistema fonador saudável. Porém, conforme anteriormente explicitado, esse sistema é suscetível a patologias de diferentes naturezas, que podem afetá-lo seriamente, as quais podem ser de ordem neurológica, motora ou psicoemocional, dentre outras. Nas seções seguintes, serão descritas as patologias consideradas nesta pesquisa e como cada uma delas afeta o sistema de produção da fala.

2.2.1 Paralisia

A Paralisia é uma patologia pertencente à categoria das patologias neurológicas, as quais são caracterizadas por interrupções na inervação da laringe. Trata-se da patologia neurológica mais comum. Ocorre devido a lesões em ramificações do *nervo vago* da laringe, denominadas *nervo laríngeo superior* e *nervo laríngeo recorrente*. Este último recebe este nome pelo fato de passar pela laringe duas vezes: partindo do cérebro, passa pelo pescoço, chega ao peito e volta à laringe. Quanto maior seu tamanho, maior sua suscetibilidade a lesões, razão pela qual lesões neste

nervo são muito mais frequentes (STEMPLE; GLASE; KLABEN, 2010; KOHLER, 2011; PARRAGA, 2002; BRANDT, 2012).

A lesão nestes nervos ocasiona a paralisia de uma ou ambas as dobras vocais. Se a Paralisia ocorrer em apenas uma dobra, dá-se o nome de Paralisia *Unilateral*. Se ocorrer em ambas, denomina-se *Bilateral*. Várias são as causas da Paralisia Unilateral, dentre as quais podem ser citadas trauma de parto, cirurgias intratorácicas, pós-intubação endotraqueal, vírus (paralisia viral), pressão sobre o nervo devido a um tumor, neoplasia maligna do pescoço ou trauma cervical. Quanto à Paralisia Bilateral, a principal causa tem sido a tireoidectomia, mas algumas das causas anteriormente citadas também são frequentes, tais como a pós-intubação endotraqueal, o trauma cervical e doenças malignas do pescoço (STEMPLE; GLASE; KLABEN, 2010; COSTA, 2008).

O prejuízo causado depende das dobras vocais que foram afetadas e de sua posição na laringe - mediana, paramediana ou lateral. No caso de uma Paralisia Unilateral em adução, com a dobra vocal em posição mediana (na linha do meio da glote), a qualidade vocal pode ser pouco afetada, pelo fato de a glote ainda ser fechada na vibração, o que cria a pressão subglótica e, por isso, não prejudica a produção da voz. Caso a dobra vocal afetada apresente frouxidão, a criação da pressão subglótica fica prejudicada. Ainda, embora a produção da voz não seja significativamente prejudicada pela paralisia nesta posição, há permanente obstrução da passagem de ar, pelo fato de a abertura da passagem do ar ser metade do tamanho normal, o que dificulta a respiração. Isto é notado mais perceptivelmente ao serem praticadas atividades tais como esportes e trabalho pesado. Se a dobra vocal estiver em posição paramediana - de 1 a 2 mm da linha do meio da glote, já é possível perceber alguns sintomas e a alteração na qualidade vocal, embora também seja possível a outra dobra vocal se estender um pouco além do normal, produzindo fechamento suficiente para uma boa fonação. Esta é a situação mais comum de Paralisia. Se a Paralisia ocorrer em abdução, a dobra vocal permanecerá em posição lateral (3 a 4 mm da

linha mediana) e o fechamento não ocorrerá, o que acarretará muita dificuldade na produção da voz e em diversas outras atividades, tais como a deglutição, sendo necessária a cirurgia ou a alimentação via sonda (STEMPLE; GLASE; KLABEN, 2010; BRANDT, 2012).

A Paralisia Bilateral é bem mais prejudicial, principalmente em posição mediana, pois prejudica seriamente a passagem de ar, vital ao ser humano. Mesmo em posição paramediana ou lateral, o prejuízo à produção da fala é muito maior do que aquele associado à Paralisia Unilateral.

Quanto aos sintomas, é comum os pacientes apresentarem elocução vocal com ruído de fundo[6], fadiga vocal, diplofonia (quando as dobras vocais vibram independentemente, em frequências diferentes), respiração ruidosa e falta de ar, dentre outros. É muito difícil para o paciente falar em ambientes com muito ruído, principalmente devido à baixa intensidade da fala (DANIEL; BOONE; McFARLANE, 1994; COLTON; CASPER, 1996 apud PARRAGA, 2002; STEMPLE; GLASE; KLABEN, 2010; KOHLER, 2011; PATIL; BALJEKAR, 2012).

2.2.2 Edema de Reinke

O Edema de Reinke é uma patologia estrutural. Patologias desta categoria são caracterizadas por mudanças na estrutura histológica das dobras vocais, afetando sua massa, tensão, flexibilidade e, consequentemente, seu padrão vibratório. No caso do Edema de Reinke, a mudança consiste no aumento de tamanho de uma (em estágios iniciais) ou ambas as dobras devido ao surgimento de um fluido viscoso em seu interior, mais especificamente no espaço de Reinke (primeiro anatomista a registrar as dobras vocais), o que modifica radicalmente o espaço da glote (STEMPLE; GLASE; KLABEN, 2010; HIRANO, 1981).

A principal causa desta patologia é o fumo, independentemente da idade, associado a outras causas, tais como uso excessivo da fala, ingestão demasiada de cafeína e/ou ingestão reduzida de água e refluxo

[6] Tradução usada neste documento para o termo em inglês *soprosity*.

gastresofágico. Há registros desta patologia em crianças com o hábito do fumo, embora seja muito mais comum em adultos entre 45 e 70 anos (KLEINSASSER, 1997; HOCEVAR-BOLTEZAR; RADSEL; ZARGI, 1997; PAPARELLA; SHUMRICK, 1982; BENJAMIN, 2000).

Devido ao surgimento do fluido viscoso em seu interior, as dobras vocais adquirem dimensões superiores àquelas normalmente exibidas, o que compromete o padrão vibratório e, consequentemente, a qualidade vocal. O paciente sente dificuldade de falar, uma vez que ocorre uma redução drástica em sua qualidade vocal, tanto em termos de frequência quanto de intensidade, de modo que sua voz torna-se bastante grave e rouca. Em mulheres, a emissão vocal pode até mesmo ser confundida com aquela de um indivíduo do sexo masculino (COSTA, 2008).

2.2.3 Outras

Outras patologias foram consideradas nesta pesquisa, a exemplo de Nódulo, Pólipo e Cisto. Entre elas, há em comum o fato de serem classificadas como lesões de massa nas dobras vocais, causando disfonias organofuncionais. Suas especificidades serão descritas brevemente, a seguir.

Os *Nódulos* nas dobras vocais também pertence à categoria das patologias estruturais. É uma das lesões benignas mais comuns e se caracteriza como protuberâncias bilaterais[7] simétricas de tamanho variável[8] nas dobras vocais, surgidas devido ao abuso vocal. Por esta razão, é a patologia mais frequente entre profissionais que fazem uso demasiado da voz, tais como professores, locutores, cantores (não treinados) e operadores de telefonia ou *telemarketing*, principalmente do sexo feminino. É também a mais frequente entre crianças em idade escolar, especialmente aquelas agitadas, que costumam fazer uso constante de gritos, fala excessiva, vocalizações explosivas, choro

[7] Há discordância quanto à existência da forma unilateral desta patologia. A maioria dos autores afirmou não existir, porém há autores que afirmaram existir esta forma, tais como Case (1996), Gonzáles (1990) e Wilson (1993).

[8] Pode variar do tamanho de uma cabeça de alfinete ao tamanho de uma ervilha.

prolongado, pigarro e falta de hidratação. Há dois tipos de Nódulos: Agudo e Crônico. Os Nódulos Agudos são mais gelatinosos, enquanto os Crônicos são mais rígidos. Um dos locais no qual podem surgir é no ponto de maior amplitude da vibração das dobras vocais. Seus sintomas incluem rouquidão e elocução vocal com ruído de fundo, devido principalmente ao fechamento falho da glote e à vibração irregular das dobras vocais (COSTA, 2008; GREEN, 1989; HERSAN, 1991; CASE, 1996; WILSON, 1993; STEMPLE; GLASE; KLABEN, 2010).

O *Pólipo* também pertence à categoria de patologias estruturais e, muitas vezes, é confundido com o Edema, diferenciando-se pelo fato de ser mais localizado e mais frequentemente unilateral (80% dos casos), enquanto o Edema é mais generalizado, i.e., atinge a totalidade das dobras vocais (BENJAMIN, 2000). Fisicamente, pode se parecer com um nódulo, sendo uma lesão composta por material gelatinoso que se desenvolve na camada superficial da lâmina própria, devido ao aumento da permeabilidade dos vasos. Similarmente ao nódulo, o uso excessivo da voz é a principal causa de surgimento desta patologia, sendo usual em indivíduos que costumam fazer uso da fala por longos períodos em ambientes ruidosos. A forma hemorrágica desta patologia se origina a partir de uma ruptura em um capilar da dobra vocal, com posterior sangramento e formação do Pólipo. O principal sintoma é a disfonia severa, mas rouquidão e elocução vocal com ruído de fundo também podem surgir (DANIEL; BOONE; McFARLANE, 1994). É a patologia que mais comumente exige remoção cirúrgica, principalmente se não se constata melhora rápida após conservação rigorosa da voz. Entre os pacientes, a incidência é duas vezes maior em homens do que em mulheres e a maioria apresenta entre 20 e 60 anos de idade (raramente aparece em crianças) (COSTA, 2008; DANIEL; BOONE; McFARLANE, 1994; STEMPLE; GLASE; KLABEN, 2010).

Por fim, o *Cisto* é uma patologia que pode aparecer principalmente devido à má formação congênita[9], mas também pode ser adquirida durante a vida, devido a uma obstrução na glândula mucosa ou abuso vocal. Pode ser encontrada a forma epidermóide, de retenção (também conhecido como intracordal) ou pseudocisto da dobra vocal. Caracteriza-se pelo aparecimento de um fluido viscoso localizado, fazendo surgir uma pequena protuberância, e por ser séssil (preso à dobra vocal diretamente pela base). Dentre os sintomas perceptíveis, listam-se disfonia, dificuldade ao falar, elocução vocal com ruído de fundo e instabilidade da fala. É mais comum em mulheres entre 20 e 50 anos (BOUCHAYER et al., 1985; MONDAY et al., 1983; PASSEROTI, em http://www.otorrinousp.org.br; STEMPLE; GLASE; KLABEN, 2010).

2.3 Análise de Sinais de Voz

Tendo sido apresentadas as principais patologias pesquisadas, nas próximas seções serão apresentados alguns dos indicadores mais comumente usados na sua identificação, a exemplo da Energia, Taxa de Cruzamento por Zero, Frequência Fundamental, *Jitter*, *Shimmer*, Entropia entre outros. Eles podem ser classificados em 3 categorias: temporais, acústicos e estatísticos.

Porém, primeiramente é válido mencionar que, embora os sinais de voz sejam **estacionários** em segmentos que duram entre 16 e 32 ms (suas propriedades estatísticas não variam com o tempo caso sejam consideradas segmentos dentro deste intervalo) (RABINER; SCHAFER, 1978; SOTOMAYOR, 2003), o trato vocal apresenta natureza dinâmica[10], o que afeta os parâmetros que representam a voz e, consequentemente, sua produção. Sendo assim, para não enviesar o processo de extração das medidas, i.e., para que os valores extraídos representem a realidade do sinal de voz manipulado, as medidas devem ser extraídas em segmentos dentro daquele intervalo, tais como as medidas apresentadas nas

[9] Durante a vida intra-uterina, na formação da laringe.
[10] Sua configuração varia com o tempo durante a produção da fala.

subseções a seguir, de modo que um arquivo seja representado por um conjunto de valores de determinada medida.

Essas medidas são chamadas de medidas de curto intervalo de tempo. Algumas medidas também são apresentadas como um valor único (média ou mediana), referente a todo o sinal, obtida a partir dos valores associados a cada segmento (COSTA, 2008; FECHINE, 2000). Elas são denominadas medidas de longo intervalo de tempo. Um exemplo de medida comumente apresentada em longo intervalo de tempo é a Frequência Fundamental.

2.3.1 Análise Temporal

Na categoria de análise temporal, são comumente utilizadas 4 medidas: Energia, Taxa de Cruzamento por Zero, Número Total de Picos e Diferença no Número de Picos.

2.3.1.1 Energia

A **Energia** de um segmento é obtida a partir da Equação 1 (FECHINE, 2000).

$$E_{seg} = N_A * E\{[s(n) - \mu_{s(n)}]^2\}, \quad (1)$$

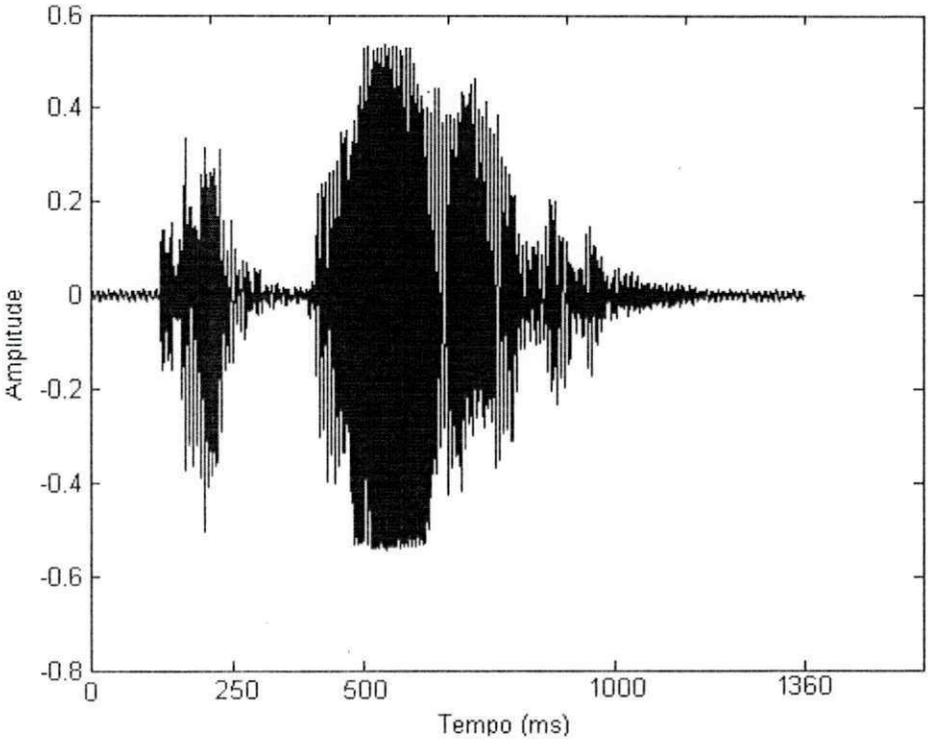
em que N_A é o tamanho do segmento, $s(n)$ representa a amplitude da n-ésima amostra de um segmento de um sinal de voz e $\mu_{s(n)}$ denota o valor médio de todas as amostras do segmento. Porém, para sinais de voz, que são considerados ergódicos e estacionários (conforme mencionado no início da Seção 2.3) no sentido amplo, com média nula, a Equação 1 pode ser simplificada. O resultado é mostrado na Equação 2.

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

$$E_{seg}(dB) = 10 \cdot \log(E_{seg}).$$

A Energia é uma medida utilizada principalmente na distinção entre sons vozeados e sons não vozeados, ou de modo mais geral, sons sonoros e sons surdos, haja vista que sons sonoros apresentam energia significativamente maior que sons surdos (RABINER; SCHAFER, 1978). É possível até mesmo distinguir entre trechos vozeados, como no caso da pronúncia de sílabas tônicas. Por exemplo, na Figura 4, destaca-se a diferença entre amplitudes em regiões diferentes da pronúncia da palavra *aplausos*. Pode-se verificar que a região referente à sílaba tônica (*pla*) se faz corresponder aos maiores valores de amplitude, fato decorrente da força necessária para pronunciá-la. Sendo assim, os quadros próximos a essa região contêm os valores de energia mais elevados.

Figura 4 - Forma de onda de um sinal de voz pronunciando a palavra *aplausos*



Os gráficos exibidos nas Figuras 5 a 7 foram construídos a partir da extração da Energia de todos os quadros de 20 ms[11] (com sobreposição

[11] Equivale a 1.000 amostras em sinais de vozes normais, pelo fato de estes terem sido amostrados a 50 kHz. Os sinais de vozes patológicas, por sua vez, foram amostrados, em sua maioria, a 25 kHz, de modo que este intervalo de tempo equivale a 500 amostras. Mais detalhes da base de dados utilizada estão disponíveis na Seção 6.1.

de 50%) de todos os arquivos que contêm sinais de vozes normais, com Edema e com Paralisia utilizados nesta pesquisa (na elocução da vogal */ah/* sustentada). É possível perceber que não há uma distinção clara entre as energias de sinais de vozes pertencentes a estas classes, embora os menores valores sejam sempre associados a vozes com alguma patologia e os maiores, a vozes normais.

Figura 5 - Contraste dos valores de Energia entre vozes Normais e com Paralisia

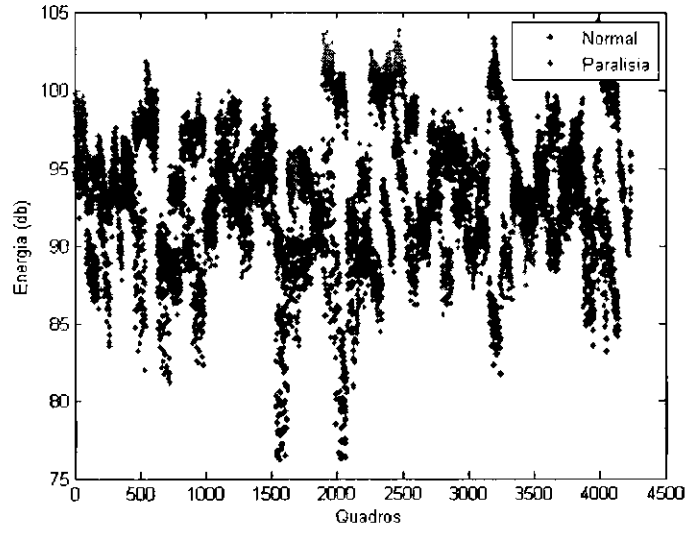
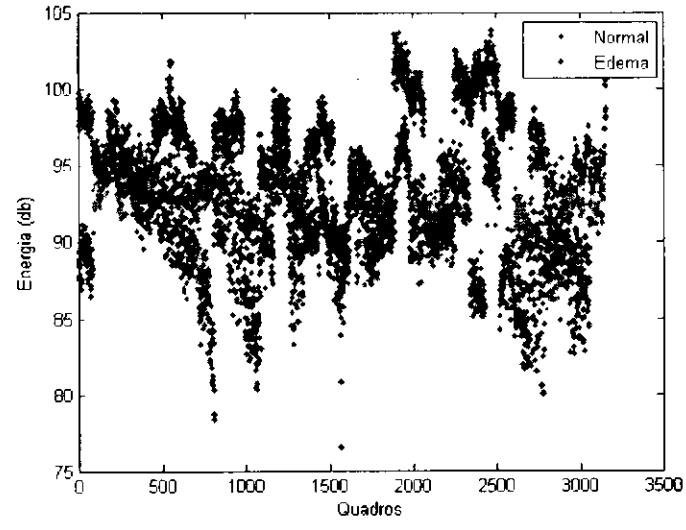


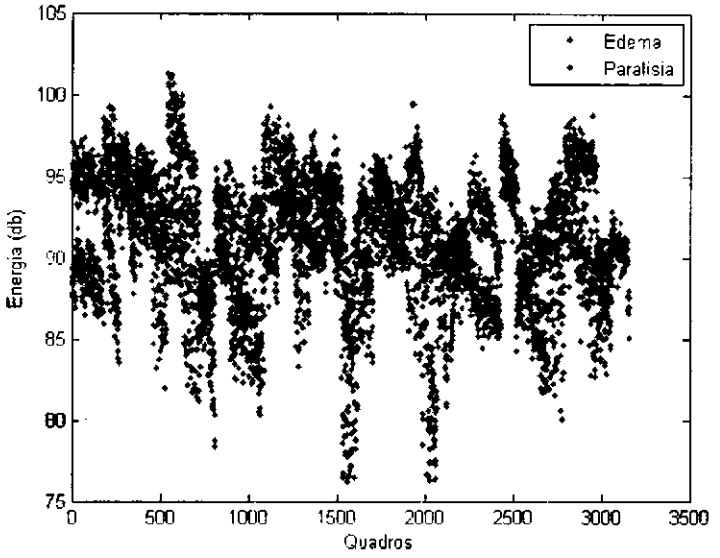
Figura 6 - Contraste dos valores de Energia entre vozes Normais e com Edema



Pela Figura 7, é possível perceber que os efeitos de ambas as patologias relacionadas no indicador Energia são muito semelhantes, o

que pode dificultar a discriminação (nas figuras anteriores, há faixas pertencentes a apenas uma das classes).

Figura 7 – Contraste dos valores de Energia entre vozes com Edema e Paralisia



2.3.1.2 Taxa de Cruzamento por Zero

Trata-se de uma medida associada ao número de vezes em que a forma de onda cruza o eixo das abscissas (tempo). No sinal de voz ilustrado na Figura 3, este eixo não é exibido explicitamente, mas é possível observar que as transições sobre o valor de amplitude 0 (representado do lado esquerdo da figura) ocorrem 28 vezes.

Esta também é uma medida utilizada principalmente na distinção entre sons vozeados e sons não vozeados, pelo fato de sons vozeados apresentarem menor Taxa de Cruzamento por Zero (RABINER; SCHAFER, 1978). Ela é obtida a partir das Equações 3 e 4 (FECHINE, 2000).

$$TCZ = N_A \cdot E\{\text{sgn}[s(n)] - \text{sgn}[s(n-1)]\} = \sum_{n=1}^{N_A-1} |\text{sgn}[s(n)] - \text{sgn}[s(n-1)]|, \quad (3)$$

em que

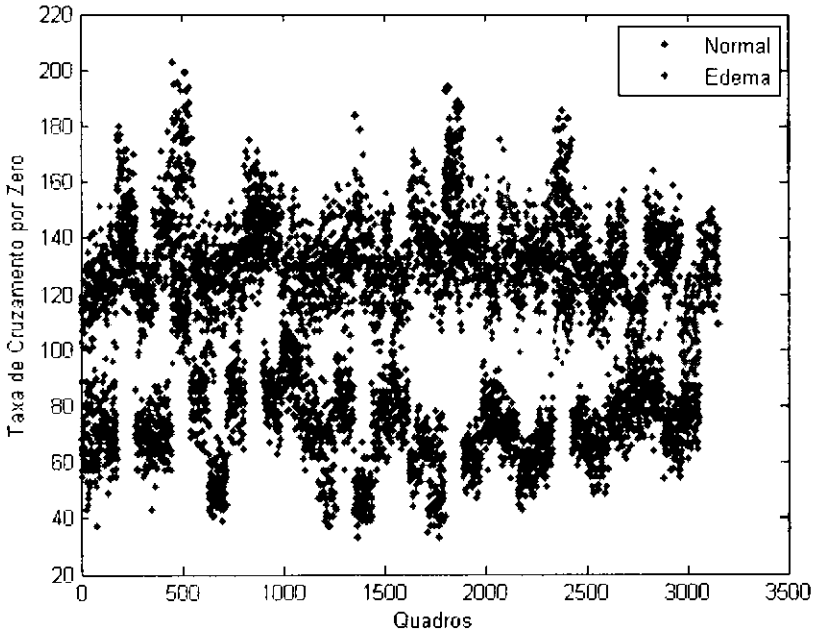
$$\text{sgn}[s(n)] = \begin{cases} 1, & \text{se } s(n) \geq 0 \\ -1, & \text{se } s(n) < 0. \end{cases} \quad (4)$$

Na Equação 3, N_A é o tamanho do quadro analisado e *sgn[s(n)]*, conforme definição na Equação 4, é uma função composta, em que *s(n)*

representa a amplitude na n-ésima amostra de um quadro de um sinal de voz.

Nas Figuras 8 a 10, são exibidos gráficos construídos a partir da obtenção dos valores de TCZ de todos os segmentos de todos os sinais de vozes normais e apresentando Edema e Paralisia utilizados nesta pesquisa, semelhantemente aos gráficos exibidos nas Figuras 5 a 7 (as considerações sobre o tamanho dos quadros, da sobreposição e a elocução do sinal também são válidas). No gráfico da Figura 8, que confronta os valores de TCZ entre sinais de vozes normais e apresentando Edema, é possível perceber que esta medida apresenta boa diferenciação entre estes diagnósticos, sendo possível identificar diversas faixas nas quais é possível afirmar que um valor pertencente a ela é proveniente de determinada classe de arquivos (principalmente no caso de Edema entre 40 e 90).

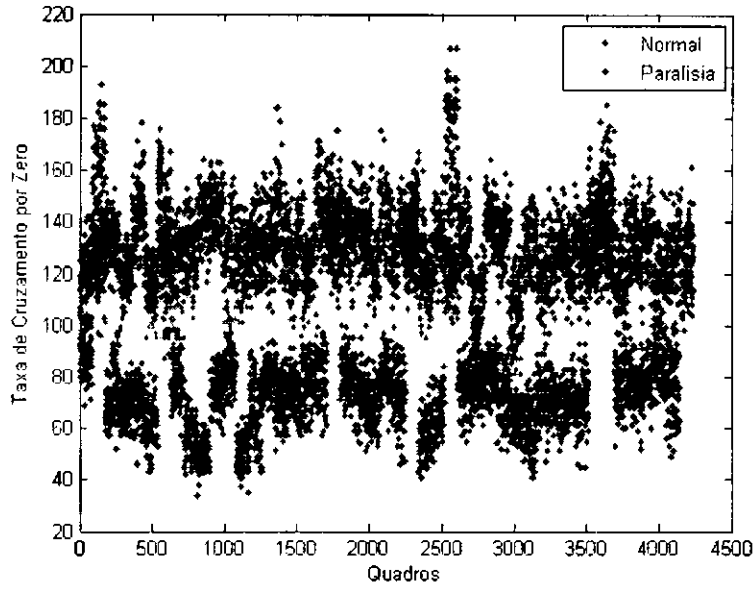
Figura 8 - Contraste dos valores de Taxa de Cruzamento por Zero entre vozes normais e com Edema



Cenário parecido pode ser percebido ao ser examinado o gráfico da Figura 9, mas dessa vez no confronto entre sinais de vozes normais e apresentando Paralisia. Embora haja faixas em que é possível afirmar que determinado valor é proveniente de sinais de voz com Paralisia (entre 40

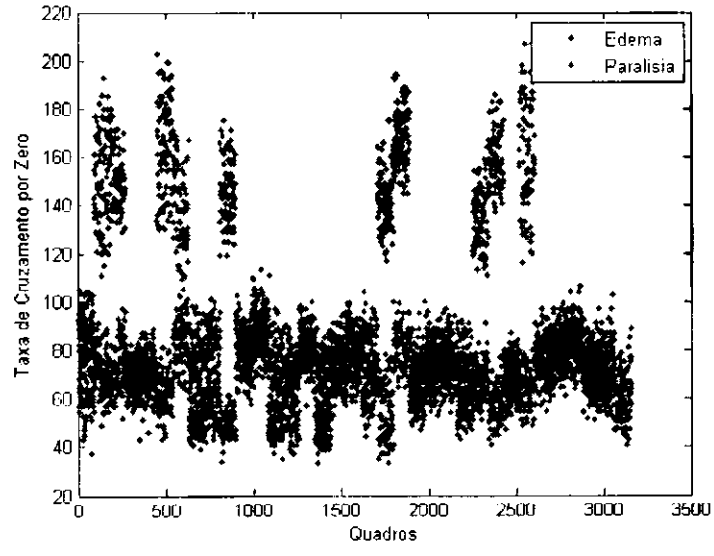
e 90), o mesmo não pode ser encontrado quando tratando de vozes normais.

Figura 9 - Contraste dos valores de Taxa de Cruzamento por Zero entre vozes normais e com Paralisia



A diferenciação deixa de existir quando confrontando sinais de vozes patológicas (Edema e Paralisia), como é possível perceber ao ser examinado o gráfico da Figura 10. Há completa mistura na faixa mais baixa (entre 40 e 100) e na mais alta (120 e 180).

Figura 10 - Contraste dos valores de Taxa de Cruzamento por Zero entre vozes com Edema e com Paralisia

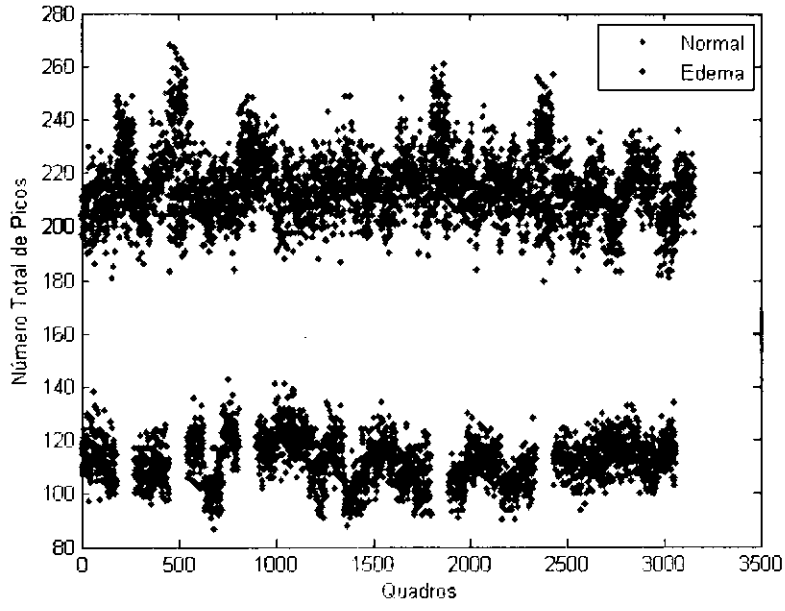


2.3.1.3 Número Total de Picos

A forma de onda exibida na Figura 3 apresenta diversos picos, tanto positivos quanto negativos. O ***Número Total de Picos*** (NTP) é uma medida que visa a calcular a quantidade de picos - positivos e negativos - existente em um dado número de quadros. Há 37 picos na forma de onda apresentada na Figura 3.

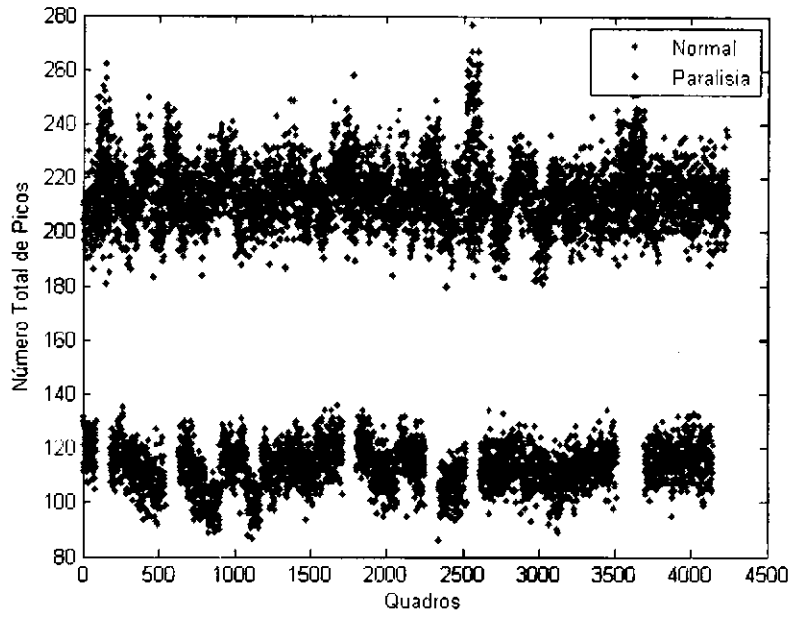
Nos gráficos exibidos nas Figuras 11 a 13 (construídos de forma semelhante aos anteriores), percebe-se o seguinte cenário: na faixa entre 85 e 140, é possível afirmar que os valores que se encontram nela são provenientes de sinais de vozes patológicas, mas não há essa faixa para sinais de vozes normais.

Figura 11 - Contraste dos valores de Número Total de Picos entre vozes normais e com Edema



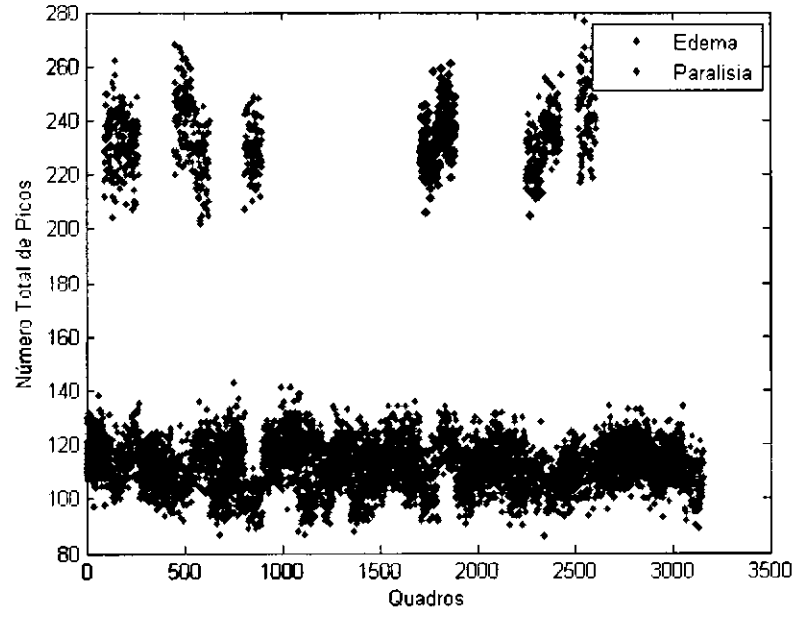
Pela Figura 12, é possível perceber que a extensão dos intervalos relacionados à classe de voz patológica em ambas as figuras são semelhantes e que um intervalo relacionado apenas a vozes normais não é possível de ser identificado.

Figura 12 - Contraste dos valores de Número Total de Picos entre vozes normais e com Paralisia



A diferenciação novamente deixa de existir quando confrontando apenas sinais de vozes patológicas, conforme pode ser visto no gráfico exibido na Figura 13. Tanto no intervalo mais baixo (entre 85 e 140) quanto no mais alto (entre 200 e 270).

Figura 13 - Contraste dos valores de Número Total de Picos entre vozes normais e com Edema

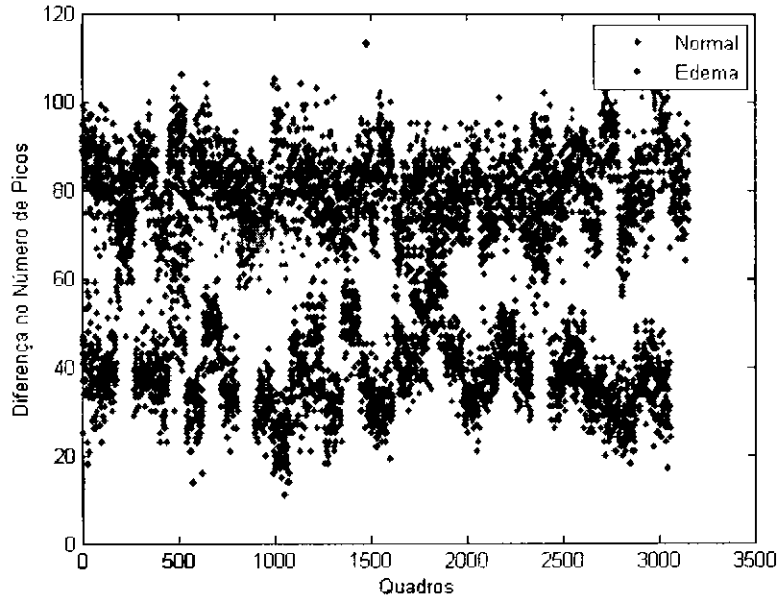


2.3.1.4 Diferença no Número de Picos

A diferença entre o cálculo desta medida e o da medida anterior é simplesmente o sinal empregado na operação: enquanto no NTP somam-se as quantidades de picos, na ***Diferença no Número de Picos*** (DNP) subtraem-se dos picos positivos os picos negativos. Esta medida é utilizada principalmente na distinção entre sons fricativos[12] e vogais de pequena intensidade. Na forma de onda da Figura 3, há 18 picos positivos e 19 picos negativos, de modo que sua DNP é -1.

Com relação aos gráficos de contraste, o cenário encontrado é um pouco diferente do encontrado quando se tratando de NTP. Quando confrontando sinais de vozes normais com sinais de vozes apresentando Edema, é possível identificar faixas pertencentes a ambas as classes: para normais, acima de 90; para Edema, entre 20 e 50. Muito embora a faixa de interseção seja considerável (entre 50 e 80). Esses dados estão exibidos na Figura 14.

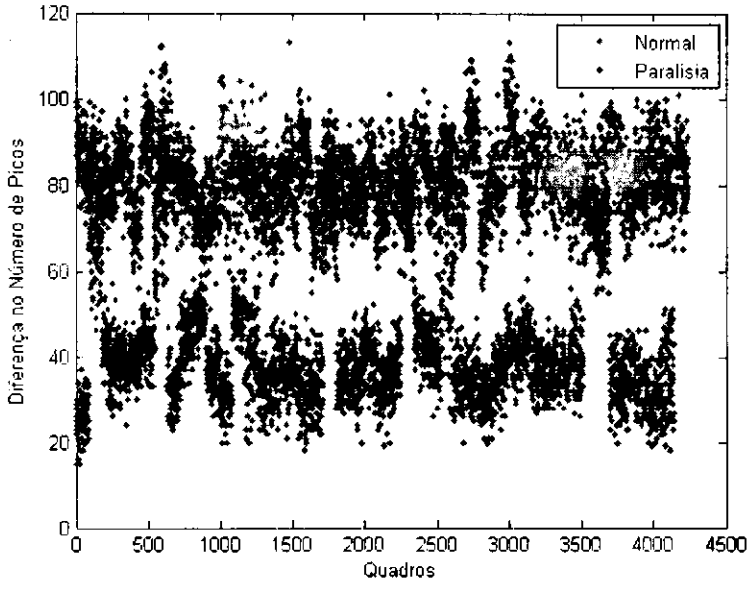
Figura 14 - Contraste dos valores de Diferença no Número de Picos entre vozes normais e com Edema



[12] Formados ao constringir algum ponto do trato vocal e forçar a passagem do ar por essa constrição. São caracterizados por serem ruidosos, ao invés de periódicos. Exemplos de sons fricativos: /j/, /z/ e /f/.

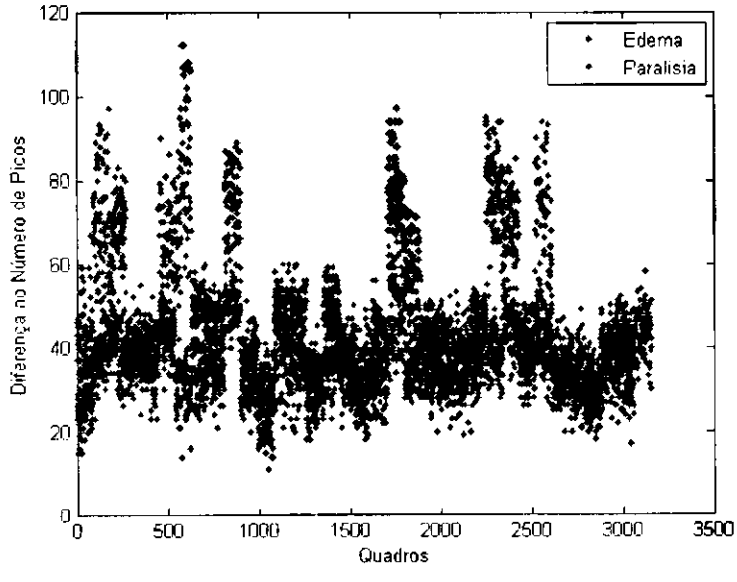
Quando a patologia confrontada é Paralisia, apenas a faixa pertencente a Paralisia é possível de ser identificada (entre 20 e 50). Acima dela, há sempre mistura entre ambas as classes. Isso está ilustrado na Figura 15.

Figura 15 - Contraste dos valores de Diferença no Número de Picos entre vozes normais e com Paralisia



A diferenciação praticamente deixa de existir quando contrastando sinais de vozes patológicas, conforme exibido na Figura 16.

Figura 16 - Contraste dos valores de Diferença no Número de Picos entre vozes com Edema e com Paralisia



2.3.2 Análise Acústica de Sinais de Voz

Na categoria de análise acústica, são utilizadas 5 medidas: Frequência Fundamental, *Jitter*, *Shimmer*, Relação Harmônico-Ruído e Codificação por Predição Linear.

2.3.2.1 Frequência Fundamental (F_0)

A audição é a capacidade de captar e traduzir informações dos sons existentes no entorno do sistema auditivo. O som é decorrente de vibrações mecânicas que se propagam por meio da interação com um meio físico líquido (e.g., sistema auditivo submerso em água), sólido (e.g., sistema auditivo em contato com o solo) ou gasoso (e.g., sistema auditivo em contato com o ar que se respira).

Ao ser recebido pelo ouvido, o som é encaminhado para o interior do canal auditivo, no qual se localiza uma fina membrana denominada tímpano, sensível a pequenas variações de pressão. Em seguida, as vibrações do tímpano são transmitidas a uma tríade de ossos denominados de martelo, bigorna e estribo. Durante este processo, as vibrações são ampliadas, permitindo que o ouvido perceba sons de intensidades muito baixas.

Se há um padrão nas vibrações sonoras, diz-se que o som tem uma forma de onda *periódica*. Se não há nenhum padrão, o som é classificado como ruído. À repetição de uma forma de onda periódica dá-se o nome de *ciclo*. O número de ciclos por segundo que ocorre na transmissão de um som representa a ***Frequência Fundamental*** desse som (ROADS, 1995).

A *Frequência Fundamental* é uma medida empregada em muitos contextos, um dos quais é a escala cromática de notas musicais, que contém notas mais graves (e.g., as primeiras oitavas de um piano ou teclado), correspondentes a valores de Frequência Fundamental mais baixos e notas mais agudas (e.g., as últimas oitavas), correspondentes a valores de Frequência Fundamental mais elevados.

A notação F_0, correspondente a esta medida, decorre do fato de um som apresentar infinitos harmônicos e a Frequência Fundamental ser o

primeiro deles, o mais forte, aquele que o identifica. Os demais harmônicos basicamente contribuem para a composição do som resultante.

No contexto de sinais de voz, F_0 representa a frequência na qual as dobras vocais vibram (abrem-se e fecham-se a cada ciclo) e influi diretamente na tonalidade da voz. Homens apresentam valores de Frequência Fundamental mais baixos (entre 100 e 137 Hz) e mulheres, mais altos (entre 177 e 244 Hz), enquanto crianças apresentam valores ainda mais elevados (entre 206 e 281 Hz).

Esta medida também é empregada para distinguir indivíduos do mesmo sexo, já que há homens com voz mais grave (baixos e barítonos) do que outros (tenores) e, portanto, que apresentam valores mais baixos de Frequência Fundamental (TEIXEIRA; FERREIRA; CARNEIRO, 2011).

Porém, esses valores ocorrem apenas em vozes saudáveis. Conforme discutidas na Seção 2.2, as patologias da laringe afetam o padrão vibratório das dobras vocais, reduzindo sua velocidade de vibração e, consequentemente, o valor da Frequência Fundamental da voz. Por exemplo, Bennett, Bishop e Lumpkin (2011) realizaram uma pesquisa com um grupo de homens e mulheres que apresentavam Edema e percebeu que as mulheres tinham uma frequência fundamental média de 108 Hz, enquanto o grupo de homens com a mesma patologia apresentou Frequência Fundamental média de 91 Hz. Por outro lado, Scalassara (2009a), Costa (2008), Andrade Sobrinho (2011), Marinus (2010) e Stemple, Glase e Klaben (2010) relataram que as patologias Nódulo, Edema, Paralisia e Cisto afetam o valor da Frequência Fundamental, sempre rebaixando-o, i.e., os pacientes acometidos por estas patologias apresentam *pitch* rebaixado e dificuldade de alcançar notas agudas, devido a mudanças na velocidade de vibração nas dobras vocais.

Porém, conforme mostrado nos gráficos das Figuras 17 a 22, nos quais são mostrados os valores de Frequência Fundamental confrontados a cada duas classes de arquivos e separados por gênero, essa distinção, quando não nula, é insignificante. O diferencial destes gráficos com

relação aos anteriores está na distinção de gênero, necessária pelo fato de haver grande distância entre as faixas de F_0 relacionadas a crianças, mulheres adultas e homens adultos. A forma de construção dos gráficos foi a mesma que a dos gráficos exibidos anteriormente: todos os segmentos de todos os sinais utilizados, segmentos de 20 ms e sobreposição de 50%. As diferenças nos números de quadros das figuras se deve à desproporção que há na base entre sinais de vozes masculinas e femininas.

Figura 17 - Contraste dos valores de Frequência Fundamental entre vozes masculinas Normais e com Edema

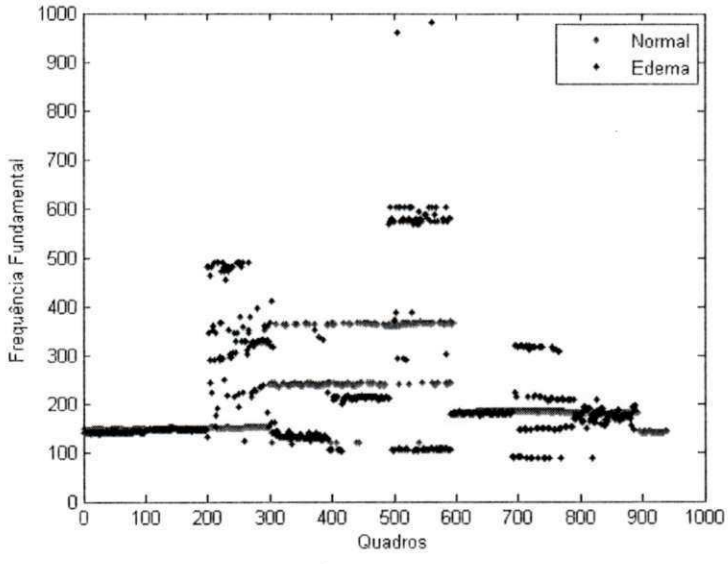
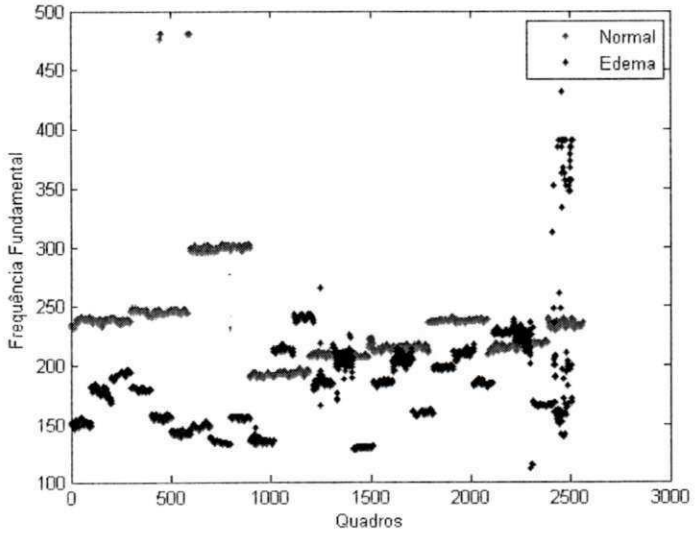
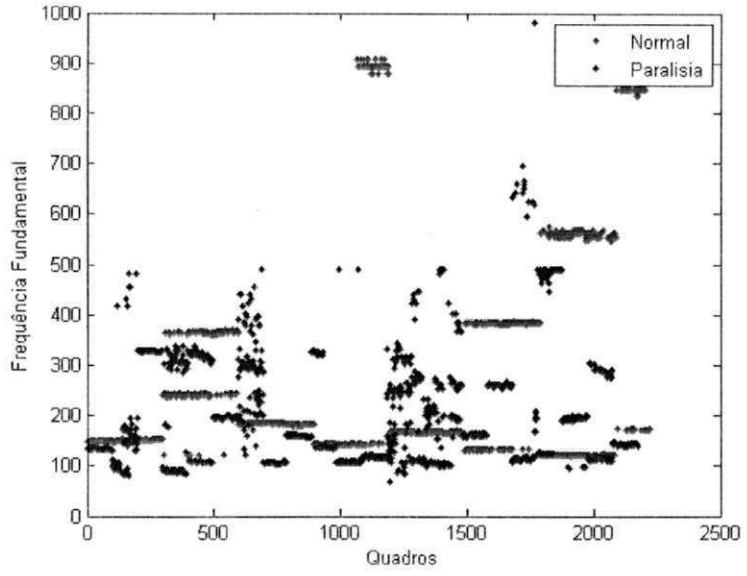


Figura 18 - Contraste dos valores de Frequência Fundamental entre vozes femininas Normais e com Edema



Na Figura 19, é possível identificar valores pontuais que podem ser atribuídos a apenas uma classe, muito embora o intervalo até 500 seja bastante misturado.

Figura 19 - Contraste dos valores de Frequência Fundamental entre vozes masculinas Normais e com Paralisia



Na Figuras 20, o intervalo em que é encontrada vai até 240.

Figura 20 - Contraste dos valores de Frequência Fundamental entre vozes femininas Normais e com Paralisia

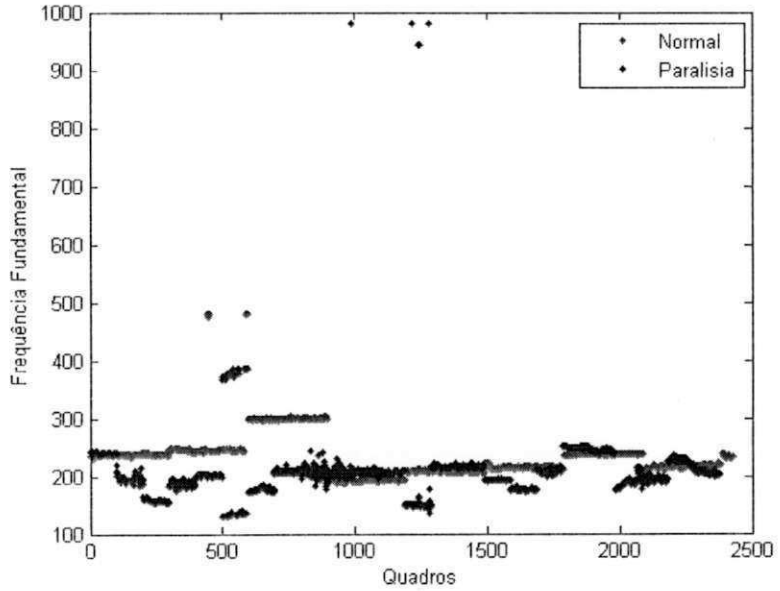


Figura 21 - Contraste dos valores de Frequência Fundamental entre vozes masculinas com Edema e com Paralisia

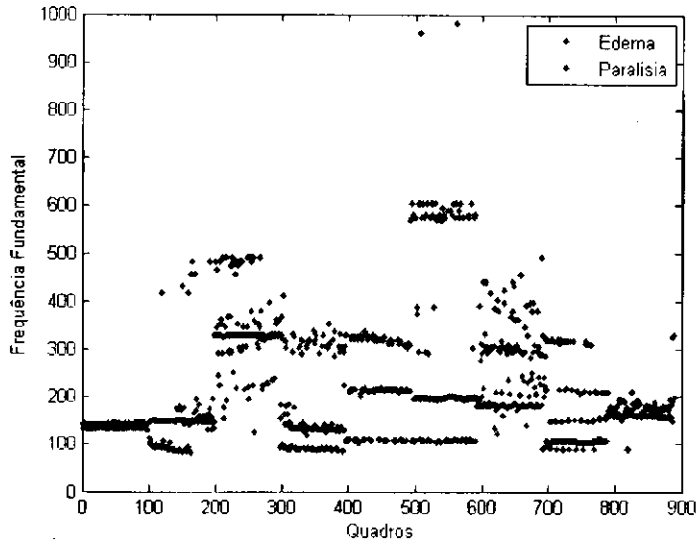
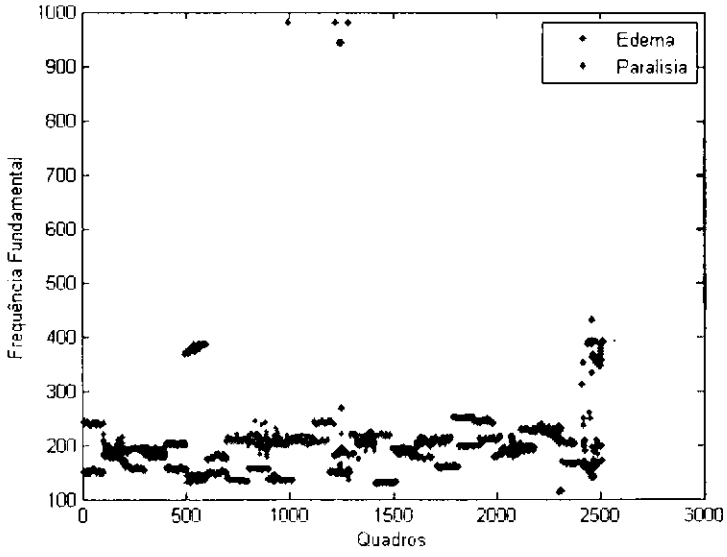


Figura 22 - Contraste dos valores de Frequência Fundamental entre vozes femininas com Edema e com Paralisia



Ao analisar as Figuras 17 a 22, percebe-se que a distinção entre classes de arquivos pela Frequência Fundamental é muito deficiente, sendo necessário o auxílio de outras medidas.

2.3.2.2 *Jitter*

O ***Jitter***, no geral, consiste na perturbação dos valores de período fundamental (T_0) do sinal. Sendo assim, para obtê-lo, devem ser extraídos

primeiramente os valores de T_0 (inverso de F_0) de pequenos segmentos do sinal e em seguida verificar quanto difere cada valor de seus vizinhos. O *Jitter* consiste no somatório dessas diferenças, sendo útil para verificar a estabilidade do sistema fonador, mais especificamente da vibração das dobras vocais, reduzida na presença de patologias (TEIXEIRA; FERREIRA; CARNEIRO, 2011).

Há 4 tipos de *Jitter* comumente usados. O primeiro é denominado *jita* e representa a diferença média absoluta entre dois períodos consecutivos. O segundo, *jitt*, é calculado a partir da razão do *jita* pela média dos períodos fundamentais do sinal. O terceiro, *rap*, representa a perturbação de um valor de período em relação à média da soma deste com seus dois vizinhos mais próximos. Por fim, o *ppq5* é semelhante ao *rap*, mas considera quatro vizinhos (dois anteriores e dois posteriores). As Equações 5 a 8 (TEIXEIRA; FERREIRA; CARNEIRO, 2011) representam as formas de obtenção destas medidas, em que N representa o número de segmentos do sinal e T_i o valor de período fundamental do i-ésimo quadro.

$$jitta = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|. \tag{5}$$

$$jitt = \frac{\frac{1}{N-1}\sum_{i-1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N}T_i}. \tag{6}$$

$$rap = \frac{\frac{1}{N-1}\sum_{i-1}^{N-1}|T_i - (\frac{1}{3}\sum_{n=i-1}^{i+1}T_n)|}{\frac{1}{N}\sum_{i=1}^{N}T_i}. \tag{7}$$

$$ppq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left| T_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i}. \quad (8)$$

Segundo Scalassara (2009a), a patologia Nódulo acarreta aumento dos valores de *Jitter*. Costa (2008), por sua vez, afirmou que, além dos Nódulos, o Pólipo e a Paralisia também elevam esses valores. De acordo com Isshiki (1980), a rouquidão é resultante da vibração aperiódica (irregular) das dobras vocais. Com base na definição de *Jitter* apresentada nesta seção, pode-se afirmar que a diminuição da periodicidade implica em aumento no *Jitter*. Costa (2008) e Andrade Sobrinho (2011) relataram que a patologia Edema acarreta rouquidão do paciente. Sendo assim, é razoável afirmar que esta patologia também implica o aumento do *Jitter*. Porém, nos gráficos exibidos nas Figuras 23 a 25, nas quais se observam valores da medida *jitt*, é possível perceber que o aumento do *Jitter* não acontece em todos os casos. Boa parte dos valores obtidos, em todos os casos, se encontra na base da figura (em torno de 0). A construção dos gráficos se deu de forma semelhante aos anteriores: segmentos de 20 ms, sobreposição de 50%, sinais de vozes na elocução da vogal */ah/* sustentada.

Figura 23 - Contraste dos valores de *Jitt* entre vozes Normais e com Paralisia

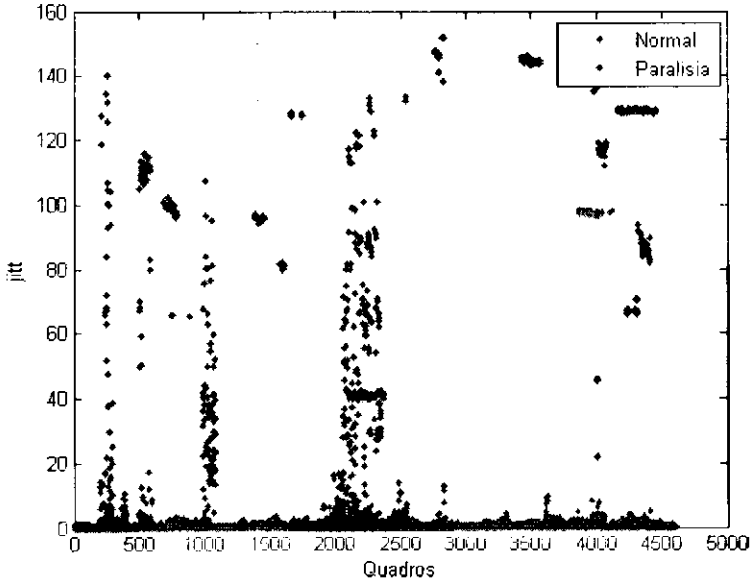


Figura 24 - Contraste dos valores de *Jitt* entre vozes Normais e com Edema

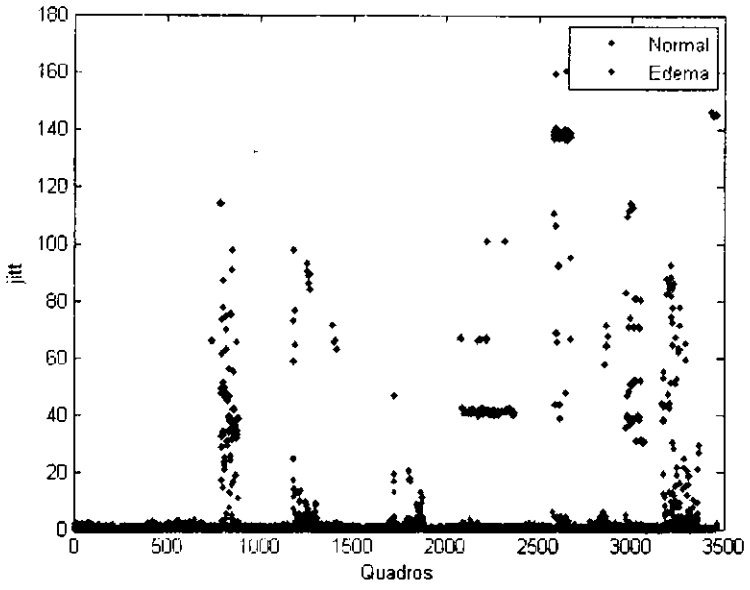
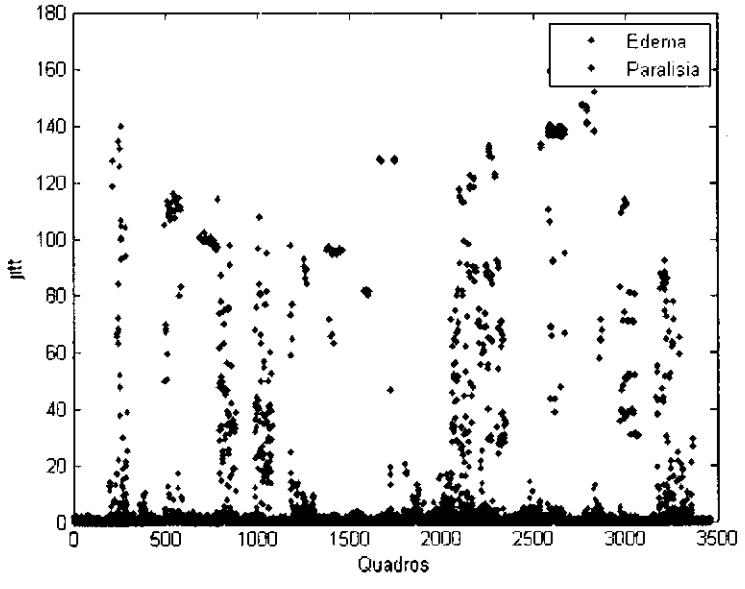


Figura 25 - Contraste dos valores de *Jitt* entre vozes com Edema e com Paralisia



2.3.2.3 *Shimmer*

O ***Shimmer*** é uma medida similar ao *Jitter*. No entanto, é empregada na análise da perturbação existente nos valores de amplitude dos picos do sinal. Por conseguinte, a forma de obter estes valores para o cálculo também é similar: a partir de segmentos do sinal, é obtida a distância entre os dois picos desse segmento (o mais alto e o mais baixo). Esta medida é útil para verificar a estabilidade da intensidade vocal, afetada

pela pressão subglótica e, por sua vez, influenciada pela amplitude de vibração e pela tensão das dobras vocais (FARRÚS; HERNANDO, 2008).

Semelhante ao *Jitter*, há vários tipos de *Shimmer*: o *Shim*, equivalente ao *Jitt*, mas é baseado em amplitudes; o *ShdB*, baseado na diferença proporcional entre dois valores consecutivos de amplitude, em decibéis (dB). Para o cálculo das outras medidas, com denominações sempre iniciadas por APQ, são consideradas várias amplitudes adjacentes. Foram encontradas várias quantidades de vizinhos consideradas: 3, 5, 11 e 55. Nas Equações 9 a 11 (TEIXEIRA; FERREIRA; CARNEIRO, 2011) são expressas algumas destas medidas.

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} | A_i - A_{i+1} |}{\frac{1}{N}\sum_{i=1}^{N} A_i}. \quad (9)$$

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1} | 20 \cdot \log(A_{i+1}/A_i) |. \quad (10)$$

$$apq3 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} | A_i - (\frac{1}{3}\sum_{n=i-2}^{i+2} A_n) |}{\frac{1}{N}\sum_{i=1}^{N} A_i}. \quad (11)$$

Nas Equações 9 a 11, N representa a quantidade de segmentos do sinal sendo analisado e A_i o valor da amplitude do i-ésimo segmento.

Scalassara (2009a) e Costa (2008) afirmaram que as patologias Nódulo, Pólipo e Paralisia nas dobras vocais implicam o aumento do *Jitter* e do *Shimmer*. Andrade Sobrinho (2011), por sua vez, afirmou que o Edema acarreta somente o aumento do *Shimmer*. Na presente pesquisa, verificou-se, a partir de gráficos como aqueles exibidos nas Figuras 26 a 28, nos quais são ilustrados valores de *ShdB* extraídos por arquivo, que o

aumento do *Shimmer* em sinais de vozes patológicas acontece, embora timidamente (principalmente para Paralisia).

Figura 26 - Contraste dos valores de *ShdB* entre vozes Normais e com Paralisia

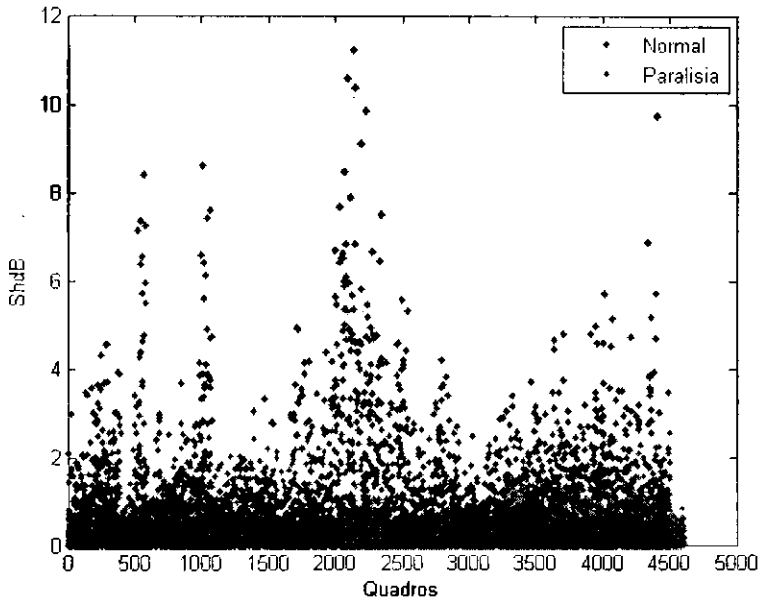
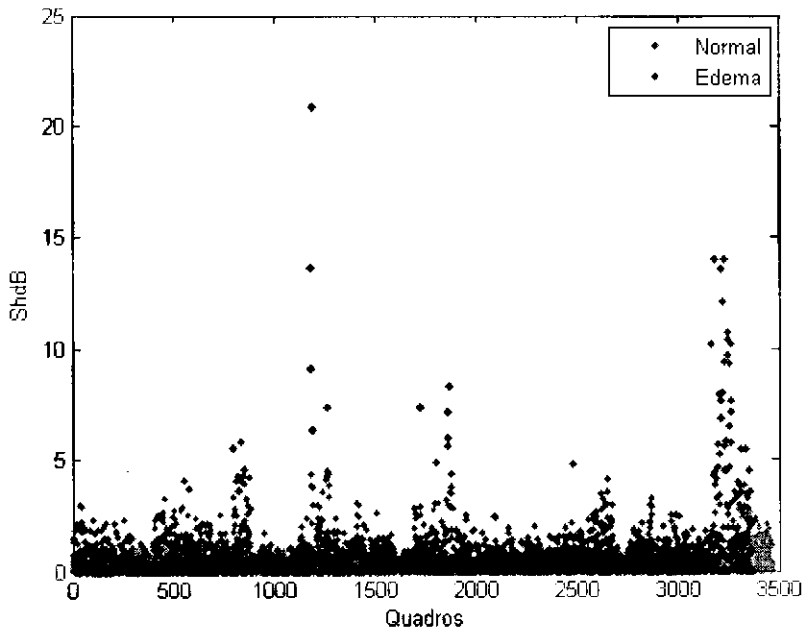
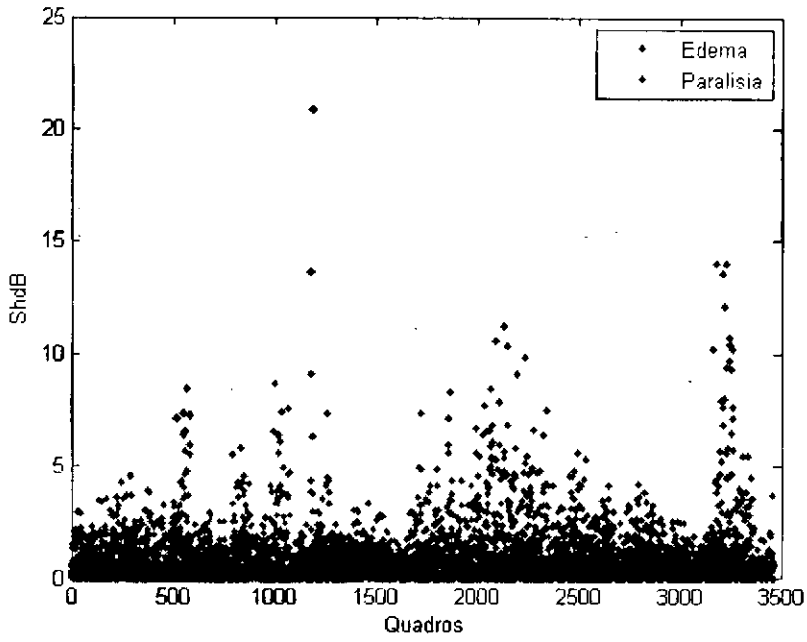


Figura 27 - Contraste dos valores de *ShdB* entre vozes Normais e com Edema



Como esperado, quando confrontando sinais de vozes patológicas (Figura 28), a distinção é praticamente nula.

Figura 28 - Contraste dos valores de *ShdB* entre vozes com Edema e com Paralisia



2.3.2.4 Relação Harmônico-Ruído

A ***Relação Harmônico-Ruído*** (*Harmonic-to-Noise Ratio* - HNR) possibilita a avaliação matemática da quantidade de ruído presente em um sinal, em comparação a sua componente periódica. É uma medida adotada em outros contextos, tal como redes sem fio, mas também se mostra adequada ao contexto do processamento digital de sinais de voz, pelo fato de a componente periódica ser decorrente da vibração das dobras e a aperiódica, do ruído glótico (LOPES et al., 2008).

A HNR, neste contexto, mede a eficiência do processo de fonação: quanto maior a eficiência da utilização do fluxo de ar expelido dos pulmões, como energia para a vibração das dobras vocais, e quanto mais completo o seu ciclo vibratório, maior é a componente periódica em relação à componente aperiódica, o que implica em maior valor de HNR (LOPES et al., 2008). Lopes et al. (2008) apresentam mais detalhes sobre a HNR.

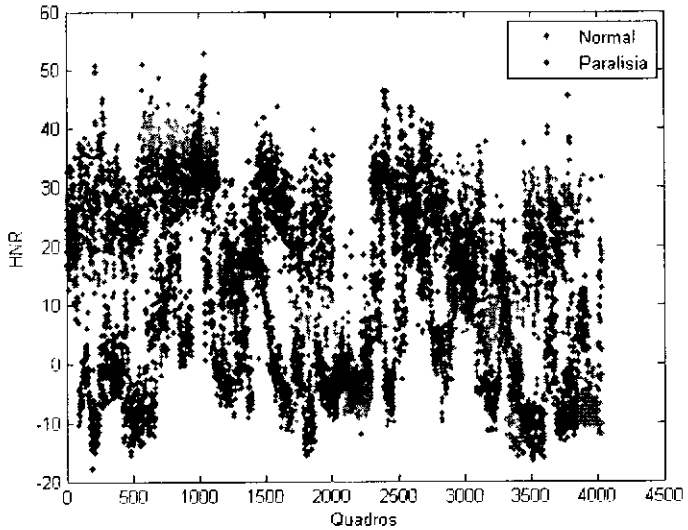
As patologias, conforme discutidas na Seção 2.2, interferem diretamente no padrão vibratório das dobras vocais, deixando a voz com um aspecto ruidoso, de modo que uma baixa HNR representa um forte

indício da presença de uma patologia. Parsa e Jamieson (2000) afirmaram que 83% das vozes patológicas apresentam baixa HNR e que quanto mais ruído um sinal de voz apresenta, mais avançado é o estágio da patologia, de modo que esta medida visa não somente a detectar a presença de uma patologia, como também o estágio em que ela se encontra.

Segundo Scalassara (2009a), Costa (2008), Andrade Sobrinho (2011), Stemple, Glase e Klaben (2010) e Marinus (2010), as patologias Nódulos, Pólipo, Cisto e Paralisia deixam a voz do paciente com ruído de fundo, o que implica dizer que reduzem o valor da HNR da elocução. Esta afirmação já tinha sido destacada por Yumoto, Gould e Baer (1982), que mostraram que a HNR de vozes normais é quase sempre maior do que aquela associada a vozes patológicas (não foi feita associação a patologias específicas) antes de ser feita alguma cirurgia, embora haja alguns casos em que isto não ocorre. Ao mesmo tempo, em vozes de indivíduos submetidos a uma cirurgia, com o intuito de tratar a patologia presente, essa distinção é praticamente impossível de ser verificada (os valores de HNR deste grupo quase sempre estão dentro da faixa de HNR de vozes normais).

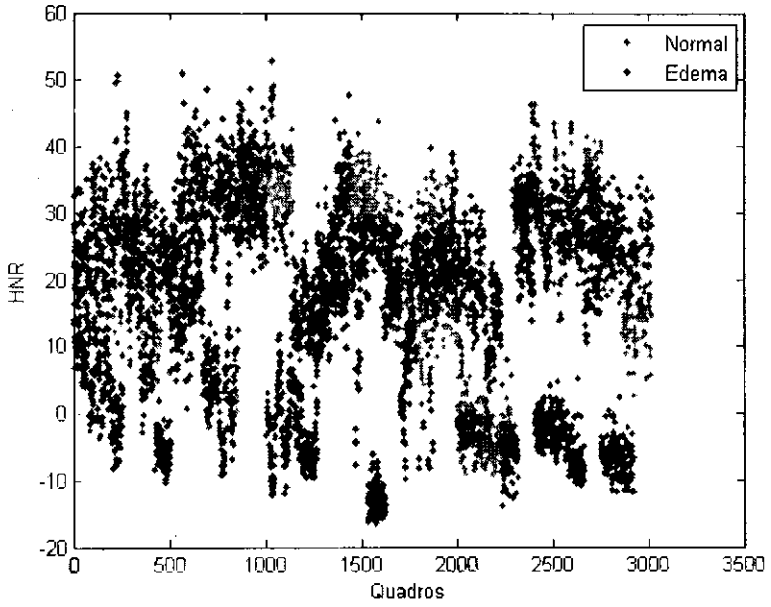
Nos gráficos exibidos nas Figuras 29 a 31, é possível observar uma separação mais clara apenas entre vozes normais e vozes apresentando Paralisia (Figura 29).

Figura 29 - Contraste dos valores de *HNR* entre vozes Normais e com Paralisia



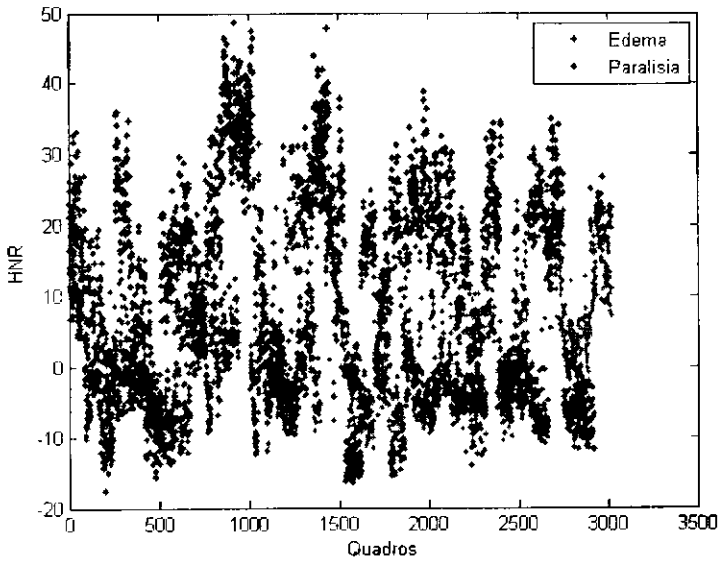
Quando se trata de Edema (Figura 30), observa-se muita interseção entre os valores de HNR.

Figura 30 - Contraste dos valores de *HNR* entre vozes Normais e com Edema



A distinção se torna praticamente impossível quando confrontando classes de arquivos de sinais de vozes patológicas, como mostrado na Figura 31.

Figura 31 - Contraste dos valores de *HNR* entre vozes com Edema e com Paralisia



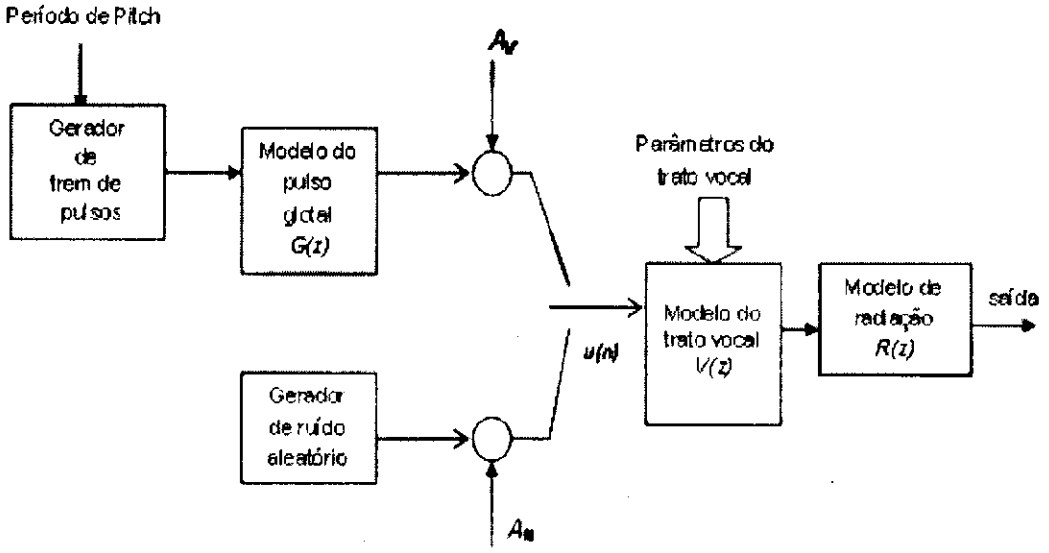
2.3.2.5 Análise por Predição Linear LPC

A ***Análise por Predição Linear*** é considerada uma das técnicas mais poderosas de análise da fala (RABINER; SCHAFER, 1978). Por meio dela, é possível estimar diversos parâmetros relacionados à voz, dentre os quais o *pitch* e frequências formantes, de modo preciso e relativamente rápido. A ideia consiste em estimar uma amostra de voz a partir de uma combinação linear de amostras passadas. Assim, o sinal como um todo pode ser reconstruído a partir da estimação das amostras. Essa estimativa deve ser feita de modo a minimizar a soma das diferenças quadradas entre as amostras reais e amostras preditas, o que permite encontrar um conjunto único de p coeficientes preditores. Quanto maior o valor de p, mais precisas serão as estimativas e, consequentemente, mais íntegro o sinal reconstruído.

Portanto, é possível representar todo o sinal apenas com esses coeficientes, o que caracteriza um processo de compressão (neste caso, de sinais de voz), já que é obtido um fluxo menor do que o original, a partir do qual pode ser gerado o sinal completo. Como não é possível o sinal reconstruído ser idêntico ao original, este pode ser considerado um processo de compressão com perdas. É comum seu uso na transmissão da fala, com o intuito de possibilitar a transmissão a baixas taxas de bits (RABINER; SCHAFER, 1978).

A Análise por Predição Linear é originária da concepção da produção da fala como um sistema linear variante no tempo, excitado por um trem de pulsos quase periódicos (a periodicidade é determinada pela frequência fundamental do sinal da fala) e ruído aleatório. O trem de pulsos é aplicado a um filtro G(z), que simula o efeito dos pulsos glotais, e em seguida, a um controle de ganho A_V. O ruído aleatório é aplicado a um controle de ganho A_N. Os últimos elementos do modelo, os quais podem receber sons periódicos ou aperiódicos, são o trato vocal e a radiação dos lábios e do próprio trato vocal (RABINER; SCHAFER, 1978). Este modelo é ilustrado na Figura 32.

Figura 32 - Modelo Linear de produção da fala



FONTE: Rabiner e Schafer (1978)

A predição linear provê métodos robustos e precisos para estimar os parâmetros que caracterizam este sistema, dentre eles a Codificação por Predição Linear (*Linear Predictive Coding – LPC*). Nesta pesquisa, a obtenção de coeficientes LPC foi feita a partir do MATLAB® (MathWorks Inc., 2007). É válido ressaltar, que este software utiliza o método de autocorrelação no processo de obtenção de coeficientes LPC. Mais detalhes sobre o cálculo desses coeficientes podem ser encontrados em Costa (2008), Marinus (2010), Fechine (2000) e Rabiner e Schafer (1978).

2.3.3 Análise Estatística de Sinais de Voz

A Estatística é a ciência que visa a colecionar, organizar e interpretar fatos numéricos, denominados *dados*, os quais devem ser analisados para que se tenha compreensão sobre um fenômeno. Qualquer conjunto de dados contém informações sobre um grupo de indivíduos. Essas informações são organizadas em variáveis, que são características dos indivíduos e podem conter diferentes valores para diferentes indivíduos. Pela distribuição de uma variável, é possível saber seus possíveis valores e a probabilidade de ocorrência de cada um. Se os valores forem numéricos (ou puderem ser

associados a números) e representarem um fenômeno aleatório, diz-se que a variável é *aleatória* (MOORE; McCABE, 1998).

A Teoria da Informação, por sua vez, é um campo que mostra como se pode medir a informação contida em uma mensagem. Intuitivamente, ela é proporcional à surpresa ocorrida ao recebê-la. Se o conteúdo da mensagem não causa nenhuma surpresa em quem a recebe, então se considera que a mensagem não continha nenhuma informação. Quanto maior o nível de surpresa ocorrido, maior a quantidade de informação transmitida (SALOMON, 2004).

De um ponto de vista estatístico, a fonte de informação que gerou uma mensagem pode ser considerada uma variável aleatória discreta, com um alfabeto associado, em que cada símbolo deste alfabeto contém uma probabilidade de surgimento. As mensagens originárias desta fonte nada mais são que produtos desta fonte de informação (variável aleatória discreta) e os símbolos tenderão a surgir respeitando as probabilidades associadas. Por exemplo, a mente humana pode ser considerada uma fonte de informação, que gera mensagens de acordo com o vocabulário conhecido pelo indivíduo.

A quantidade de informação contida em uma mensagem resultante desta fonte é chamada de ***Entropia***, que é comumente referida também como uma medida da incerteza desta variável aleatória, sendo expressa a partir da equação:

$$H = -\sum_{i=0}^{N} p_i \log_2(p_i) \,, \tag{12}$$

na qual p_i indica a probabilidade de um evento da distribuição de probabilidades de uma variável aleatória discreta. H atinge valor máximo quando todas as probabilidades são iguais e valor mínimo quando apenas um símbolo é gerado ($p_i = 1$, $\log_2(1) = 0$). O emprego do logaritmo na base 2 indica que a entropia é expressa em bits/símbolo.

A entropia é uma medida fundamental para a Teoria da Informação, sendo empregada como o limiar de compressão máximo possível de ser

obtido na compressão de uma mensagem proveniente de uma fonte de informação[13].

Para a detecção de patologias em sinais de voz, há diversos relatos da sua utilização em Tavares et al. (2011), Scalassara (2009b) e Maciel, Pereira e Stewart (2010). Os autores supracitados descobriram que quanto maior a entropia de um sinal de voz, maior a probabilidade de a voz apresentar uma patologia e obtiveram, com isso, detectores da presença de patologias baseados em entropia.

Um tipo especial de entropia tem sido usado em investigações de classificação: a entropia *condicional*, expressa na Equação 13.

$$F_n(P) = -\sum_{i=1}^{M^n}\sum_{j=1}^{M} P(x_i^{n-1}, a_j)\log_2 P(a_j \mid x_i^{n-1}), \qquad (13)$$

em que M é a quantidade de símbolos do alfabeto da fonte, M^n a quantidade de mensagens diferentes que podem ser geradas com n caracteres, $P(x_i^{n-1}, a_j)$ representa a probabilidade da sequência x_i^{n-1} surgir concatenada com o símbolo a_j e $P(a_j \mid x_i^{n-1})$ representa a probabilidade de surgimento do símbolo a_j dado que antes foi lida a sequência x_i^{n-1}.

2.4 O método de Predição por Casamento Parcial

Tendo sido apresentados os principais indicadores, utilizados em diversas pesquisas relacionadas revisadas (as quais serão apresentadas brevemente na Seção 3.1), nesta seção será apresentado um classificador utilizado em diversos outros contextos de classificação.

O método *PPM* é um método computacional utilizado na compressão de fluxos de dados, tal como o algoritmo Deflate[14], de uso difundido nos dias atuais. Do ponto de vista do poder de compressão, o PPM é um compressor de dados bastante eficaz, conforme o estado da arte na área da compressão sem perdas (SALOMON, 2004; HONÓRIO; BATISTA;

[13] A compressão de uma mensagem de tamanho n nunca ultrapassa nH bits, em que n é o tamanho da mensagem e H é a entropia da fonte geradora (SALOMON, 2004).

[14] Usado para criar arquivos em formato *ZIP*.

DUARTE, 2009; BARUFALDI et al., 2009; MEDEIROS et al., 2011). Seu desempenho, no que se refere à Razão de Compressão e ao tempo de execução, foi comparado empiricamente a outros métodos bastante populares e outras aplicações comerciais de *software* destinadas à compressão de dados. Os resultados estão disponíveis no Apêndice A.

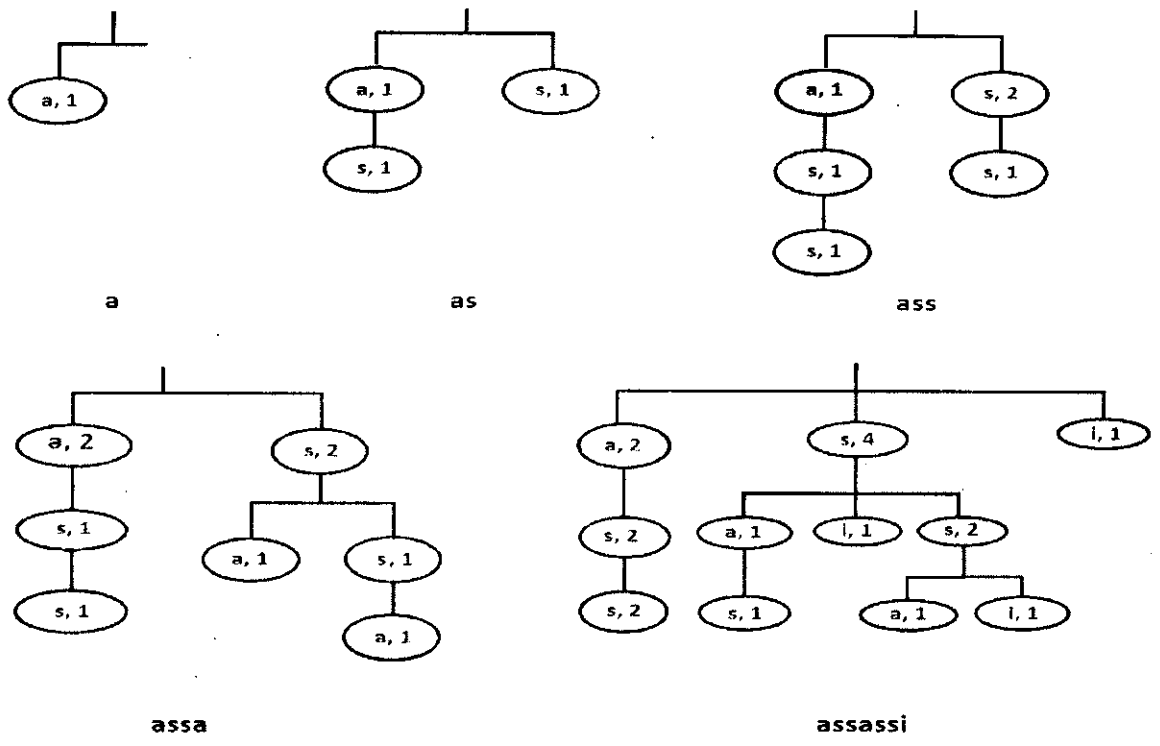
Entretanto, o seu uso em escala comercial ainda é muito limitado, restringindo-se, prioritariamente, ao âmbito da pesquisa acadêmica, pelo fato de ser armazenado, durante sua execução, um modelo muito preciso da fonte de dados sendo comprimida, o que acarreta alto consumo de memória e velocidade de execução relativamente baixa, principalmente se comparado com os compressores mais utilizados.

Métodos estatísticos destinados à compressão de dados, tal como o PPM, podem ter sua operação dividida em duas etapas: *modelagem* e *codificação* (SALOMON, 2004). A *modelagem* consiste no armazenamento das probabilidades de símbolos originários de um fluxo de dados. Tendo em vista que o modelo PPM é contextual, também podem ser armazenadas as probabilidades de sequências dos símbolos. A intenção é estimar, o mais precisamente possível, as probabilidades da fonte de dados estacionária abstrata que gerou o fluxo de dados.

A contribuição da proposição do algoritmo, feita por Cleary e Witten (1984), foi restrita ao campo teórico. Os autores se mostraram pessimistas com a implementação computacional do método, tendo relatado terem sido necessários cerca de 50 ms por caractere. Todavia, com base na proposição de Cleary e Witten (1984), diversas propostas foram formuladas, tendo sido a primeira implementação prática denominada *PPM-C*, tida por Moffat (1990) como aquela que obtinha os melhores resultados de compressão.

Na Figura 33, é mostrado um passo a passo da alimentação de um modelo *PPM-C* ao ser lido o trecho *assassi* da palavra *assassinar*.

Figura 33 - Alimentação de um modelo PPM caractere por caractere



O modelo exibido na Figura 33 é chamado *trie*, sugerido por Salomon (2004) para o armazenamento de um modelo *PPM-C*. É possível perceber que no primeiro nível (filhos do nó raiz) são contabilizados os símbolos isoladamente - sem relação com contextos. Pelo fato de a letra *s* ter aparecido 4 (quatro vezes) no fluxo, o contador associado a esse caractere indica 4. O mesmo vale para as letras *a* e *i*, que apareceram 2 e 1 vez, respectivamente.

Nos nós "netos" do nó raiz, por sua vez, estão armazenados e contabilizados os símbolos que sucedem contextos compostos por um único símbolo (o nó ascendente). Na palavra lida, a letra *s* sucede as letras *a* e *s* duas vezes cada, fato indicado nos contadores associados. E além de *s*, outras letras sucedem a letra *s* na palavra, tais como *a* e *i*, 1 (uma) vez cada. A letra *i*, pelo fato de ter aparecido apenas no fim da leitura (na etapa exibida), não é contexto para nenhuma outra letra, razão pela qual não tem nós filhos.

Por fim, os nós do último nível sucedem contextos compostos por dois símbolos (representados pelos dois nós ascendentes). As letras *a* e *i* sucedem uma vez, em diferentes momentos, o dígrafo *ss*, fato representado pelo nós mais à direita no modelo. O contexto *as* sempre é sucedido pela letra *s* (duas vezes), fato representado pelo nó mais à esquerda do modelo.

Caso o tamanho de contexto máximo armazenado fosse maior, essa *trie* seria maior em profundidade - sua largura depende unicamente da quantidade de símbolos do alfabeto considerado no processo.

No Quadro 1, é apresentado o estado de um modelo *PPM-C* após a leitura completa da palavra *assassinar*.

Quadro 1 - Exemplo de um modelo PPM após a leitura da palavra *assassinar*

Ordem k = 2			Ordem k = 1			Ordem k = 0			Ordem k = -1		
Predição	c	p	Predição	c	p	Predição	c	P	Predição	c	p
as → s	2	2/3	a → s	2	2/5	→ a	3	3/15	→ A	1	1/|A|
→ Esc	1	1/3	→ r	1	1/5	→ s	4	4/15			
ss → a	1	1/4	→ Esc	2	2/5	→ i	1	1/15			
→ i	1	1/4	s → s	2	2/7	→ n	1	1/15			
→ Esc	2	2/4	→ a	1	1/7	→ r	1	1/15			
sa → s	1	1/2	→ i	1	1/7	→ Esc	5	5/15			
→ Esc	1	1/2	→ Esc	3	3/7						
si → n	1	1/2	i → n	1	1/2						
→ Esc	1	1/2	→ Esc	1	1/2						
in → a	1	1/2	n → a	1	1/2						
→ Esc	1	1/2	→ Esc	1	1/2						
na → r	1	1/2									
→ Esc	1	1/2									

Este modelo é dividido em quatro colunas. Em cada coluna, estão agrupados (i) contextos diferentes de mesmo tamanho, sendo 2 o tamanho máximo guardado e (ii) os símbolos que os sucedem. A escolha do tamanho máximo de contexto armazenado é livre mas, vale salientar que o consumo de memória cresce exponencialmente com esse valor. Além disto, contextos muito altos guardam muitas informações que aparecem com pouca frequência e, segundo Salomon (2004), a curva de aprendizado do modelo PPM cessa o crescimento a partir de um dado

tamanho de contexto, o que ocasiona redução na taxa de compressão. Na prática, o tamanho máximo de contexto adotado não se distancia muito daquele que é mostrado no Quadro 1, tendo o tamanho 5 apresentado ótimos resultados para textos (TEAHAN, 1997; COUTINHO et al., 2005; BARUFALDI et al. 2009).

No Quadro 1, ao lado de cada símbolo, são exibidos um inteiro e uma fração, os quais representam, respectivamente, a frequência do referido símbolo em determinado contexto no fluxo lido e a probabilidade estimada de tal símbolo surgir nesse contexto. Por exemplo, a probabilidade estimada associada ao aparecimento isolado da letra *s* é 4/15. No contexto *a*, ou seja, antecedida por esta letra, a probabilidade sobe para 2/5 e no contexto *as*, aumenta para 2/3.

A coluna mais à direita representa o que se denomina *ignorância absoluta* ou *contexto -1*. Enquanto os contextos maiores ou iguais a zero representam de fato o que foi observado na leitura do fluxo original[15], o contexto -1 considera que a fonte é de distribuição *uniforme*, atribuindo probabilidades iguais a todos os símbolos. Por esta razão, tal contexto é denominado *ignorância absoluta*. No início do processo de leitura, todas as colunas devem estar vazias, com exceção da coluna referente ao contexto -1, que deve conter todos os símbolos do alfabeto. Sendo assim, um símbolo que ainda não tenha sido lido do fluxo deverá, certamente, encontrar-se nesse contexto, devendo ser retirado quando surgir pela primeira vez no fluxo.

Esse modelo é construído adaptativamente, ou seja, a cada novo símbolo lido os contadores (e, consequentemente, as probabilidades) são atualizados ou novas entradas (sequências) são criadas e armazenadas.

No modelo do Quadro 1 está presente o símbolo *Esc*, o qual não pertence ao conjunto de símbolos do alfabeto. Denominado *símbolo de Escape*, *Esc* é um mecanismo usado pelo algoritmo para sinalizar a substituição para o contexto imediatamente menor e foi criado

[15] São estimativas reais das probabilidades dos símbolos e sequências na fonte.

originalmente para sincronizar a substituição entre contextos nos processos de codificação e decodificação. Porém, como no âmbito do reconhecimento de padrões não há decodificação, ele pode ser interpretado como a representação de todos os outros símbolos do alfabeto considerado não listados em determinado contexto. Por exemplo, sob o contexto *s*, *Esc* (probabilidade 3/7) pode ser interpretado como representante das letras restantes deste alfabeto (*n* e *r*) e sua probabilidade é a probabilidade de um desses símbolos surgir pela primeira vez sob este contexto. Quando todos os símbolos do alfabeto têm suas probabilidades estimadas em determinado contexto (assim como no contexto de tamanho 0), torna-se desnecessário armazenar no modelo os valores relacionados ao escape neste contexto, de modo que eles podem ser descartados. O contador de *Esc* em um contexto, por sua vez, é incrementado a cada novo símbolo surgido neste contexto, de modo que pode sempre representar também a quantidade de símbolos já listados neste contexto.

A segunda etapa de um processo de compressão estatístico é a codificação. Enquanto na modelagem são mantidas as probabilidades de símbolos e sequências de símbolos, com o intuito de identificar os mais frequentes, na codificação são atribuídos códigos a cada símbolo, visando a gerar um fluxo de dados menor que o original[16]. Para tanto, os símbolos e sequências mais frequentes devem receber os menores códigos, com o mínimo de bits possível, enquanto os menos frequentes recebem os maiores. Os codificadores estatísticos funcionam desta forma, a exemplo do código de Huffman (mais detalhes em Salomon (2004)). Os códigos atribuídos têm sempre quantidades não-fracionárias de bits. No entanto, a Teoria da Informação mostra que a função que quantifica informação é o logaritmo, de modo que a quantidade ideal de bits usada para codificar símbolos com base em suas probabilidades, visando a gerar o menor fluxo comprimido possível, deve ser fracionária. Por exemplo, todos os

[16] Se o propósito fosse criptografia, a codificação seria projetada com fins a dificultar a recuperação da mensagem original, independente do tamanho da mensagem final.

primeiros 100 números (0 a 99) podem ser representados com algarismos de log 100 = 2 dígitos. Os primeiros 1000, log 1000 = 3 dígitos. Em outra situação, a quantidade de iterações em uma busca binária depende do tamanho da lista em que se faz a busca. Em uma lista com 100 elementos, são necessárias $\log_2 100 = 6{,}64 \approx 6$ iterações.

Sendo assim, para que seja obtido um código comprimido de tamanho ideal, é necessário que o codificador atribua quantidades fracionárias de bits a cada símbolo. O codificador denominado Aritmético alcança este objetivo ao atribuir um código único (longo) a toda a mensagem, o que acarreta quantidades fracionárias de bits por símbolo do fluxo original. Esse código, na verdade, é um número, com quantidade ilimitada de casas decimais, dentro de um intervalo. No início do processo, o intervalo é [0, 1). À medida que mais probabilidades são recebidas, este intervalo diminui, mas o número de casas decimais de cada limite aumenta. No fim, tem-se um intervalo pequeno, mas com limites contendo várias casas decimais. O fluxo comprimido pode ser qualquer número dentro deste intervalo.

Devido à riqueza de informação deste modelo, são encontrados diversos relatos de uso do PPM visando à classificação de objetos, tais como textos, imagens e sinais de eletrocardiograma. Nestes experimentos, em linhas gerais, é construído um modelo para cada classe a ser considerada na classificação. O treinamento consiste em alimentar os modelos com dados referentes à classe que representam. A classificação, por sua vez, consiste em tornar os modelos estáticos (de modo que os contadores armazenados não são alterados/atualizados), comprimir fluxos de dados não utilizados no treinamento com todos os modelos construídos e obter a Razão de Compressão para cada caso. Considera-se que o fluxo de dados pertence à classe do modelo que gerou o menor fluxo comprimido (com o qual foi obtida a maior Razão de Compressão) (COUTINHO et al., 2005; BARUFALDI et al., 2009; MEDEIROS et al., 2011). Mais detalhes sobre os referidos experimentos e os seus resultados são apresentados na Seção 3.3.

2.5 Discussão

O sistema de produção da fala compreende uma etapa de captação de energia (inspiração de ar pelas narinas) e de interferência de determinados órgãos na passagem do ar, durante a expiração, principalmente a laringe e órgãos que fazem parte dos tratos vocal e nasal, os quais dão à voz as características usuais. Porém, o funcionamento deste sistema pode ser comprometido por diversos tipos de patologias, seja impedindo as dobras vocais de vibrarem (paralisia) ou alterando sua anatomia (patologias estruturais), o que altera a massa das dobras vocais (no caso das patologias estruturais), consequentemente sua velocidade de vibração, e interfere no fechamento da glote.

Seus principais sintomas são rouquidão e elocução vocal com ruído de fundo. Todas as patologias estudadas no curso desta pesquisa incluem um ou ambos destes sintomas, o que pode tornar difícil sua caracterização (distinção em relação às demais). Também é possível encontrar (mas em menor frequência) sintomas tais como diplofonia, fadiga e instabilidade vocal. Elas são causadas principalmente por abuso vocal e consumo excessivo de tabaco e álcool, mas há também as que podem ser adquiridas congenitamente ou como sequelas de cirurgias.

É comum que o surgimento dessas patologias acarrete alterações em características extraídas da voz, algumas das quais são consideradas nesta investigação. Há várias categorias de indicadores do sinal da voz, entre elas as medidas temporais (e.g., Energia e Taxa de Cruzamento por Zero), acústicas (e.g., Frequência Fundamental e Relação Harmônico-Ruído) e estatísticas (e.g., Entropia). Essas características devem ser utilizadas para alimentar um classificador, o qual deve decidir sobre a presença ou ausência de uma patologia, ou até mesmo sobre qual patologia está presente.

Um exemplo de classificador é o PPM, um método estatístico de compressão de dados, escolhido para esta pesquisa devido ao seu desempenho em outros contextos de classificação. A aplicação de um método de compressão no contexto de classificação de sinais de voz se

mostra adequada, o que pode ser observado a partir da descrição da operação do método, que cria um modelo da fonte de informação, armazenando as probabilidades de surgimento dos símbolos do alfabeto desta fonte. Esse modelo pode ser usado em atividades de classificação.

No capítulo seguinte, serão apresentados trabalhos relacionados, tanto no que se refere à detecção por computador de patologias da fala quanto à aplicação do PPM.

Capítulo 3

Trabalhos Relacionados

Neste capítulo, são apresentados trabalhos relacionados em todas as esferas que envolvem esta pesquisa, ou seja, a detecção por computador de patologias da fala e a aplicação do método PPM com fins à classificação. São apresentadas também pesquisas que envolvem a utilização do método PPM com fins à compressão de dados, com o intuito de apresentar seu potencial.

3.1 Detecção de Patologias da Fala

Uma das técnicas mais adotadas na detecção de patologias da fala é a *análise acústica*. O interesse pela análise acústica para avaliar a voz sempre existiu (dadas as pesquisas de Lieberman (1963) e Costa (2008)), pelo fato de ser possível utilizá-la como meio para detectar patologias ou, até mesmo, para determinar alterações da função vocal, avaliações de cirurgias, tratamentos farmacológicos e de reabilitação (COSTA, 2008). O principal objetivo do uso desta técnica na avaliação vocal tem sido prover uma ferramenta não-invasiva destinada à detecção de patologias (TAVARES et al., 2011), visando, pelo menos, a reduzir a frequência do uso de exames invasivos, os quais são tradicionalmente administrados neste tipo de avaliação.

Na pesquisa de Costa (2008), foi demonstrado que o uso de LPC, combinada a modelos de Markov escondidos (*Hidden Markov Models - HMM*) e Quantização Vetorial, proporciona excelentes percentuais de acerto, tais como (i) 100% de acerto na distinção entre vozes com Edema e vozes Normais utilizando LPC e HMM; (ii) 96% na diferenciação entre vozes com Edema e Outras patologias, utilizando LPC e HMM; e (iii) 99%

na distinção entre vozes Normais e Patológicas, utilizando LPC, coeficientes Delta-Cepstrais ponderados e Cepstrais ponderados, cada um combinado com HMM. Foram considerados apenas os resultados relacionados à métrica *Eficiência*, definida no trabalho. Na obtenção destes resultados, foi adotada a base de dados *Kay Elemetrics*, sendo utilizados 120 sinais, sendo 25 para treinamento (todos afetados por Edema) e o restante para teste. Todos eles representam a elocução dos pacientes da vogal */ah/* sustentada.

Por sua vez, Tavares et al. (2011), considerando apenas coeficientes obtidos a partir da utilização de LPC, obtiveram percentuais médios de acerto de até 92%, utilizando Quantização Vetorial alimentado com coeficientes Delta-Cepstrais, na distinção entre vozes com Edema de Reinke e vozes Normais, sendo considerados novamente apenas os resultados relacionados à métrica *Eficiência*. A base de dados adotada também foi a *Kay Elemetrics*, sendo utilizados 44 sinais apresentando Edema e 53 de vozes normais. No treinamento, foram utilizados 50% dos sinais de vozes apresentando Edema e o restante (totalizando 75 sinais) foram utilizados na fase de testes. Todos eles representam a elocução dos pacientes da vogal */ah/* sustentada.

Aguiar Neto et al. (2007) executaram uma modelagem acústica de vozes com Edema, considerando coeficientes LPC, Cepstrais e Mel-cepstrais, utilizando quantização vetorial de dimensão 12 e 64 níveis para cada característica extraída de quadros de 20 ms (a extração, porém, se deu a cada 10 ms), visando a classificar um sinal de voz entre Saudável, com Edema e com Outras patologias. O classificador, baseado em quantização vetorial, foi treinado apenas com vozes com Edema, semelhantemente aos trabalhos anteriormente citados. Como resultado, constatou-se que a utilização desta abordagem ainda precisa ser melhorada (devido à falsa rejeição mínima de 23% em todas as classificações executadas), mas também que elas se mostraram promissoras, pelo fato de terem sido obtidos 96% de correta rejeição na distinção entre Edema e Outras patologias utilizando LPC e 100% de

correta rejeição utilizando as três técnicas na distinção entre Normal e Patológico e Normal e Edema (nesta última, porém, a correta aceitação também foi baixa, entre 49 e 76%). A base de dados adotada também foi a *Kay Elemetrics*, sendo empregados 44 sinais de vozes com Edema, 53 de vozes normais e 23 de outras patologias, tais como Cisto, Nódulos e Paralisia, todos representando a elocução vocal da vogal */ah/* sustentada. Conforme já mencionado, apenas 20 sinais de Edema foram utilizados no treinamento e o restante apenas na fase de testes. Os sinais de vozes normais foram subamostrados a 25 kHz, de modo a ficarem equivalentes aos sinais de vozes patológicas.

Arias-Londoño et al. (2011) utilizaram *modelos de misturas Gaussianas* (*Gaussian Mixture Models - GMM*) combinados a *máquinas de vetores de suporte* (*Support Vector Machines - SVM*) para distinguir sinais de voz entre Normal e Patológico. Foram utilizados dois classificadores, baseados em GMM, cada um recebendo dados de tipos diferentes: um caracterizado por 11 (onze) medidas de complexidade, dentre as quais *expoente de Lyapunov* e *dimensão de correlação* e o outro caracterizado por medidas de ruído e coeficientes Cepstrais. As saídas de ambos os classificadores foram combinadas a partir de uma abordagem discriminativa baseada em SVM. A precisão das classificações realizadas chegou a 98,23% (usando SVM e GMM). Também foi utilizada a base de dados *Kay Elemetrics*, sendo 53 sinais de vozes normais e 173 de vozes patológicas, todas na elocução da vogal */ah/* sustentada, semelhante ao que fizeram Parsa e Jamieson (2000). Os sinais de voz que eram amostrados originalmente a 50 kHz (todos os sinais de vozes normais e alguns de patológicos) foram subamostrados a 25 kHz. A estratégia de treinamento consistiu em Validação Cruzada (será explicado na Seção 4.2.4) de 10 partições.

Patil e Baljekar (2012) utilizaram características de fase TEO (*Teager Energy Operator*) e coeficientes Mel-cepstrais na classificação de um sinal de voz em Saudável ou Patológico. Segundo os autores, este é o primeiro trabalho que empregou esforços na investigação de

características analíticas de fase na caracterização do efeito das patologias na fonte do processo de produção da voz. O melhor resultado obtido nesta investigação foi 97,5% de acurácia (usando ambas as técnicas mencionadas). Nesta pesquisa também foi utilizada a base de dados Kay Elemetrics e também seguindo o que foi feito por Parsa e Jamieson (2000) - 53 de vozes normais e 173 de vozes patológicas, todos na elocução da vogal */ah/* sustentada. Nesta pesquisa porém, diferentemente da apresentada por Arias-Londoño et al. (2011), foi utilizada Validação Cruzada com 4 parcelas, repetidas 12 vezes. O resultado final apresentado consistiu da média das classificações executadas.

Raju et al. (2012) propuseram a extração de parâmetros, tais como *pitch* e frequências formantes, para diferenciar vozes Patológicas (com Disartria) de vozes Saudáveis, utilizando SVM. Ambos os parâmetros foram extraídos por meio do uso de LPC que, segundo os autores, permite distinguir satisfatoriamente vozes Normais de Patológicas, mas não fornece valores mensuráveis para que seja tomada uma decisão. Os resultados apresentados consistiram na constatação da possibilidade de distinção visual do traçado gráfico de ambos os parâmetros e de que o método proposto provê boa distinção entre normal e patológico (nenhum resultado percentual foi fornecido). Os autores mencionaram que utilizaram a base de dados *Kay Elemetrics*, mas não mencionaram quantos arquivos utilizaram, apenas que os sinais de vozes patológicas utilizados apresentam Disartria.

Monteiro et al. (2011) utilizaram técnicas semelhantes àquelas usadas em outras investigações: coeficientes LPC e Cepstrais, classificados por quantização vetorial. O objetivo era classificar entre Normal, Edema ou Paralisia. Como resultado, constatou-se que as maiores taxas de reconhecimento advieram da alimentação do modelo com coeficientes LPC (87,36%). A base de dados utilizada também é da *Kay Elemetrics*, sendo 50 sinais de vozes normais, 57 apresentando Paralisia e 44 apresentando Edema, todos na elocução da vogal */ah/* sustentada.

Outro conceito empregado na detecção de patologias é a entropia (Seção 2.3.3). Sua utilização é apropriada neste contexto pelo fato de os padrões vocais patológicos apresentarem-se bem mais complexos e imprevisíveis do que aqueles associados a vozes normais. Esse resultado também foi obtido por Patil e Patel (2012). Na pesquisa de Tavares et al. (2011), já mencionada anteriormente, os percentuais de acerto obtidos com o uso da entropia combinada a coeficientes Cepstrais chegam a 98%, com alguns casos de 100% de correta aceitação, o que mostra uma melhora com relação aos resultados anteriormente apresentados, que consistiam na utilização de apenas um tipo de coeficiente cepstral.

Scalassara et al. (2009b) citou trabalhos em que foram utilizadas a entropia *aproximada* (MOORE et al., 2004; MOORE; MANICKAM; SLEVIN, 2006), a entropia *relativa* (SCALASSARA et al., 2009c) e a *entropia de Shannon* (SCALASSARA et al., 2008) no contexto de classificação de padrões vocais e, neste relato, fizeram uso da taxa de entropia. Como resultado, utilizando 14 sinais de uma base de dados própria (ainda restrita ao laboratório em que foi feito o estudo na época de sua publicação), sendo 7 Saudáveis e 7 com Nódulos. Todos os padrões vocais com Nódulos apresentaram entropia mais alta do que todos os padrões vocais Saudáveis. Este resultado reforça a teoria de que padrões vocais Patológicos são mais complexos e imprevisíveis do que padrões vocais Saudáveis.

A entropia relativa, também conhecida como entropia de *Kullback-Leibler*, foi empregada por Scalassara et. al. (2009c), como mencionado anteriormente. Os autores forneceram uma informação útil: a entropia pode ser utilizada com qualquer distribuição de probabilidade que se harmonize com a noção intuitiva de medidas da informação. Sendo assim, a entropia relativa possibilita mensurar a dificuldade de se discriminar duas distribuições. Maciel, Pereira e Stweart (2010), por sua vez, afirmaram que a entropia relativa permite mensurar a diferença entre duas distribuições. Foram extraídos valores médios e de desvio padrão de medições de entropia dos sinais, seguida da execução de um teste *t* de

Student não pareado. Os registros de ambos os trabalhos não apresentaram percentuais de acerto, mas seus autores constataram a superioridade do valor das amostras com patologia. Scalassara et. al. (2009c) utilizaram 48 sinais de uma base de dados própria, dividida igualmente entre Normal, Edema e Nódulo. Maciel, Pereira e Stweart (2010), por sua vez, não mencionaram a origem dos sinais de voz utilizados, se limitando a dizer que eles foram avaliados por um profissional da área.

Costa et al. (2007) utilizaram a entropia de Shannon e a entropia relativa, visando a discriminar sinais de voz entre Saudáveis e apresentando Edema de Reinke. As amplitudes dos sinais de voz foram normalizadas (entre 0 e 1) e quantizadas, a partir de um processo de quantização linear de 32 níveis. Os valores que alimentaram os algoritmos dos cálculos das entropias advieram de histogramas de segmentos desses sinais e a discriminação foi feita a partir do erro médio quadrático mínimo. Os resultados, por sua vez, foram modestos: não ultrapassaram 82%. Foi adotada também a base de dados da *Kay Elemetrics*, sendo utilizados 43 sinais de vozes apresentando Edema e 50 sinais de vozes normais, todos na elocução da vogal */ah/* sustentada.

Redes neurais foram utilizadas nas pesquisas de Behroozmand e Almasganj (2005), Salhi, Talbi e Cherif (2008) e Marinus et al. (2009). Na pesquisa de Behroozmand e Almasganj (2005), o intuito era distinguir entre Edema nas dobras vocais, Nódulos e Pólipo. Para tanto, foi conduzida uma etapa de pré-processamento (pré-ênfase e janelamento), seguida de uma etapa de extração de características dos sinais, usando coeficientes Mel-Cepstrais e decomposição *wavelet*, sendo a classificação realizada a partir de redes neurais com retropropagação, SVM e algoritmos genéticos (para a seleção do índice do vetor menos correlacionado), todos separadamente, i.e., não houve combinação entre as técnicas. O melhor percentual de acerto (94,12%) foi obtido utilizando a entropia de transformadas *wavelet*, uma das variantes de SVM e algoritmos genéticos. A base de dados adotada também foi a da Kay

Elemetrics, sendo utilizados 83 sinais de voz, divididos em 44 apresentando Edema, 19 de Nódulos e 20 de Pólipos, todos na elocução da vogal */ah/* sustentada. Na fase de treinamento, foram utilizados 60% dos sinais listados e o restante na fase de testes.

Na pesquisa de Salhi, Talbi e Cherif (2008), a técnica de redes neurais foi combinada com transformadas *wavelet*. Semelhantemente à pesquisa de Behroozmand e Almasganj (2005), transformadas *wavelet* discreta e contínua foram empregadas na extração de características dos sinais, tendo sido utilizada uma rede neural multicamada no processo de classificação. Os percentuais de acerto obtidos foram diferentes para a classificação entre padrões vocais Saudáveis e Patológicos. Para padrões Saudáveis, foram obtidos até 100% de acerto, enquanto para padrões Patológicos, até 90%. Os melhores resultados foram obtidos com o uso de coeficientes *wavelet* contínuos e também a partir da entropia dos coeficientes extraídos. Foram utilizados 100 sinais provenientes de uma base de dados preparada com a ajuda da Universidade de Los Angeles e de um hospital tunisiano, sendo 80 para treinamento (divididos igualmente entre Normal e Patológico, sendo esta composta por diferentes patologias) e o restante para teste (também divididos igualmente entre estas classes). Estes sinais, diferentemente dos sinais utilizados nas pesquisas anteriormente apresentadas, representam a elocução de palavras, ao invés da vogal */ah/* sustentada.

A metodologia adotada no estudo de Marinus et al. (2009) assemelha-se àquela descrita por Behroozmand e Almasganj (2005): uma etapa de pré-processamento precedeu a etapa de extração de características com coeficientes Cepstrais, a qual, por fim, deu sequência a uma etapa de classificação de padrões a partir de redes neurais. As diferenças entre as metodologias adotadas nas referidas pesquisas residem nas duas últimas etapas: Marinus et al. (2009) empregaram 12 coeficientes Cepstrais para extrair características dos sinais e uma rede neural do tipo *perceptron* multicamada com retropropagação para realizar a classificação. Behroozmand e Almasganj (2005), por sua vez,

empregaram coeficientes Mel-cepstrais e decomposição *wavelet* para extrair características e SVM, redes neurais com retropropagação e algoritmos genéticos para realizar a classificação.

Marinus et al. (2009) testaram várias arquiteturas de redes neurais, cada uma das quais considerando um número diferente de neurônios na camada escondida. Os autores conseguiram 99,36% no reconhecimento de vozes Normais (usando 46 neurônios), 96,09% na detecção de vozes com Edema (usando 48 neurônios) e 93,93% na detecção de vozes com Outras patologias (usando 50 neurônios). Foi adotada a base de dados da *Kay Elemetrics*, tendo sido usados 44 sinais de vozes com Edema, 23 com outras patologias, tais como Nódulo, Cisto e Paralisia e 53 de vozes normais, totalizando 120 sinais. Os autores dão a entender erroneamente que esta base de dados é composta de 1400 sinais de voz na elocução da vogal */ah/* sustentada (na Seção 5.1 é mostrada em detalhes a composição desta base de dados), levando a crer que os sinais utilizados representam esta elocução vocal.

Costa et al. (2012) utilizaram gráficos de recorrência para classificar um sinal de voz como Saudável ou apresentando Edema ou Paralisia (ECKMANN; KAMPHORST; RUELLE, 1987). Para tanto, foram empregadas representações gráficas, a partir das quais foram extraídas 7 (sete) medidas de quantificação de recorrência que auxiliaram no processo de tomada de decisão. Isto significa que os dados manipulados não são extraídos diretamente dos sinais de áudio, mas de gráficos que os representam. Segundo os autores, os gráficos de recorrência permitem que sejam identificadas propriedades que podem ser observadas a partir do emprego de técnicas não-lineares. Como resultados, os autores constataram que quanto maior a recorrência, menor a taxa de classificação. As melhores taxas são obtidas quando se tem 1% de recorrência. Porém, o principal resultado está nas altas taxas de acerto obtidas pelo uso de algumas das medidas, como *Entropia* e *Tamanho máximo das linhas verticais*. No caso da distinção entre vozes Saudáveis e com Paralisia, foram obtidos até mesmo índices de 100%, considerando

as métricas Sensibilidade e Especificidade (para Edema, o índice máximo foi 95% de Especificidade), porém, considerando a Acurácia, que mede a performance geral, os índices máximos foram 95% para Normal x Paralisia e 80,56% para Normal x Edema. Não foi feita a distinção entre vozes com Edema e vozes com Paralisia. Os autores também adotaram a base de dados da *Kay Elemetrics*, embora não tenham especificado o tipo de elocução vocal dos sinais. Foram utilizados 45 sinais com Edema (29 para treinamento e 16 para teste), 55 com Paralisia (35 para treinamento e 20 para teste) e 53 de vozes normais (33 para treinamento e 20 para teste).

Uma investigação muito semelhante foi conduzida por Vieira, Costa e Costa (2012), na qual os gráficos de recorrência foram empregados visando a classificar um sinal de voz como Normal ou apresentando Paralisia. Foram empregadas as mesmas medidas de quantificação de recorrência (MQR) em ambas as investigações. A principal diferença metodológica entre ambas as investigações está na variação do raio de vizinhança, a qual ocorreu entre 1 e 15% nesta, enquanto na investigação conduzida por Costa et al. (2012) não foram mencionadas variações neste parâmetro. Os autores constataram que os valores de raios de vizinhança dos melhores resultados são sempre baixos (entre 2 e 6%) e que a MQR *comprimento máximo das linhas diagonais* (L_{max}) é a que fornece os melhores resultados de acurácia, seja na análise discriminante linear ou quadrática: 91,73 ± 9,05% e 93,64 ± 6,14%, respectivamente. Em um segundo momento, algumas dessas medidas foram combinadas, havendo melhoria nos resultados. O melhor deles foi 98,18 ± 3,83% quando da combinação de três medidas e sendo feita a análise discriminante quadrática. Também foi adotada a base de dados da Kay Elemetrics, tendo sido utilizados 108 sinais no total, divididos em 53 de vozes normais e 55 de vozes com Paralisia, todos representando a elocução da vogal */ah/* sustentada. A divisão entre treinamento e teste foi feita utilizando Validação Cruzada de 10 parcelas.

Orozco et al. (2012) propuseram o uso de dinâmica não linear (*Non-Linear Dynamics - NLD*) para detectar a presença de patologias em fala contínua, i.e, a fala registrada enquanto o paciente pronuncia uma frase (ao invés da fala sustentada, modalidade mais comum nas pesquisas desta área). Foram utilizadas 4 (quatro) medidas de complexidade, dentre as quais a *complexidade de Lempel-Ziv* e a *dimensão de correlação*, todas extraídas após um processo de reconstrução do espaço de estados. Após esse processo, compôs-se um vetor de características contendo descritores estatísticos referentes aos conjuntos dessas medidas, e.g., desvio padrão e média. Para classificar os dados obtidos, foram experimentadas SVM, redes neurais e o algoritmo de k vizinhos mais próximos, sendo o melhor resultado obtido com SVM (95% ± 3,54%). Também foi adotada a base de dados da *Kay Elemetrics*, mas desta vez, os sinais de voz utilizados fazem a elocução da *Rainbow Passage*[17]. Foram utilizados 396 sinais de voz, sendo 360 de vozes patológicas (com uma variedade de patologias) e o restante de vozes normais. A seleção entre treinamento e teste foi feita utilizando Validação Cruzada com 10 parcelas.

Lima et al. (2012) utilizaram o expoente de Hurst para fazer a distinção entre vozes Saudáveis, com Edema e com Paralisia. Segundo os autores, esse parâmetro é adequado a processamentos dessa natureza por representar o comportamento estocástico da voz, além de apresentar baixo custo computacional, podendo ser obtido em tempo real. Os arquivos foram divididos em segmentos e de cada segmento foi extraído um valor deste parâmetro. Usando análise de discriminantes lineares (*Linear Discriminant Analysis - LDA*), foi obtida em todos os casos de classificação correta rejeição de 100%, com um máximo de 98,39% de medida de Eficiência. Também foi adotada nesta investigação a base de dados da *Kay Elemetrics*, tendo sido utilizados 53 sinais de vozes normais,

[17] Esta base de dados, conforme descrição na Seção 5.1, contém sinais na elocução da vogal */ah/* sustentada e os 12 primeiros segundos da *Rainbow Passage*.

44 com Edema e 52 com Paralisia, todos na elocução da vogal */ah/* sustentada.

O Quadro 2 contém uma síntese das pesquisas discutidas nesta seção.

Quadro 2 - Resumo das pesquisas relacionadas à detecção de patologias da fala revisadas nesta dissertação

Autor	Abordagem adotada	Percentuais de acerto obtidos
Costa (2008)	Codificação por predição linear, coeficientes cepstrais e Modelos de Markov escondidos	99% entre voz Normal e voz Patológica, 100% entre voz Normal e voz com Edema e 96% entre voz com Edema e voz com Outras patologias
Arias-Londoño et al. (2011)	GMM e SVM	98,23% na distinção entre voz Normal e voz Patológica, utilizando Validação Cruzada de 10 parcelas
Patil e Baljekar (2012)	Fase TEO	97,5% na distinção entre voz Normal e voz Patológica, utilizando Validação Cruzada com 4 parcelas e 12 repetições
Tavares et al. (2011)	Entropia de Shannon e Coeficientes Cepstrais	98% de Eficiência na distinção entre voz Normal e voz com Edema, sendo usados 50% dos sinais com Edema no treinamento
Behroozmand e Almasganj (2005)	Pacote *wavelet* e SVM	94,12%, sendo utilizados 60% dos sinais no treinamento
Salhi, Talbi e Cherif (2008)	Redes Neurais, Coeficientes *wavelet* contínuos e entropia	100% para voz Normal e 90% para voz Patológica, sendo utilizados 80% dos sinais no treinamento
Marinus et al. (2009)	Coeficientes Cepstrais e Rede Neural perceptron multi-camada com retropropagação	99,36% de acerto na detecção de vozes normais, 96,04% na detecção de vozes com Edema e 93,78% na

		detecção de vozes com Outras patologias, sendo utilizados 50% dos sinais no treinamento, 25% para validação e 25% para testes
Costa et al. (2012)	**Gráficos de recorrência com extração de 7 medidas**	**95% de Acurácia para Saudável x Paralisia e 80,56% de Acurácia para Saudável x Edema, sendo utilizados 60% dos arquivos para treinamento**
Vieira, Costa e Costa (2012)	**Gráficos de recorrência**	**98,18 ± 3,83% na distinção entre voz Normal e voz com Paralisia, utilizando Validação Cruzada de 10 parcelas**
Orozco et al. (2012)	**Dinâmica Não-Linear e SVM**	**95% ± 3,54% na distinção entre voz Normal e voz Patológica, utilizando Validação Cruzada de 10 parcelas**
Lima et al. (2012)	**Expoente de Hurst e Análise Discriminante Linear**	**94,91 na distinção entre voz Normal e voz Patológica, 89,29 na distinção entre voz Normal e voz com Edema e 98,39 na distinção entre voz Normal e voz com Paralisia**

Na próxima seção serão apresentadas aplicações do PPM em compressão de dados. O intuito desta seção é mostrar seu poder de compressão, razão pela qual é considerado um dos mais eficazes compressores da atualidade, e os pré-processamentos que tem sido feitos para melhorar seus resultados.

3.2 Aplicações do PPM em Compressão de Dados

Drinic e Kirovski (2002) empregaram PPM para comprimir arquivos binários executáveis da plataforma Windows. Contudo, não empregaram o

procedimento original, uma vez que empregaram heurísticas em etapas de pré-processamento, tais como a reorganização das instruções e mistura do que foi chamado de *subfluxos*, além de ser aplicado conhecimento sobre a estrutura das instruções contidas nesses arquivos, tal como o fato de variarem em tamanho entre 1 e 16 bits (ao invés dos 8 bits fixos de textos) e da composição e correlação entre seus campos. Como resultado, foram reportados resultados 16 a 22% superiores àqueles obtidos a partir do emprego do PPMD (uma das variações do PPM, como o PPM-C, utilizado nesta pesquisa), considerado o melhor compressor pelos autores.

Souza (2008) construiu um sistema completo de aquisição e compressão de sinais de eletrocardiograma, que incluía conversor analógico-digital e dispositivo FPGA (*Field-programmable Gate Array*). O processo ainda incluía a conversão para código Gray (ver Salomon (2004)) e decomposição dos sinais em planos de bits, antes da compressão, a partir de um PPM adaptado ao alfabeto binário. A intenção era coletar os dados e reduzir os custos de transmissão e armazenamento. Um fato curioso sobre essa adaptação é que não havia alocação dinâmica de memória: a estrutura do modelo já continha os espaços para todos os possíveis contextos, o que, segundo o autor, reduziu a demanda de memória. O melhor resultado, em termos da razão de compressão, foi obtido com tamanho de contexto 7: 2,46:1. Porém, com tamanho de contexto 1, foi obtida uma razão de compressão de 2,38:1.

Farias (2010) desenvolveu um *holter*, simulando sinais de eletrocardiogramas (ECG) com séries de Fourier, empregando o processador Nios II e FPGA, codificação Gray e decomposição em planos de bits para os sinais das amostras e PPM, também adaptado para alfabeto binário, a fim de comprimir os sinais codificados, igualmente no intuito de reduzir os custos de transmissão e armazenamento. Várias compressões foram executadas e várias razões de compressão foram obtidas a partir daí. A melhor foi 3:1, com tamanho de contexto 5. Porém,

similarmente à pesquisa de Souza (2008), com contexto 1 a melhor razão de compressão obtida foi 2,87:1, o que faz suscitar a dúvida de qual é o melhor tamanho de contexto a ser empregado.

Marques et al. (2006) propuseram uma abordagem para a compressão de imagens mamográficas utilizando PPM, também adaptado para alfabeto binário, acrescentando ao processo a decomposição em planos de bits. O objetivo da pesquisa, conforme ressaltado diversas vezes pelos autores, foi mostrar que a decomposição em planos de bits, em associação com um esquema de modelagem avançado, produz um esquema de compressão eficaz e traz uma série de benefícios adicionais. Como resultado, foram obtidas taxas de compressão de 36,7%, com tamanhos de contexto entre 8 e 11, i.e., a imagem comprimida apresentava 36,7% das dimensões originais.

O Quadro 3 contém uma síntese das pesquisas discutidas nesta seção e a próxima seção apresenta registros do emprego do método PPM na classificação de diversos tipos de padrões.

3.3 Usos do PPM em Processos de Classificação de Padrões

Pelo fato de o método PPM ter sido originalmente concebido com fins à compressão de dados (CLEARY; WITTEN, 1984; SALOMON, 2004), sua aplicação à classificação de padrões vocais pode parecer, à primeira vista, inadequada. Todavia, é importante ter em mente que os algoritmos modernos de compressão sem perdas, em geral, constroem modelos precisos, os quais representam adequadamente a distribuição de probabilidades dos símbolos na fonte do fluxo a ser lido. Desde a concepção da técnica, foi investigada sua aplicação à classificação de diferentes tipos de informação, dentre as quais textos nas mais diversas línguas e sinais de eletrocardiogramas, conforme será discutido a seguir.

Coutinho et al. (2005) e Barufaldi et al. (2009) utilizaram o método PPM na classificação de textos em língua portuguesa, visando a identificar autoria e período literário, dentre outros itens, com taxas de acerto que variam entre 45,7% e 100%, dependendo dos tamanhos do contexto

(variantes entre 0 e 11) e dos arquivos usados no treinamento (variantes entre 8 Kb e 128 Kb). Dentre as vantagens apresentadas para sua utilização, são destacadas: (i) a descrição do conteúdo como um todo, sem a necessidade de descarte de informação;

Quadro 3 - Resumo das pesquisas relacionadas à compressão de dados revisadas nesta dissertação

Autores	Objeto de compressão	Etapa acrescentada	Resultados
Drinic e Kirovski (2002)	Binários executáveis da plataforma Windows	Reorganização das instruções e mistura de subfluxos; Heurísticas sobre os campos das instruções	16 a 22% superiores ao emprego do PPMD neste mesmo contexto
Souza (2008)	Sinais de eletrocardiograma (ECG)	Conversão para código Gray; Decomposição em planos de bits; Adaptação ao alfabeto binário	Contexto 7 - 2,46:1 Contexto 1 - 2,38:1
Farias (2010)	Sinais de eletrocardiograma (ECG)	Conversão para código Gray; Decomposição em planos de bits; Adaptação ao alfabeto binário	Contexto 5 - 3:1 Contexto 1 - 2,87:1
Marques et al. (2006)	Imagens mamográficas	Adaptação ao alfabeto binário; Decomposição em planos de bits;	Imagem comprimida com 36,7% das dimensões originais utilizando contextos entre 8 e 11

(ii) a isenção de considerações simplificadoras a respeito das distribuições de probabilidades; e (iii) a capacidade de adaptação do modelo à distribuição de probabilidades apresentada na mensagem, o que oferece um modo uniforme de classificar diferentes fontes de informação. Na

pesquisa de Coutinho et al. (2005), foram utilizadas 48 obras, divididas em 12 (doze) classes, enquanto na pesquisa de Barufaldi et al. (2009) foram utilizadas 37 obras, divididas em 4 (quatro) períodos literários.

Mahoui et al. (2008) foram além e estudaram a eficácia do PPM na mineração de dados. O objetivo era usar a classificação, a partir da compressão via PPM, para identificar trechos que se referiam a funções de genes dentro de resumos de trabalhos da área biomédica (em inglês). Para tanto, a etapa de treinamento foi precedida por uma etapa de pré-processamento, na qual se empregou a aplicação de *software* TMT[18] (*Text Mining Toolkit*), a fim de separar trechos referentes às funções daqueles que não se referiam a esses diretamente. Diversas variantes foram testadas e o melhor percentual de acerto obtido foi de 97,89%. Não foi mencionado o tamanho da base de dados usada, mas esta se denomina *GeneRIF* e pertence ao Centro Nacional de Informação de Biotecnologia, NCBI, do governo americano.

Burbey e Martin (2008) também utilizaram texto, mas na forma de tuplas, a fim de predizer localizações futuras de um indivíduo, juntamente com a hora. A base de dados usada se constituiu de 28.000 pares <tempo, localização>, extraídos de *logs* de acesso sem fio do padrão IEEE 802.11. Similarmente ao estudo de Mahoui et al. (2008), diversas variantes foram testadas e os percentuais de acerto chegaram a 98% (aplicando em terceira ordem, i.e, dado um par <tempo, localização> e outro dado de tempo, predizer a localização).

Chaiwanarom e Lursinap (2008) empregaram a classificação a partir do PPM em um domínio bastante peculiar - a completude da autoria de artigos científicos, cujo propósito era predizer se determinado autor encontrava-se entre os autores de um artigo científico. Para o teste, um dos autores de cada artigo foi retirado propositalmente e artigos publicados por apenas um autor não foram considerados.

A peculiaridade desta aplicação da classificação via PPM reside no fato de que se baseia em grafos denominados *grafos de autoria*, os quais

[18] Disponível em http://www.informatics.bangor.ac.uk/~wjt/software/TMT-0.08.tar.gz

representam a autoria de artigos científicos e nos quais cada vértice representa um autor e uma aresta entre dois vértices rotulada por um *ID* representa um artigo publicado por ambos os autores. Um artigo publicado por mais de dois autores é representado por várias arestas com o mesmo *ID* conectando diferentes vértices. Para encaixar esse tipo de dado na tabela, os autores transformaram as conexões em sequências de identificadores de autores e empregaram essas sequências como entradas para o modelo PPM. Além disto, armazenaram as sequências de autores na direção normal, na direção inversa e de maneira híbrida (ambas as direções).

Os percentuais de acerto variaram entre menos de 60% (contexto 3, faltando o primeiro ou o último autor nos artigos) e 100% (8 e 9 autores, contexto 2 a 3[19] e usando modelagem híbrida). A base de dados considerada pelos autores consistiu em 10.053 artigos de Ciência da Computação, extraídos da base de publicações científicas DBLP (*DataBase systems and Logic Programming*).

Honório, Batista e Duarte (2009) apresentaram um sistema de classificação de texturas em imagens usando PPM e equalização de histogramas. As imagens com histograma equalizado foram lidas horizontalmente e, em seguida, verticalmente. Para cada modalidade de leitura, foi criado um modelo, para posterior comparação no estágio de classificação. Visando a diminuir os requisitos computacionais, o codificador aritmético não foi utilizado. Ao invés disto, foi empregada a entropia condicional: as probabilidades da textura de uma imagem foram calculadas a partir de modelos referentes a outras imagens. Desta forma, verificou-se a similaridade da textura de uma imagem com relação a determinada classe de texturas. Quanto menor a entropia, maior a similaridade. As taxas de acerto obtidas variaram entre 99,73 e 100%.

Malheiros et al. (2012) empregaram PPM como base para a construção de um sistema de recomendação, no intuito de auxiliar novatos em projetos de desenvolvimento de *software* a executarem suas

[19] O significado deste tamanho de contexto não foi explicado pelos autores.

atividades sem o auxílio de um mentor, i.e., um membro experiente do projeto, de modo a utilizar seu tempo em atividades mais importantes. Este tipo de ferramenta é útil em grandes projetos de software livre, tais como GIMP ou *Mozilla Firefox*, em que qualquer desenvolvedor pode contribuir, mas se juntará ao time em um estágio já avançado do projeto.

Os dados utilizados para alimentar o modelo são originários do resumo, descrição e comentários sobre requisições de mudanças, concatenados em uma única string. Para cada requisição, foi criado um modelo PPM e a similaridade entre as requisições foi calculada a partir da entropia condicional, semelhante ao que foi realizado por Honório, Batista e Duarte (2009).

Os autores testaram a abordagem proposta nos documentos relativos aos projetos GIMP, GTK+ e Hadoop, separadamente, e compararam com os resultados obtidos pela utilização da técnica LSI (*Latent Semantic Indexing*), empregada na aplicação de software *Hipikat*[20], um dos programas de recomendação de artefatos úteis mais conhecidos. Para tanto, os autores criaram uma versão similar do sistema, alterando apenas a técnica de atribuição da similaridade. Como resultado, todos os índices de precisão (máximo de 17,84%), revocação (máximo de 46,52%) e taxa de revocação (máximo de 66,8%) obtidos a partir do emprego do PPM foram superiores àqueles obtidos a partir da utilização de LSI.

Adicionalmente, vários experimentos têm mostrado a superioridade da eficácia da classificação via PPM, quando comparado a outros classificadores, devido à capacidade de construir modelos precisos (MEDEIROS et al., 2011). Na pesquisa de Medeiros et al. (2011), em particular, a partir do emprego de amostras de tacogramas, foram obtidas taxas de acerto de até 99,51%.

Pavelec et al. (2009) compararam PPM com SVM, na identificação da autoria de textos em português e também constataram sua superioridade. Sua pesquisa difere daquela realizada por Coutinho et al. (2005), pelo fato

[20] Disponível em http://www.cs.ubc.ca/labs/spl/projects/hipikat/

de a base de dados ser composta de artigos coletados da Internet sobre os mais diversos temas (e.g., fofoca, vinho, economia), ao invés de obras literárias de autores renomados. Os autores obtiveram 100% de acerto em vários casos, tendo justificado os altos e baixos percentuais de acerto como dependentes do estilo de redação e do tema descrito pelos autores. O percentual médio de acerto foi 84,3%.

O Quadro 4 contém uma síntese das pesquisas discutidas nesta seção.

3.4 Discussão

A pesquisa ora relatada se encontra no campo de detecção de patologias da fala por computador, o qual é composto por pesquisas que visam a detectar a presença ou até mesmo o tipo de patologia presente em um sistema fonador (algumas delas são apresentadas na Seção 3.1), com o intuito principalmente de diminuir a frequência e necessidade dos exames tradicionais, que podem ser invasivos ou subjetivos. Porém, se diferencia destes ao apresentar a utilização de um método ainda não empregado neste contexto, o PPM. A justificativa para sua escolha se encontra nas Seções 3.2 e 3.3, que contêm relatos de resultados de aplicações deste método em diferentes contextos, sempre com bons resultados.

A vantagem do PPM em relação a outros métodos de compressão se encontra no seu conhecido poder de compressão, que é considerado superior ao dos outros métodos da literatura (razão pela qual é o estado da arte na área). Foi levantada a hipótese, portanto, de que seu superior poder de compressão poderia levar a uma boa capacidade de classificação. Algumas das vantagens de sua utilização neste contexto foram mencionadas na Seção 3.3.

O próximo capítulo apresenta a modelagem aplicada no emprego do PPM no auxílio ao diagnóstico de patologias da fala.

Quadro 4 - Resumo das pesquisas relacionadas à detecção de patologias da fala revisadas nesta dissertação

Autores	Objeto de classificação	Propósito da classificação	Resultados
Coutinho et al. (2005)	Obras literárias	Atribuição de autoria	Máximo de 83,33%, utilizando contexto 6
Barufaldi et al. (2009)	Obras literárias	Identificação de período literário	Máximo de 91,89%, utilizando contexto 4
Mahoui et al. (2008)	Artigos científicos	Identificação de trechos referentes a funções de genes	97,89% de precisão e 96,87% de revocação utilizando contexto 5
Burbey e Martin (2008)	Tuplas <tempo, localização> de logs de acesso sem fio	Predição de localizações futuras de uma pessoa	Máximo de 98% utilizando modelo de segunda ordem (dado um par localização e hora, predizer a próxima localização)
Chaiwanarom e Lursinap (2008)	Grafos de autoria	Completude da autoria de artigos científicos	Até 100%, utilizando contextos 2 e 3
Honório, Batista e Duarte (2009)	Imagens	Classificação de texturas	Até 100%, utilizando contexto 1
Malheiros et al. (2012)	Resumo, descrição e comentários sobre requisições de mudanças em projetos de software	Auxílio a novatos em projetos de desenvolvimento de *software* na execução de suas atividades sem o auxílio de um mentor	Até 66,8% de taxa de revocação (não foi mencionado o tamanho de contexto utilizado)
Medeiros et al. (2011)	Tacogramas	Classificação de arritmias cardíacas	Até 99,51%, utilizando contexto 2
Pavelec et al. (2009)	Textos coletados da Internet	Identificação de autor	Até 100% (também não foi mencionado o tamanho de contexto utilizado)

Capítulo 4

Descrição da Modelagem Aplicada

Este capítulo trata das abordagens metodológicas adotadas nos experimentos conduzidos ao longo da realização da pesquisa, mais especificamente do histórico das abordagens experimentadas e da descrição da abordagem escolhida para a obtenção e análise dos resultados finais que são apresentados e discutidos no Capítulo 5 (**Apresentação e Discussão dos Resultados**).

4.1 Histórico das Abordagens Experimentadas

Nesta investigação, o maior foco recaiu na parte técnica, principalmente no que se refere às abordagens de utilização do método e à manutenção do modelo. Todas as abordagens experimentadas, até a configuração final selecionada, serão brevemente descritas nas próximas seções.

4.1.1 Organização da Base de Dados

Nas fases iniciais da pesquisa, o objetivo primordial foi testar a viabilidade do emprego do método PPM na detecção de patologias da voz, principalmente em termos de taxas de acerto, mas também considerando o tempo/custo computacional do processo de classificação. Sendo assim, não foi necessária nenhuma reorganização da base de dados (a qual está caracterizada na Seção 5.1). Considerou-se a divisão original em sinal Normal e Patológico e foram utilizados todos os arquivos da base de dados.

Em um segundo estágio da pesquisa, buscou-se a distinção entre patologias. Então, foram consideradas apenas aquelas patologias que dispunham de uma quantidade adequada de arquivos para representá-las, o que resultou em um total de 23 patologias, além da classe de arquivos

com vozes Normais. Pelo fato de ser muito comum um dado paciente apresentar mais de uma patologia (não é raro encontrar, na base de dados utilizada, arquivos de vozes contendo até 8 patologias), muitos arquivos se repetiam na estrutura de diretórios montada, de modo que se essa configuração chegasse a ser utilizada, o número de arquivos ultrapassaria 1.300.

Porém, após consulta a uma fonoaudióloga (BRANDÃO, 2012), constatou-se que o impacto que cada patologia causa no sistema fonador é diferente, além do que algumas patologias são derivadas de outras. Sendo assim, de um conjunto de patologias da fala associadas ao mesmo sistema fonador, apenas uma pode ser considerada a principal, sendo essa a que direcionará o tratamento, em detrimento das demais. Portanto, esta patologia é a que deve ser informada por um sistema de detecção automática de patologias da fala.

Foram selecionadas, do conjunto de 23 patologias, 6 classes de arquivos com vozes Patológicas, incluindo a classe Outras. Destas, porém, muitas não continham número suficiente de arquivos representando-as (algumas dispunham de menos de 10 arquivos), de modo que não seria possível realizar um estudo utilizando-as, pelo fato de não haver significância estatística suficiente para que se tenha uma boa estimativa de erro dessas classes. Por essa razão, foram consideradas apenas 4 classes de arquivos: Normal, Edema, Paralisia e Outras. Todas dispõem de mais de 30 arquivos, número considerado estatisticamente significativo para a realização da pesquisa. Os arquivos utilizados estão listados no Apêndice B.

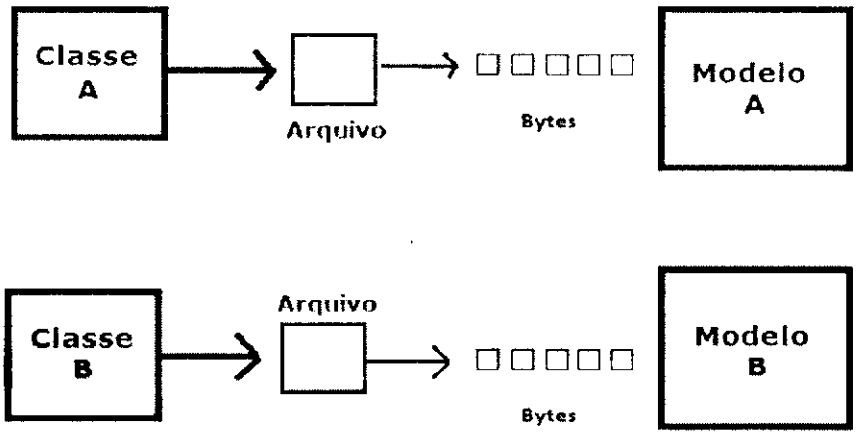
4.1.2 Abordagens de Utilização e Manutenção do Modelo

Diversas estratégias de utilização e manutenção foram testadas até que se chegasse ao modelo final utilizado nesta investigação. Cada uma delas é descrita brevemente a seguir. Tal descrição permite um melhor entendimento das vantagens e limitações do método PPM.

4.1.2.1 Leitura de Bytes e Armazenamento em Memória

A primeira estratégia de utilização do modelo consistiu em alimentá-lo com os Bytes lidos dos arquivos da forma como foram originalmente armazenados (diretamente do fluxo de entrada de dados). Um diagrama em blocos desta abordagem é mostrado na Figura 34.

Figura 34 - Diagrama de blocos da abordagem de alimentação com bytes e manutenção em memória



A ideia era que o modelo PPM fosse composto diretamente pelos bytes lidos do dispositivo de armazenamento, semelhantemente ao que foi descrito na Seção 2.4. Esta é a estratégia mais natural, tendo em vista que a técnica foi originalmente projetada e tem sido mais utilizada para a compressão de dados.

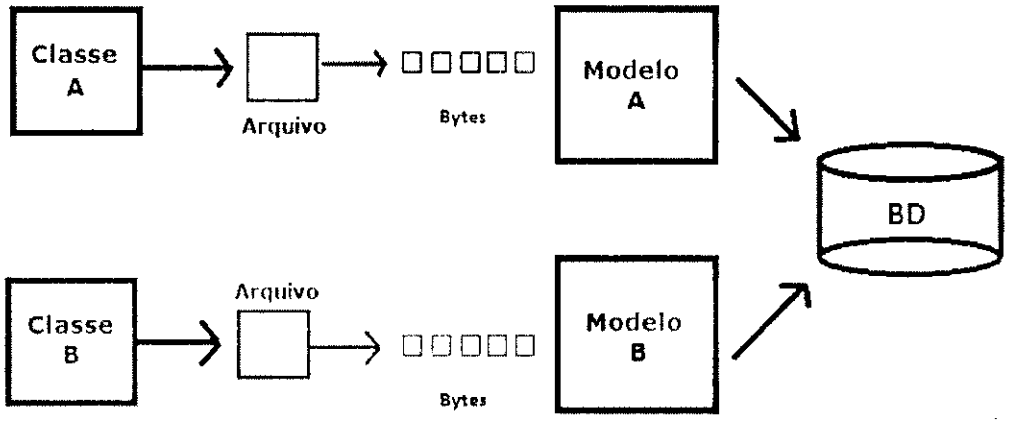
Como pode ser visto nas Seções 3.2 e 3.3, é cada vez mais comum a inclusão de etapas antes e após a utilização do PPM, como pré-processamento dos dados e aquisição do conhecimento sobre sua organização, com fins à obtenção de melhores resultados. Porém, a inclusão destas etapas pouco afeta o tamanho do modelo, já que o fluxo de dados que o alimenta é derivado do fluxo original e aproximadamente do tamanho deste, de modo que ainda assim há chances de ocorrer problemas de estouro de memória. Este tipo de problema ocorreu

constantemente nos experimentos iniciais[21] quando utilizados contextos maiores que 2, devido à grande extensão da base de dados utilizada (666 arquivos, principalmente aqueles referentes a vozes Patológicas) e à extensão do alfabeto considerado (256 símbolos) e, consequentemente, grande quantidade de dados[22] que podia estar presente no modelo ao ser utilizada esta abordagem. Por essa razão, ela teve que ser substituída.

4.1.2.2 Leitura de Bytes e Armazenamento em Banco de Dados

Com o intuito de manter o modelo completo, mas sem que houvesse ocupação excessiva dos recursos de memória disponíveis, foi considerada a utilização de bancos de dados - diretamente em *JDBC* ou por intermédio do *Hibernate* -, como mostrado na Figura 35.

Figura 35 - Diagrama de blocos da abordagem de alimentação com bytes e manutenção em banco de dados



Esta abordagem é muito semelhante à abordagem descrita anteriormente, com a diferença de que dessa vez foi tentada a manutenção do modelo em banco de dados, de modo a evitar ocupação excessiva dos recursos de memória disponíveis. Porém, devido às diversas

[21] Na ocasião, foi utilizada uma máquina com 2 GB de memória principal e processador Dual Core

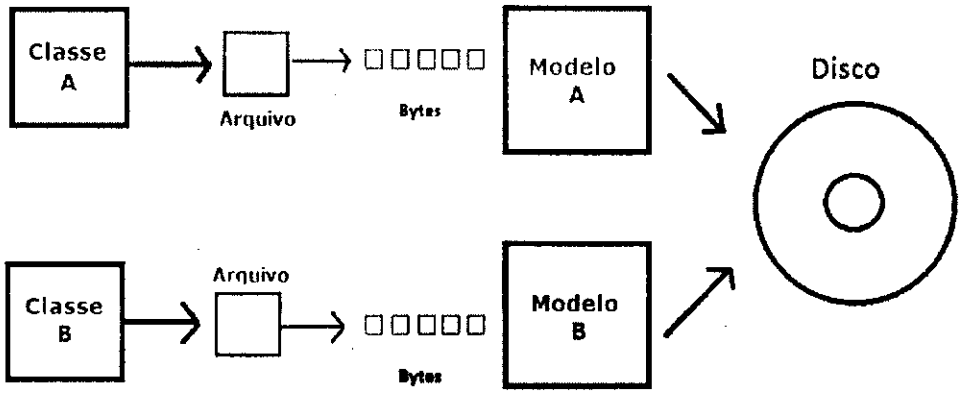
[22] $256^k * 256^{k-1} \ldots * 256$ contextos, sendo k o tamanho máximo de contexto adotado. Para cada contexto, é possível também que estejam associados mais 256 símbolos e o contador de cada um. Quando classificando textos (o contexto de utilização mais comum do PPM), o alfabeto considerado é composto de aproximadamente 40 símbolos apenas.

operações de segurança feitas pelos sistemas de gerenciamento de bancos de dados (SGBD), com o intuito de preservar a consistência e integridade dos dados, constatou-se que a classificação feita dessa forma seria excessivamente lenta, de modo que a ideia também foi descartada.

4.1.2.3 Leitura de Bytes e Armazenamento em Disco

Em seguida, visando a evitar as operações de SGBDs (desnecessárias neste contexto) e, com isso, diminuir a lentidão na execução, experimentou-se o armazenamento do modelo em disco, como descrito na Figura 36.

Figura 36 - Diagrama de blocos da abordagem de alimentação com bytes e manutenção em disco



No armazenamento em disco, foi considerado tanto a serialização dos objetos quanto o uso de planilhas. Porém, a primeira destas estratégias também se mostrou excessivamente lenta e a segunda, insuficiente em virtude das estruturas das planilhas não serem capazes de conter o modelo completo.

Ainda foi levada em consideração a "limpeza" do modelo, i.e., a exclusão de partes consideradas não importantes ou até mesmo sua exclusão total, sugerida na pesquisa de Drinic, Kirovski e Potkonjak (2003). Todavia, tal estratégia não chegou a ser implementada pelo fato de, em determinado estágio da pesquisa, conforme mencionado na seção anterior, uma fonoaudióloga ter sido consultada sobre as patologias que

mais afetam o sistema fonador e que usualmente direcionam a intervenção de um fonoaudiólogo[23]. Esta consulta resultou na eliminação de problemas referentes à manutenção do modelo, porém se constatou que a alimentação do modelo a partir de dados diretamente provenientes dos fluxos de entrada não contribuía para a detecção de patologias, de modo que foi necessário formular uma nova estratégia de classificação.

4.1.2.4 Predição por Casamento Parcial Aproximado

Com base na investigação de Zhang e Adjeroh (2008), foi levada em consideração uma técnica derivada do PPM, denominada PPAM (*Prediction by Partial Approximate Matching*), que reúne informações consideradas próximas em uma única entrada, ao invés de armazená-las em diferentes entradas, como no modelo original. Porém, PPAM exige a definição de uma métrica de proximidade. Por exemplo, considerando as sequências [100 67 89 205] e [98 69 91 204], os autores apresentaram duas métricas de aproximação: em uma delas, que utiliza a distância máxima entre dois elementos de posições correspondentes, se diria que a distância entre elas é de 2 unidades; na outra, na qual é feita a soma das distâncias entre elementos de posições correspondentes, se diria que elas se distanciam em 7 unidades.

Um exemplo de um modelo PPAM é mostrado no Quadro 5, após a leitura da sequência 56857568. Neste modelo, m é o tamanho do contexto (equivalente ao k no Quadro 1), k é a medida de proximidade de contextos e <.> denota a união de contextos k-aproximados (e.g., 56 e 57 ou 85 e 75). O diferencial do PPAM em relação ao PPM se dá a partir de k = 1. A primeira entrada no Quadro 5 é a união dos contextos 56 e 57 (aproximação possível pelo fato de a distância entre eles ser 1), de modo que os símbolos que sucedem o contexto desta entrada são os símbolos que sucedem o contexto 56 (o símbolo 8) e 57 (o símbolo 5) na coluna anterior. O mesmo se observa na terceira entrada desta coluna, que é a

[23] A consulta a pelo menos outros dois profissionais da área se mostra importante para validar a seleção dos sinais utilizados.

união dos contextos 85 e 75 da coluna anterior, pelo fato de a distância entre eles também ser 1. Na coluna seguinte, que representa o modelo para k = 2, a última entrada é a união de todos os contextos distintos registrados, pois todos eles são próximos do contexto 75 para k =2.

Quadro 5 - Exemplo de um modelo PPAM após a leitura da palavra 56857568

Ordem m = 2, k = 0			Ordem m = 2, k = 1			Ordem m = 2, k = 2		
Predição	c	p	Predição	c	p	Predição	c	P
56 → 8	2	2/3	<.> → 5	1	1/5	<.> → 5	2	2/8
→ Esc	1	1/3	→ 8	2	2/5	→ 6	1	1/8
			→ Esc	2	2/5	→ 8	2	2/8
						→ Esc	3	3/8
68 → 5	1	1/2	<.> → 5	2	2/3	<.> → 5	2	2/6
→ Esc	1	1/2	→ Esc	1	1/3	→ 8	2	2/6
						→ Esc	2	2/6
85 → 7	1	1/2	<.> → 6	1	1/4	<.> → 6	1	1/4
→ Esc	1	1/2	→ 7	1	1/4	→ 7	1	1/4
			→ Esc	2	2/4	→ Esc	2	2/4
57 → 5	1	1/2	<.> → 5	1	1/5	<.> → 5	2	2/8
→ Esc	1	1/2	→ 8	2	2/5	→ 6	1	1/8
			→ Esc	2	2/5	→ 8	2	2/8
						→ Esc	3	3/8
75 → 6	1	1/2	<.> → 6	1	1/4	<.> → 5	2	2/10
→ Esc	1	1/2	→ 7	1	1/4	→ 6	1	1/10
			→ Esc	2	2/4	→ 7	1	1/10
						→ 8	2	2/10
						→ Esc	4	4/10

FONTE: Zhang e Adjeroh (2008)

Para a pesquisa ora registrada, seria necessária a definição do limiar de aproximação dos símbolos, que auxilia na definição de "em quais entradas uma dada sequência deve ser armazenada". Informações com nível de proximidade igual ou menor que o limiar definido devem ser incorporadas a uma única entrada; caso contrário, devem ser destinadas a entradas distintas. Quanto maior o limiar, maior a quantidade de informações reunidas, porém é maior também a quantidade de informações perdidas.

A estratégia de utilização do PPAM foi considerada, porém logo em seguida descartada, devido a contradições na explicação da técnica e à

falta de esclarecimentos convincentes dos autores acerca das dúvidas que surgiram durante a leitura do documento no qual Zhang e Adjeroh (2008) descreveram a técnica. O principal esclarecimento necessário é ilustrado pelo exemplo a seguir: sejam $k > 0$ e C um conjunto de contextos distintos (e.g., 56 e 57). É possível perceber, analisando o modelo do Quadro 5, que os contextos pertencentes a C podem estar presentes em diferentes entradas no modelo (no caso do modelo de exemplo, para $k = 2$, os contextos 56 e 57 somente não estão presentes na terceira entrada). Quando encontrado um contexto pertencente a C na etapa de testes, que consiste em consultas às probabilidades armazenadas no modelo para a tomada de decisão, qual entrada deve ser considerada?

4.1.2.5 Classificação Unária

Por fim, a última abordagem descartada consistia em alimentar um modelo PPM com vetores de características e procurar estabelecer um limiar com o qual fosse possível separar duas classes de arquivos, isto é, caso a Razão de Compressão (RC) obtida em uma compressão fosse maior que o limiar utilizado, significaria que o arquivo de testes pertencia à determinada classe; caso contrário, à outra classe testada. Essa abordagem foi baseada em Marinus (2010), é exibida no diagrama em blocos da Figura 37 e foi descartada pelo fato de não ter sido encontrado um limiar que fizesse a devida separação entre duas classes de arquivos de sinais de vozes patológicas. Quando experimentando com sinais de vozes normais, foi possível perceber que não houve mistura entre as Razões de Compressão obtidas: na compressão de um arquivo de teste com o modelo ao qual se sabia que pertencia, qualquer que fosse ele, era obtida uma alta RC; na compressão com o outro modelo, uma baixa RC.

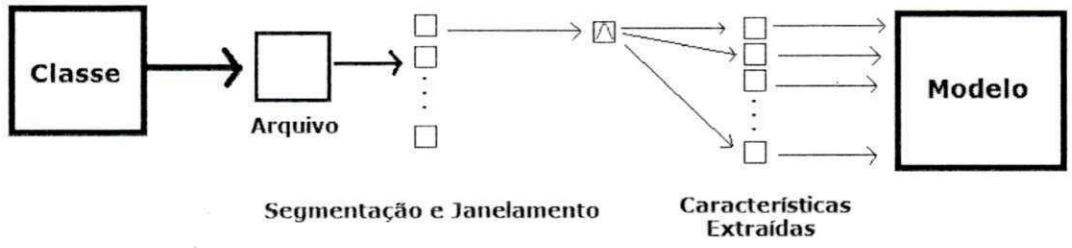
A utilização da métrica Razão de Compressão como base para a tomada de decisão foi fundamentada em diversos trabalhos que fazem uso do PPM com fins de classificação, a exemplo de Coutinho et al. (2005), Barufaldi et al. (2009) e Medeiros et al. (2011).

Figura 37 - Modo unário de classificação

4.1.3 Abordagem Selecionada

A abordagem selecionada, elaborada após o descarte do PPAM, consiste em utilizar o PPM como classificador tomando-se como base a predição ao invés da compressão e alimentando-se o modelo com vetores de características extraídas de segmentos dos sinais de voz, como mostrado na Figura 38.

Figura 38 - Diagrama de blocos da abordagem de alimentação por vetores de características



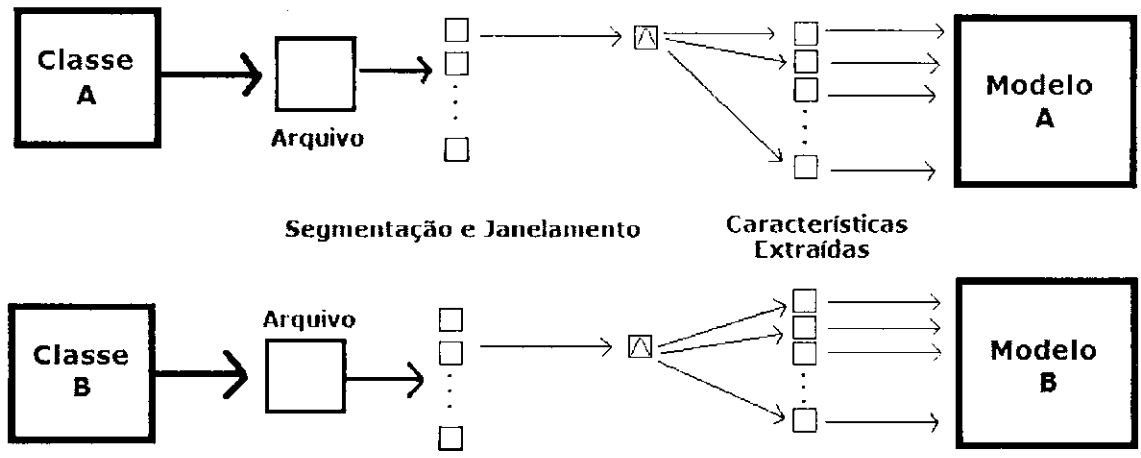
Esta abordagem consiste em segmentar o arquivo em quadros de 20 ms, com superposição de 50% (procedimentos explicados em mais detalhes na Seção 4.2.1), executar um procedimento de janelamento em cada quadro, extrair medidas desta janela e alimentar o modelo com estas medidas, semelhante ao que é feito em diversas outras pesquisas da área, a exemplo de Marinus (2010), Tavares et al. (2011) e Aguiar Neto et al. (2007). O diferencial desta pesquisa em relação às demais é que são utilizados parâmetros temporais, acústicos e estatísticos (os quais são devidamente explicados nas Seções 2.3.1 a 2.3.3) e o classificador é o PPM (apresentado na Seção 2.4).

O vetor de características montado é composto de: Energia, Taxa de Cruzamento por Zero, Número Total de Picos, Diferença no Número de Picos, Entropia, Frequência Fundamental, *jita*, *jitt*, *rap*, *ppq5*, *shim*, *ShdB*, *apq3*, *apq5* e HNR.

4.2 Execução do Experimento

Na execução de uma classificação, são construídos dois modelos, cada um referente a uma das classes de arquivos consideradas, como mostrado na Figura 39.

Figura 39 - Construção dos modelos utilizados em uma classificação



No experimento completo foram consideradas 4 (quatro) classes, totalizando 189 arquivos (13,6% do total da base). A distribuição entre as classes não é regular, uma vez que a menor classe de arquivos originais dispõe de 34 arquivos e a maior de 56. Informações adicionais acerca das referidas classes estão disponíveis no Apêndice B.

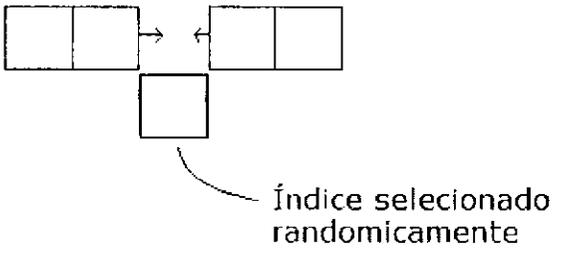
Nas Seções 4.2.1 a 4.2.4 são descritas as etapas do experimento, que incluiu: (i) a investigação do melhor tipo de entrada para cada classificação; (ii) a investigação dos impactos de atividades de pré-processamento e variação do tamanho máximo do contexto do classificador; e (iii) a obtenção dos percentuais de classificação via validação cruzada utilizando as melhores configurações.

4.2.1 Execução de uma classificação

Nas classificações executadas antes da caracterização de fato do classificador via Validação Cruzada (ver Seção 4.2.4), a divisão entre treinamento e teste foi feita conforme a descrição de Medeiros et al. (2011). Assim, cada classe foi particionada percentualmente, sendo 60% destinada ao treinamento e 40% ao teste, indiscriminadamente.

Visando a evitar qualquer tipo de viés, a seleção dos arquivos para treinamento foi feita randomicamente. Os arquivos pertencentes a uma mesma classe (armazenados no mesmo diretório) foram organizados em uma lista e a seleção de um índice da lista, i.e., de um arquivo, foi feita com base em um algoritmo padrão de geração de números randômicos. O arquivo selecionado foi posteriormente removido da lista, conforme ilustrado na Figura 40.

Figura 40 - Seleção de um arquivo para treinamento

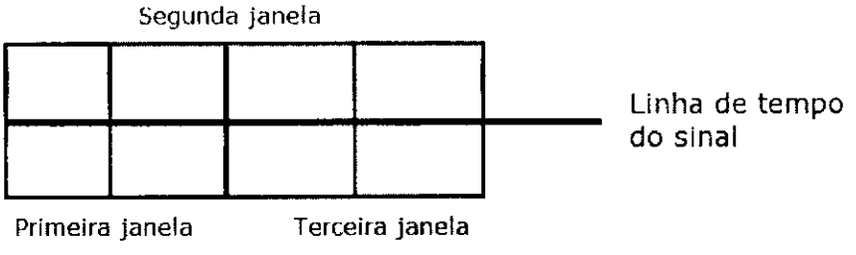


Em seguida, o referido arquivo foi submetido a uma etapa de pré-processamento. Na maioria das pesquisas, esta etapa consiste em pré-ênfase, segmentação e janelamento do sinal. Porém, inspirado em Marinus (2010), investigou-se adicionalmente nesta pesquisa a influência da aplicação do filtro de pré-ênfase no melhoramento dos resultados (além de outras atividades de pré-processamento; ver Seção 4.2.3), considerando que Zwetsch et al. (2006) não a recomendaram por prejudicar a discriminação de patologias, devido ao efeito de suavização da excitação glotal que a pré-ênfase produz.

A segmentação consistiu na divisão do arquivo em quadros temporais de 20 ms. Na área de processamento de voz, são usualmente

utilizados intervalos entre 16 e 32 ms, pelo fato de ser a voz estacionária neste intervalo, i.e, suas propriedades estatísticas não variarem no intervalo mencionado, mas somente a partir de intervalos maiores (RABINER; SCHAFER, 1978; SOTOMAYOR, 2003; RIBEIRO, 2003). É válido mencionar ainda que, visando a minimizar os efeitos da descontinuidade entre segmentos, é recomendável aplicar a superposição entre segmentos adjacentes: os últimos Bytes de um segmento devem ser iguais aos primeiros Bytes do segmento vizinho (FECHINE, 2000). No caso da superposição em 50%, utilizada nesta e em diversas outras pesquisas (e.g., as pesquisas conduzidas por Costa (2008), Fechine (2000) e Marinus (2010)), a última metade de um segmento corresponde à primeira metade do segmento seguinte, conforme ilustrado na Figura 41.

Figura 41 - Segmentação com sobreposição de 50%



Por fim, o janelamento consiste na transformação dos Bytes de um segmento, visando a minimizar a adversidade dos efeitos advindos da segmentação abrupta que causa descontinuidades no espectro do sinal de voz (FECHINE, 2000). Dentre os algoritmos existentes, os mais comuns são o retangular, Hamming e Hann. O janelamento retangular não altera o segmento. Os outros dois consistem em uma transformação cossenoidal, visando a minimizar os efeitos nas extremidades e mantendo os efeitos no centro. Esses algoritmos são bastante similares (se diferenciam apenas em duas constantes). Nas Figuras 42 e 43, ambas as janelas foram extraídas a partir da utilização do MATLAB® (Mathworks Inc., 2009), sendo possível observar que a diferença visível ocorre apenas na base dos gráficos (apenas um deles toca o eixo de origem).

Figura 42 - Forma de onda da janela de *Hamming*

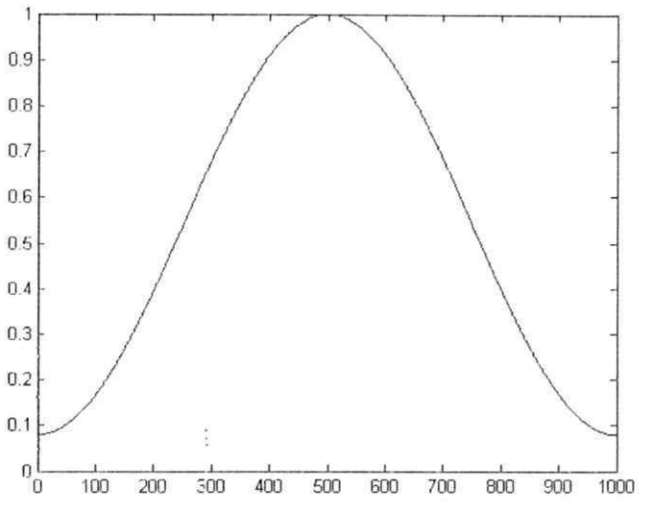
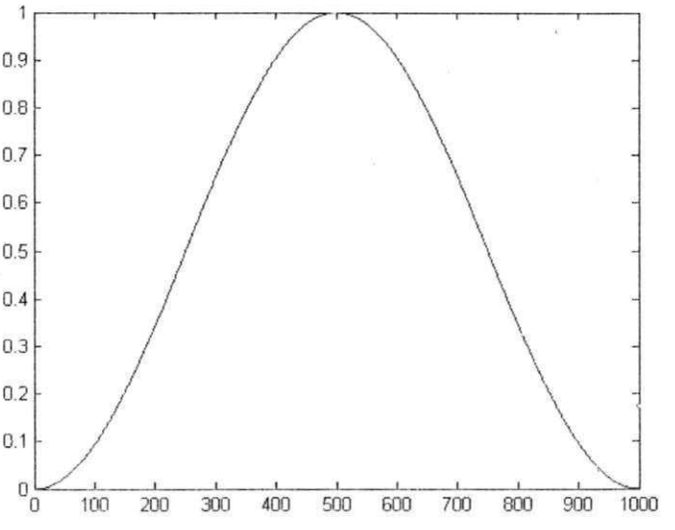


Figura 43 - Forma de onda da janela de *Hann*

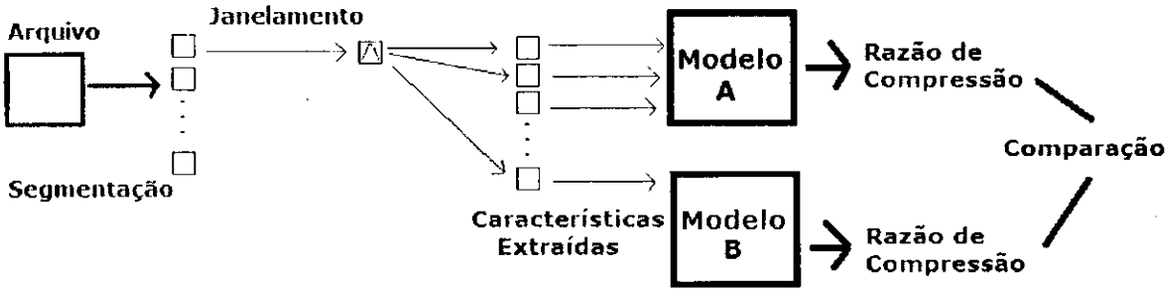


Em ambas as figuras, o eixo das abscissas representa as amostras do sinal (neste caso, da janela) e o eixo das ordenadas, o valor da amostra (o qual é multiplicado pelo valor correspondente da amostra do sinal de voz).

Em seguida, foram extraídas medidas de cada quadro, de modo a caracterizar o sinal com fins à tomada de decisão sobre a presença ou ausência de uma dada patologia. Os modelos relacionados a uma classe de arquivos foram alimentados pelas medidas extraídas, de modo a caracterizá-lo unicamente, no tocante aos demais. Esta etapa corresponde ao treinamento, sumariado graficamente na Figura 39.

Na etapa de testes, todo o processo ocorreu de modo similar. Contudo, os modelos tornaram-se estáticos, i.e, as probabilidades armazenadas não eram atualizadas e nenhuma entrada no modelo era criada. A ideia era que os modelos nada aprendessem sobre os arquivos de teste. Esta é mais uma forma de evitar vieses. Além disto, nesta etapa foi utilizado o codificador aritmético, que gerava resultados que podiam ser utilizados na tomada de decisão sobre a qual classe de arquivos pertencia um dado arquivo de teste. Os codificadores aritméticos geravam fluxos, que eram comparados em tamanho com o fluxo original. Ao modelo que gerava o menor fluxo pertencia o arquivo testado. Este processo é sumariado graficamente na Figura 44.

Figura 44 - Processo de teste



É válido ressaltar que o sistema utilizado para a execução das classificações foi desenvolvido especificamente para esta pesquisa. Isto significa que a seleção dos arquivos para treinamento e testes e estas etapas propriamente ditas, a segmentação dos sinais, a composição do vetor de características e o próprio PPM, por exemplo, foram etapas executadas por um sistema próprio, não tendo sido utilizados sistemas de terceiros.

4.2.2 Identificação do melhor tipo de entrada para cada classificação

O objetivo da investigação conduzida nesta etapa foi identificar os melhores tipos de entrada para cada classificação executada. Foram considerados quatro tipos de entrada: parâmetros temporais (ver Seção

2.3.1), parâmetros acústicos (ver Seção 2.3.2), parâmetro estatístico (ver Seção 2.3.3) e Análise Preditiva Linear (coeficientes LPC; ver Seção 2.3.2.5). Embora os coeficientes LPC sejam considerados parâmetros acústicos, semelhantemente ao *Jitter*, *Shimmer* e Frequência Fundamental, por exemplo, percebeu-se que eles são incompatíveis com os parâmetros citados, pelo fato de a Análise Preditiva Linear fornecer diversos coeficientes para cada quadro do sinal e os outros parâmetros, por sua vez, fornecerem um único valor representativo de cada quadro. Por essa razão, os coeficientes LPC foram investigados em separado dos outros parâmetros acústicos.

A identificação do melhor tipo de entrada se deu mediante a utilização de Projeto Experimental com um único fator, conforme descrito por Jain (1991) e de forma resumida no Anexo A, sendo o fator o tipo de entrada que alimenta os modelos e os níveis deste fator os diversos tipos extraídos e combinações entre eles. A utilização desta ferramenta fornece como resultado a quantificação do impacto de cada nível do fator analisado (o *efeito* de cada fator) nos resultados, isto é, o quanto os aumenta ou os diminui, o que permite a identificação do melhor tipo de entrada, que se trata do nível que retorna os melhores resultados.

4.2.3 Investigação dos Impactos dos Processamentos e Contextos

Nesta etapa, objetivou-se investigar o impacto de atividades de pré-processamento e reorganização dos sinais da base sobre os resultados, visando a analisar a viabilidade de sua aplicação. Trata-se das seguintes atividades: pré-ênfase, distinção de gêneros (utilização apenas de sinais de vozes masculinas e femininas em uma classificação) e subamostragem, que consiste em utilizar apenas a metade das amostras dos sinais que são representados em 50 kHz de frequência de amostragem (50 mil amostras por segundo), de modo a igualar a frequência de amostragem de todos os sinais utilizados em 25 kHz.

Paralelamente, nesta etapa também foi conduzida a investigação sobre o efeito da variação do tamanho máximo de contexto do modelo

PPM nos resultados, visando a investigar a viabilidade de aumentar o tamanho do modelo durante o procedimento. As classificações executadas até então foram feitas utilizando contexto 0, i.e, foi registrada apenas a quantidade de ocorrências de cada valor de medida extraído. Nesta etapa, foi investigada também a inclusão de sequências (de tamanho entre 1 e 4) de valores nos modelos e sua influência nos percentuais de acerto obtidos.

As classificações executadas nesta etapa utilizaram as entradas identificadas na etapa anterior. Por exemplo, caso na etapa anterior tenha sido observado que os melhores percentuais de acerto na classificação Normal x Edema são obtidos pelo emprego apenas de coeficientes LPC, então nesta etapa os modelos foram alimentados apenas com coeficientes LPC.

A investigação nesta etapa se deu mediante a comparação par a par entre dois casos. Retomando o exemplo anteriormente apresentado, os resultados obtidos na classificação Normal x Edema utilizando como entrada apenas coeficientes LPC foram comparados par a par com os resultados obtidos ao aplicar filtro de pré-ênfase, distinção de gênero e subamostragem e ao utilizar tamanhos de contexto entre 1 e 4. A comparação foi feita utilizando testes de comparação de resultados como Intervalos de Confiança (um procedimento a que Jain (1991) se referiu como *t-test*, o qual é explicado em detalhes no Anexo B) ou teste de Mann-Whitney, a depender da distribuição amostral dos percentuais. A utilização destes testes é justificada pelo fato de que o intuito desta investigação era verificar se a aplicação de atividades de pré-processamento ou aumento do tamanho do modelo forneciam resultados significativamente maiores que os resultados obtidos nas classificações sem essas etapas.

Segundo Jain (1991), os intervalos de confiança apenas devem ser utilizados se ambos os conjuntos de dados comparados tiverem tamanho maior que 30 ou apresentarem distribuição Normal. Porém, em cada classificação eram coletados apenas 5 percentuais de acerto. Caso pelo

menos um dos conjuntos comparados não apresentasse distribuição Normal[24], a comparação era feita utilizando teste de Mann-Whitney (não-paramétrico).

4.2.4 Obtenção dos Percentuais via Validação Cruzada

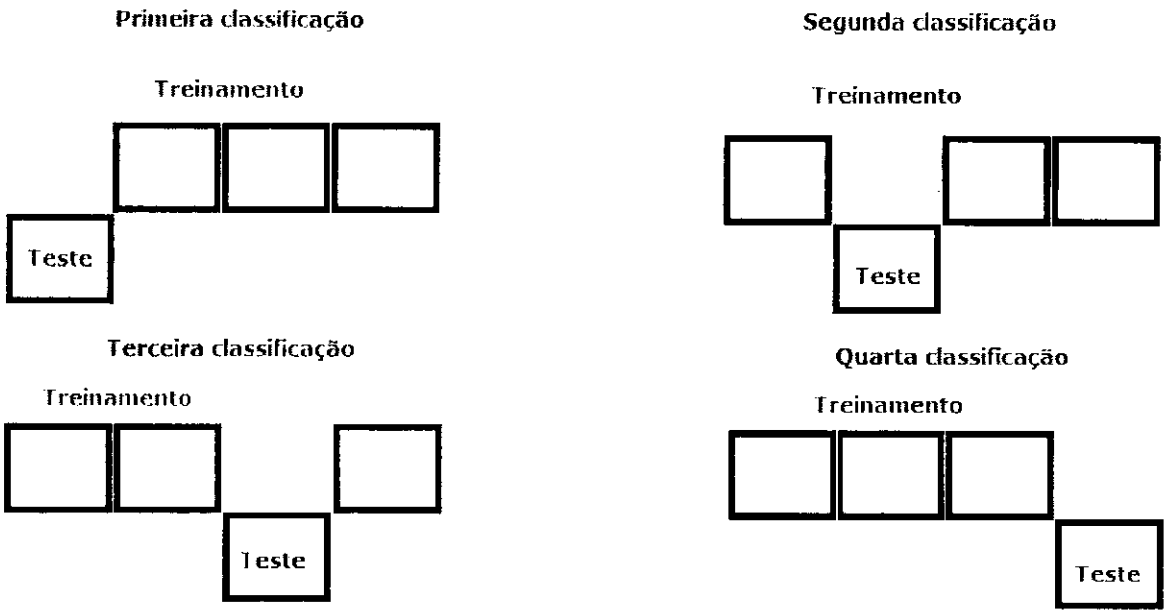
Tendo sido encontrada a melhor configuração do experimento (o que incluiu o tipo de entrada que alimenta os modelos, a execução ou não de atividades de pré-processamento e o tamanho de contexto utilizado), objetivou-se, nesta etapa, a obtenção de percentuais que avaliassem de fato o desempenho do PPM como classificador de patologias da fala. Nesta etapa, foi adotada uma metodologia bastante conhecida, denominada Validação Cruzada (*Cross Validation*) (MOSTELLER; TUKEY, 1968), que consiste em dividir uma classe de arquivos em N partes iguais e executar N classificações considerando essa classe. A cada rodada, utiliza-se uma parcela diferente nos testes, ou seja, na primeira rodada, os arquivos da primeira parcela são empregados nos testes e o restante no treinamento; na segunda, os arquivos da segunda parcela são empregados nos testes e assim por diante. Uma esquematização desta abordagem de treinamento é mostrada na Figura 45.

Nesta pesquisa, foi empregada uma validação cruzada com 4 parcelas, pelo fato de se ter percebido ser esta a quantidade mínima de parcelas que mantinha quantidade suficiente (estatisticamente significativa) de sinais de voz para serem usados nos testes. Isso porque são utilizados 34 sinais de vozes com Edema. Ao ser feita a divisão em 4 (quatro) parcelas, cada uma dispunha de 8 (oito) arquivos, sobrando 2 (dois). Uma quantidade de parcelas maior que esta não possibilitaria a manutenção de uma quantidade suficiente de arquivos por parcela para os testes.

---

[24] A verificação da distribuição amostral dos conjuntos de percentuais foi feita utilizando um teste de hipóteses de Shapiro-Wilk (SHAPIRO; WILK, 1965) a 95% de significância, recomendado para pequenos conjuntos.

Figura 45 - Validação Cruzada com 4 parcelas



4.3 Ferramentas Utilizadas

Para a execução de todo o processo descrito neste capítulo, incluindo as abordagens de manutenção do modelo descritas na Seção 4.1.2 e o cálculo de parte das características consideradas, foi utilizada a plataforma Java, por intermédio da IDE (*Integrated Development Environment* - Ambiente de Desenvolvimento Integrado) Netbeans. Para o cálculo das características restantes, foi utilizada a ferramenta Matlab®, de forma integrada ao sistema em Java. Para análise estatística dos resultados, que fundamentaram as tomadas de decisões e a condução da pesquisa, foi adotado o software R.

Entre as ferramentas descartadas no decorrer da pesquisa se encontram o SGBD PostgreSQL e a ferramenta de administração pgAdmin III, utilizados nos estágios iniciais da pesquisa quando estava em experimentação a abordagem de manutenção do modelo em banco de dados.

4.4 Discussão

Neste capítulo, foi apresentada a evolução das abordagens metodológicas pesquisadas, culminando na abordagem metodológica adotada, cuja

contribuição se caracteriza pela combinação de elementos de áreas distintas, como a compressão de dados com PPM e a análise de sinais de voz, seja com parâmetros temporais, acústicos ou estatísticos. Conforme descrito na Seção 4.2, os modelos PPM, após criados, foram alimentados com essas características e a decisão foi tomada com base na ocorrência dos valores obtidos, utilizando o codificador aritmético. Primeiramente, foi investigado o tipo de entrada que fornece os melhores resultados para cada classificação. Em seguida, utilizando as entradas identificadas em cada caso, foi investigada a influência das atividades de pré-processamento e variação do tamanho máximo de contexto sobre os resultados obtidos. Por fim, foram obtidos resultados que caracterizavam o PPM de fato, por meio de validação cruzada.

No Capítulo 5, são apresentados os resultados obtidos ao utilizar a metodologia descrita.

Capítulo 5

Apresentação e Discussão dos Resultados

Neste capítulo, são apresentados e discutidos os resultados obtidos a partir da abordagem concebida ao longo da pesquisa ora documentada. É válido mencionar que, transcendendo a apresentação de porcentuais obtidos, investigam-se e discutem-se os efeitos da utilização de diferentes tipos de entradas na alimentação dos modelos, aplicação de pré-ênfase, distinção entre gêneros, subamostragem dos sinais e diferentes tamanhos máximos de contexto. Tal investigação teve o propósito de identificar a melhor configuração de classificação para cada caso de confronto entre duas classes. Em seguida, foram obtidos percentuais de acerto que caracterizassem o desempenho do método PPM no processo de classificação de padrões patológicos por meio de validação cruzada.

Todos os processamentos estatísticos foram executados utilizando o ambiente de *software* R, destinado à computação estatística de dados e à construção de gráficos.

5.1 Base de Dados

Nesta pesquisa, foi utilizada a base de dados da Kay Elemetrics, gravada no Voice Speech Lab da Massachussets Eye and Ear Infirmary (MEEI). A gravação foi feita com distância fixa do microfone e em ambiente controlado (livre de ruído externo), de modo que não foi necessária a aplicação de técnicas de eliminação de ruído de fundo. A principal fonte de ruído é a própria voz patológica e o ruído proveniente dessa fonte deve ser considerado. As patologias foram diagnosticadas (para registro na base) após extensa análise dos sistemas fonadores dos pacientes, o que incluiu a aplicação de estroboscopia, análise aerodinâmica e análise acústica.

A base é composta por 1381 arquivos, em formato NSP[25], com 16 bits/amostra, dos quais 666 contêm o pronunciamento da vogal */ah/* sustentada. Embora também sejam oferecidos sinais de voz no pronunciamento dos 12 primeiros segundos da *Rainbow Passage* (um texto bastante conhecido nos Estados unidos para testar a habilidade de um indivíduo de produzir texto corrido), foi utilizado apenas o subconjunto mencionado pelo fato de a extensa maioria dos trabalhos relacionados na área também ter utilizado este subconjunto da base de dados adotada (conforme apresentado na Seção 3.1), o que dá mais confiança e relevância à comparação entre os resultados. Além disso, as dobras vocais vibram durante esta elocução, o que facilita a análise do comportamento do sistema fonador durante esse processo na presença de uma patologia (COSTA, 2008; MONTEIRO et al., 2011).

Dos 666 arquivos que contêm o pronunciamento sustentado, 53 correspondem a sinais de vozes saudáveis (cada um dos quais com aproximadamente 3s de duração) e o restante corresponde a sinais de vozes patológicas (cada um dos quais com duração em torno de 1s).

São disponibilizados arquivos de vozes contendo mais de 30 patologias distintas, porém, sem regularidade de representação entre elas. Isso significa que algumas patologias, tais como a Paralisia e o Edema, são representadas por dezenas de arquivos (razão pela qual foram consideradas nesta pesquisa) e muitas outras são representadas por menos de 10 arquivos, número considerado sem significância estatística, pelo fato de não fornecer uma quantidade adequada de segmentos, além de não haver significância estatística suficiente para que se tenha uma boa estimativa de erro.

É importante salientar que não há unanimidade entre as frequências de amostragem dos arquivos. Todos os arquivos contendo sinais de vozes saudáveis têm frequência de amostragem de 50 kHz, ou seja, 50 mil amostras por segundo. Os arquivos contendo sinais de vozes patológicas,

[25] Formato proprietário do *Computerized Speech Lab* (CSL) da *Kay Elemetrics*, a empresa proprietária da base de dados utilizada nesta pesquisa.

por sua vez, têm frequência de amostragem de 25 kHz, mas alguns também têm 50 kHz. Uma vez que as medidas utilizadas são extraídas a partir de quadros que correspondem a um intervalo de tempo, é essencial que no processamento destes sinais (pré-ênfase, janelamento e extração de medidas), esta característica seja levada em conta de modo uniforme, não importando o tamanho em Bytes do quadro, a fim de que os valores extraídos sejam válidos. Por exemplo, na extração do *Jitter* de um quadro de 20 ms, se o sinal gravado no arquivo tiver sido amostrado a 50 kHz, o quadro considerado deve ser composto por 1.000 amostras. Contudo, se tiver sido amostrado a 25 kHz, ele deve ser composto por 500 amostras.

5.2 Identificação do melhor tipo de entrada

Nesta etapa da investigação, objetivou-se identificar o melhor tipo de entrada para cada caso de classificação. Conforme mostrado no Quadro 6, descobriu-se que não há um tipo de entrada que retorne os melhores percentuais em todos os casos. O melhor tipo de entrada varia com a classificação executada.

É válido ressaltar que quase todos os melhores resultados são obtidos quando empregando parâmetros temporais, isoladamente ou de forma combinada, principalmente quando a classificação objetiva a detecção da presença de uma patologia (quando envolve sinais de vozes Normais), o que pode indicar sua utilidade na classificação de patologias.

Em todas as classificações com o intuito de detectar sinais de vozes Normais, foram obtidos resultados semelhantes ao utilizar parâmetros isolados e a combinação de parâmetros. Nesses casos, recomenda-se a utilização de parâmetros isolados, com o intuito de economizar memória, de modo a tornar mais fácil a implantação do sistema em dispositivos de menor recurso de armazenamento e, principalmente, comprometer menos o desempenho do sistema em que está sendo utilizado.

Em particular, na classificação Normal x Patológico, foram obtidos resultados semelhantes ao utilizar parâmetros temporais (Energia, TCZ,

NTP e DNP) e ao utilizar Entropia. Neste caso, pela mesma razão, recomenda-se a utilização exclusiva da entropia.

Quadro 6 - Percentuais obtidos e tipos de entrada utilizados

Classificação	Tipos de entrada (1ª classe)	Tipos de entrada (2ª classe)
Normal x Patológico[26]	Temporais; Entropia; Temporais e Entropia	Temporais e Entropia
Normal x Edema	Temporais; Temporais e Entropia	Temporais e Entropia
Normal x Paralisia	Temporais; Temporais e Entropia	Temporais
Normal x Outras[27]	Temporais; Temporais e Entropia	Temporais
Edema x Paralisia	LPC	Temporais e Acústicos
Edema x Outras	LPC	Acústicos
Paralisia x Outras	Temporais	Temporais, Acústicos e Entropia

5.3 Investigação do impacto de atividades de pré-processamento e variação do tamanho do contexto

A segunda etapa da investigação consistiu em analisar o efeito de atividades de pré-processamento nos percentuais de acerto obtidos, conforme descrito na Seção 4.2.3. Além disto, foi investigado também,

[26] Classe que contém todos os sinais de voz com alguma patologia, incluindo Edema e Paralisia.
[27] Classe que contém sinais de vozes patológicas, mas não incluindo Edema e Paralisia.

em separado, o efeito da variação do tamanho máximo de contexto armazenado pelo modelo PPM no treinamento (ver Seção 2.4).

Os resultados da primeira investigação podem ser resumidos em: nenhum dos processamentos empreendidos acarretaram impacto positivo nos resultados; os resultados das classificações com atividades de pré-processamento foram significativamente menores ou não houve diferença com significância estatística.

No caso da pré-ênfase, não houve surpresa com os resultados obtidos, devido ao efeito de suavização da excitação glotal produzido, conforme disse Zwetsch et al. (2006), o que pode prejudicar resultados de classificação. No caso da subamostragem, isso se deve à alteração na qualidade do sinal produzida por esse processamento (perda de metade da informação carregada pelo sinal), o que pode prejudicar, por conseguinte, a extração de medidas e, consequentemente, as classificações que se utilizam destas.

Por fim, com relação à distinção de gêneros, foi constatada uma quantidade reduzida de sinais de vozes masculinas na base adotada, o que ocasionou alta instabilidade nos percentuais obtidos (e.g., 0% em todas as repetições de um caso de classificação ou 83% e 33% em outro caso), devido ao número mínimo de arquivos disponíveis para teste - a classificação errada de apenas 1 sinal pode reduzir drasticamente o percentual. Sendo assim, pode-se afirmar que estes resultados não apresentam nível de confiança significativa, sendo, por essa razão, ignorados. No caso dos sinais femininos, não houve melhora significativa que justificasse a adoção deste artifício.

Com relação à variação dos contextos, observou-se que o aumento no tamanho do contexto, em nenhum dos casos, implicou percentuais mais elevados: o impacto foi nulo (sem diferença estatisticamente significante) ou negativo (resultados significativamente piores). É uma constatação semelhante àquela de Souza (2008) e de Farias (2010), que constataram que o aumento do tamanho do contexto implicou pequeno aumento da Razão de Compressão, sendo mais viável utilizar tamanhos

de contextos menores, já que o ganho obtido com contextos maiores não compensava o aumento do uso de recursos computacionais exigidos.

5.4 Caracterização do Classificador por Validação Cruzada

Tendo sido encontradas as configurações melhores e mais viáveis para cada caso de classificação, as quais incluem alimentar os modelos com as categorias de indicadores mostradas no Quadro 6, não aplicar nenhum pré-processamento (é recomendável a repetição desta investigação com outra base de dados mais representativa de ambos os gêneros) e utilizar contexto 0, foi utilizado o procedimento de Validação Cruzada, a fim de obter o principal resultado desta pesquisa: os percentuais de acerto para cada um dos casos de classificação, de modo a caracterizar o PPM como classificador de patologias da fala quanto à eficácia.

Na avaliação do método PPM como classificador de patologias, foram utilizadas algumas medidas de avaliação, as quais serão explicadas a seguir:

- Correta Rejeição (CR): mede a capacidade do classificador de indicar corretamente que a patologia não está presente. O complemento dessa medida é chamado de Falsa Aceitação (FA);
- Correta Aceitação (CA): mede a capacidade do classificador de indicar corretamente que a patologia está presente. O complemento dessa medida é a Falsa Rejeição (FR);
- Eficiência: representa a taxa de classificação correta do classificador, isto é, sua eficiência na indicação da presença e da ausência de uma patologia. É calculada pela Equação 14.

$$E = \frac{CR + CA}{CR + CA + FR + FA} * 100\% \qquad (14)$$

Os resultados são sumariados no Quadro 7.

Quadro 7 - Percentuais de cada classificação

Classificação	Correta Rejeição (mediana %)	Correta Aceitação (mediana %)	Eficiência (%)
Normal x Patológico	100,0	94,1	97,05
Normal x Edema	100,0	95,0	97,5
Normal x Paralisia	100,0	92,3	96,15
Normal x Outras	100,0	96,4	98,2
Edema x Paralisia	84,6	50,0	67,3
Edema x Outras	75,0	40,0	57,5
Paralisia x Outras	64,2	57,6	60,9

A coluna "Correta Rejeição", nas classificações que envolvem sinais de vozes normais, se refere aos casos em que a presença de patologias foi descartada (rejeitada), enquanto a coluna "Correta Aceitação" se refere aos casos em que a presença de patologias foi confirmada (aceita) na classificação. Quando a classificação não envolve sinais de vozes normais (3 últimas linhas), essas medidas denotam os casos em que a primeira patologia foi rejeitada e confirmada, respectivamente. Isso significa que na classificação Edema x Paralisia, a Correta Rejeição se refere aos casos em que o Edema foi corretamente rejeitado, enquanto a Correta Aceitação se refere aos casos em que sua presença foi corretamente detectada. A eficiência, por fim, representa a média entre corretas aceitação e rejeição, ou seja, a eficácia geral do classificador.

É válido ressaltar, por fim, que a escolha da mediana como índice de tendência central das porcentagens apresentadas no Quadro 7 foi embasada por Jain (1991), que afirmou que a mediana deve ser escolhida quando os dados não são categóricos (nominais), o total dos dados obtidos não é de interesse e a distribuição de probabilidade dos dados é enviesada, caso dos dados em questão.

5.5 Caracterização da Eficiência do PPM

A eficiência do PPM na classificação de patologias da fala em sinais de voz pode ser analisada quanto à velocidade de execução e à utilização de recursos computacionais, mais especificamente os recursos de memória.

Quanto à velocidade de execução, foram analisados os tempos de execução de 10 classificações com as classes mais bem representadas da base de dados: Normal e Patológico (que inclui todas as patologias consideradas nesta pesquisa), com 53 e 56 arquivos cada uma. Estas classes foram aquelas escolhidas para a análise da eficiência por exigirem mais arquivos, tanto no treinamento quanto nos testes, de modo que os tempos para as outras classificações serão certamente menores do que aqueles relatados a seguir, no Quadro 8. Os intervalos apresentados correspondem aos intervalos de confiança a 95% de significância estatística dos tempos de execução do treinamento e de testes individuais.

Quadro 8 - Tempos de execução da classificação Normal x Patológico

Etapa	Tempo médio (ms)	Intervalo de confiança (ms)
Treinamento	**421**	**[405,3; 436,6]**
Testes individuais	**11,35**	**[10,8; 12,2]**

Com relação ao uso de memória, 10 execuções da mesma classificação foram analisadas e as quantidades máximas de memória utilizadas nas etapas de treinamento foram registradas, pelo fato de ser esta a etapa na qual ocorre maior utilização de memória, haja vista que os testes não alteram o modelo. O intervalo de confiança a 95% de significância destes registros foi extraído, tendo sido obtido o intervalo [15,02; 15,77] MB. A ferramenta de monitoramento utilizada foi VisualVM, fornecida pela distribuidora da plataforma Java.

Para a obtenção destes valores de desempenho, foi utilizada uma máquina com 6 GB de memória principal e um processador Intel Core i5, executando os sistemas operacionais Windows e Linux (Ubuntu). Não é possível compará-los aos valores obtidos pelo uso de outras abordagens, pelo fato de esse tipo de análise não ser reportada nos relatos analisados na revisão de literatura.

5.6 Discussão

É possível perceber, observando o Quadro 7, que os melhores resultados foram oriundos das classificações que envolviam sinais de vozes Normais, consideradas com o intuito de detectar a presença de patologias. Nelas, obtiveram-se resultados entre 92,3 e 100%. Porém, tais resultados apenas confirmam o potencial do PPM no contexto da classificação de padrões vocais, haja vista que resultados semelhantes foram alcançados em pesquisas anteriormente conduzidas, tais como as relatadas por Costa et al. (2012), Costa (2008) e Marinus (2010).

Um resultado importante, almejado por pesquisadores da área, consiste na discriminação precisa de patologias distintas. Nesta pesquisa, foram obtidos resultados dessa natureza que podem ser considerados bons e medianos, a depender da classificação. Os piores resultados são oriundos das classificações em que o intuito é diagnosticar sinais de voz com Edema. Isto se deve, segundo Brandt (2012), à pouca representatividade desta patologia na base de dados utilizada, uma vez que dos 43 registros de arquivos de vozes com Edema, apenas 3 são de vozes que apresentam apenas Edema. Todos os outros apresentam outras patologias além desta, alguns chegando a apresentar outras 5 (cinco) patologias. Quanto mais patologias um registro de voz apresenta, menos esse registro representa a patologia principal, o que pode confundir os classificadores utilizados e comprometer seus resultados.

Contudo, bons resultados de discriminação de patologias também foram obtidos, especialmente aqueles relativos a classificações destinadas a discriminar outras patologias em confronto com sinais de vozes com

Edema, que se encontram entre 75 e 84,6%. Mesmo assim, eles podem ser considerados de pouca utilidade em um ambiente clínico real pois, conforme relatado nas Seções 2.3.1 e 2.3.2, o impacto que diferentes patologias têm sobre algumas das medidas utilizadas nesta pesquisa é praticamente o mesmo. Além disso, como mencionado na Seção 2.5, diferentes patologias apresentam vários sintomas em comum, tais como rouquidão e elocução vocal com ruído de fundo. Faz-se necessária, portanto, a busca de medidas que sejam afetadas de forma diferente para as diferentes patologias às quais o sistema de produção da fala humano é suscetível.

Em resumo, os resultados obtidos com a utilização do PPM na classificação de padrões vocais apenas mostram seu potencial neste contexto, sendo necessárias mais pesquisas com vistas a melhores resultados.

Capítulo 6

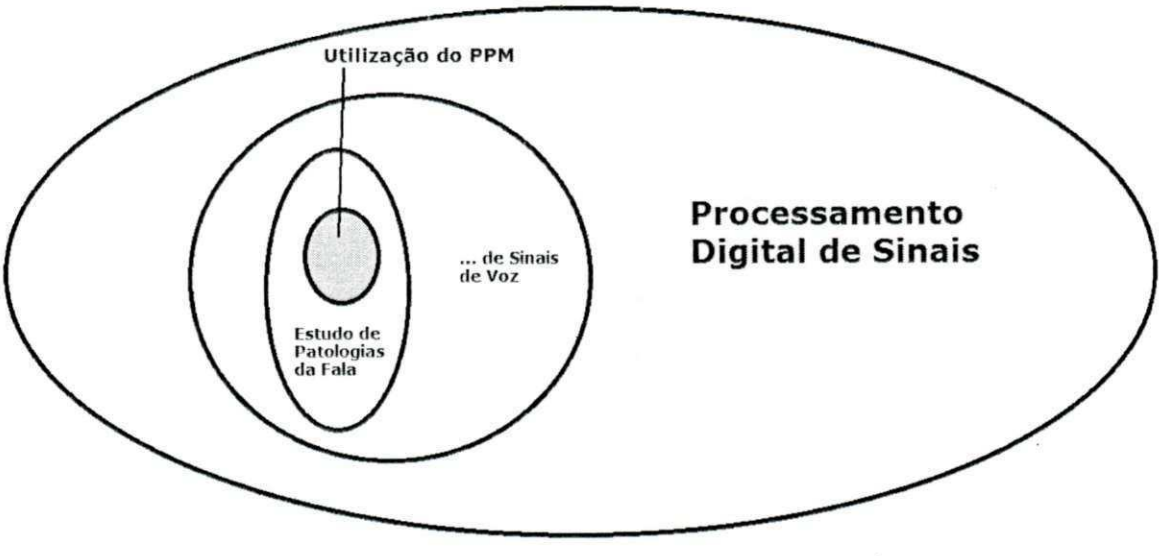
Considerações Finais

Neste capítulo, são apresentadas as considerações finais desta pesquisa, sob a forma de um sumário da pesquisa, de contribuições alcançadas e de sugestões para trabalhos futuros.

6.1 Resumo da Pesquisa

Esta pesquisa, conforme destacado na Figura 46, se insere em um ramo bem específico do Processamento Digital de Sinais. Trata-se, na verdade, de um novo contexto de utilização do método estatístico de compressão de dados PPM, constatada a partir da revisão de literatura realizada: a classificação de padrões vocais, com o intuito de diagnosticar patologias da fala.

Figura 46 - Diagrama de Venn que contextualiza o escopo da pesquisa



Conforme mostrado na Seção 3.3, o método PPM tem sido empregado na classificação de sinais de diferentes naturezas, tais como

imagens, áudio (reconhecimento de fala; ver Bao e Sim (1998)) e sinais de eletrocardiograma, mas sua aplicação no diagnóstico de patologias da fala não foi registrada na revisão bibliográfica levada a efeito no âmbito desta pesquisa.

Inicialmente, havia a intenção de discriminar o máximo de patologias possível, considerando, no âmbito do experimento, aquelas representadas em número suficiente de arquivos na base de dados utilizada (23 patologias). Porém, devido à informação, advinda de uma entrevista com uma fonoaudióloga, de que um pequeno subconjunto destas patologias, consideradas as principais, são causadoras de outras patologias ou afetam o sistema fonador em maior grau do que as demais (as quais podem ser caracterizadas como secundárias), foi utilizado um subconjunto contendo apenas 4 (quatro) classes (diagnósticos possíveis): Normal, Edema, Paralisia e Outras. Assim, esta pesquisa foi conduzida visando à discriminação precisa (pelo menos 80% de acerto) das 4 (quatro) classes de sinais de voz consideradas. A redução no número de classes de arquivos (e, consequentemente, no número de arquivos utilizados no experimento) resolveu o problema de utilização excessiva da memória ("estouro" de memória) que havia no início da pesquisa.

Nos estágios iniciais, a abordagem de utilização do classificador também era diferente daquela finalmente utilizada: os modelos eram alimentados com os Bytes extraídos diretamente dos arquivos. Devido aos resultados insatisfatórios obtidos nos experimentos preliminares, partiu-se para a alimentação do modelo com vetores de características. Somente a partir do emprego desta abordagem metodológica, que incluiu a composição dos vetores de características, a partir da extração das medidas apresentadas nas Seções 2.3.1, 2.3.2 e 2.3.3, bons resultados passaram a ser obtidos, principalmente nas classificações que envolviam registros de vozes Normais. Nestas, obteve-se 100% de acerto entre os sinais de vozes Normais (obtinha-se como resultado que as vozes destes sinais não apresentavam patologia - um indicador denominado Verdadeiro Negativo) e percentuais medianos, sempre acima de 90%, entre os sinais

de vozes patológicas. Nas classificações que envolviam exclusivamente registros de vozes patológicas, obtiveram-se percentuais medianos entre 40 (que ainda podem ser considerados baixos para ambientes clínicos reais) e 86,4%, levando a crer que os sinais de vozes patológicas se diferenciam entre si menos do que os sinais de vozes Normais se diferenciam de sinais de voz com alguma patologia, principalmente se a patologia apresentada pelo registro de voz for Edema, que correspondeu aos piores percentuais obtidos.

6.2 Contribuições da Pesquisa

A primeira contribuição advinda desta pesquisa consistiu na identificação das principais patologias, i.e., aquelas que mais afetam o sistema de produção da fala e que devem ser identificadas por sistemas de detecção automática de patologias. A maioria das pesquisas afins conduzidas nesta área considera somente algumas destas patologias, provavelmente pelo fato de serem aquelas mais comuns nas bases de dados disponíveis. Porém, com o conhecimento das patologias que mais afetam o sistema fonador, já se sabe que em trabalhos futuros não se deve criar uma classe de sinais de voz em que a patologia em comum apresentada por eles seja Hiperfunção ou Compressão Ventricular[28], por exemplo.

Outras das principais contribuições incluem: (i) a apresentação de um novo contexto de utilização do PPM (classificação de patologias da fala); (ii) a alimentação dos modelos PPM com medidas extraídas de segmentos destes sinais, de forma semelhante ao que fazem trabalhos similares utilizando outros classificadores; (iii) a identificação das melhores medidas, dentre as consideradas neste trabalho, para cada tipo de classificação executada; (iv) a obtenção de resultados bons ou promissores, a depender do tipo de classificação executada, a partir desta abordagem metodológica; (v) a obtenção de uma metodologia que

[28] Patologias representadas abundantemente na base de dados utilizadas, mas que são apenas derivadas de uma ou mais das principais patologias.

fornece bons resultados em um intervalo curto de tempo e com baixa utilização de recursos computacionais (ver Seção 5.5).

As duas primeiras contribuições refletem o caráter inovador desta pesquisa. A terceira, por sua vez, é mais importante, pois permitirá melhor direcionamento de pesquisas futuras, já que se sabe quais as medidas que melhor identificam uma dada patologia, em confronto com outra patologia ou com voz Normal. Por fim, as duas últimas refletem a necessidade de pesquisas futuras para que a utilização desta abordagem metodológica em ambientes clínicos reais possa se tornar um fato.

6.3 Sugestões para Pesquisas Futuras

A principal sugestão para pesquisas futuras consiste na busca de medidas ou combinações que permitam melhor diferenciação de patologias, já que medidas consideradas nesta investigação são afetadas de modo semelhante por diferentes patologias.

Outras sugestões de continuidade desta mesma pesquisa podem dar mais relevância aos resultados encontrados, tais como: revisão de literatura utilizando a técnica de Revisão Sistemática (GALVÃO; SAWADA; TREVIZAN, 2004; SAMPAIO; MANCINI, 2007), de modo a ser obtido um melhor conhecimento da produção literária das áreas correlatas; utilização de 4 (quatro) modelos em uma classificação (ao invés de apenas dois - classificação binária), com o intuito de analisar o comportamento do PPM quando confrontando vários modelos entre si, o que vai permitir a construção de uma Matriz de Confusão com os resultados e torná-lo de fato adequado à aplicação em ambiente clínico; execução de repetidas classificações, via Validação Cruzada, após mistura aleatória dos arquivos antes de segmentá-los para seleção entre treinamento e teste, sendo o resultado obtido pela média das várias classificações executadas; teste de redução de características utilizando PCA (*Principal Component Analysis*) (JOLLIFFE, 2002) ou redução da dimensionalidade dos dados, com base em Godino-Llorente, Gómez-Vilda e Blanco-Velasco (2006); teste dos indicadores unitariamente, ao invés de categorias de indicadores

(temporais, acústicos e estatísticos), de modo a identificar se há um ou mais, dentro de uma categoria considerada importante pela investigação relatada na Seção 5.2, que não contribuem significativamente a este resultado, podendo ser desconsiderado (a abordagem de Godino-Llorente, Gómez-Vilda e Blanco-Velasco (2006) pode ser levada em consideração); combinação do PPM com outra técnica de classificação, o que pode dar um resultado mais preciso e confiável; teste do vetor de características montado nesta pesquisa com outras técnicas de classificação, principalmente as que não envolvem compressão de dados (ex.: Redes Neurais, Redes Bayesianas etc.).

Entretanto, é válido ressaltar que há a possibilidade de que, utilizando a abordagem apresentada neste documento, resultados melhores e mais confiáveis sejam obtidos ao ser adquirida uma base de dados que: contenha mais sinais de vozes com as patologias principais que não foram diretamente consideradas nesta pesquisa, tais como Pólipo e Nódulo; disponha de mais sinais representativos de Edema, isto é, sinais de vozes que apresentem somente Edema, conforme sugestão dada por Brandt (2012) e com breve menção na Seção 5.6, e que disponha de mais sinais de vozes contendo os diferentes tipos de Edema (unilateral e bilateral, esquerda e direita); disponha também de mais sinais de vozes masculinas, de modo a dar mais significância ao resultado obtido pela distinção de gênero. Ainda não se tem conhecimento da existência e/ou disponibilidade de uma base de dados com essas características.

Por fim, uma sugestão que visa a dar mais robustez ao sistema de discriminação de patologias por computador é a identificação de um limiar de qualidade da voz gravada, de modo que caso o sinal de voz não apresente qualidade suficiente para análise, seja requerida uma nova gravação pelo paciente. Fechine (2000) sugere a análise da Frequência Fundamental por quadros e a estimativa do Coeficiente de Variação dessas frequências ao longo dos quadros. Caso ele seja maior que 40%, deve ser repetida a gravação pelo paciente.

Referências Bibliográficas

ADNENE, C.; LAMIA, B. Analysis of pathological voices by speech processing. In: INTERNATIONAL SYMPOSIUM ON SIGNAL PROCESSING AND ITS APPLICATIONS, 7, 2003, Paris. **Anais...** p.365–367

AGUIAR NETO, B. G. et al. Feature Estimation for Vocal Fold Edema Detection Using Short-Term Cepstral Analysis. In: IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOENGINEERING, 7, 2007. **Anais...** p.1158-1162.

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

ANDRADE SOBRINHO, F. A. **Medida da dispersão da periodicidade de um sinal de voz normal e voz patológica através da seção de Poincaré**. 2011. 98f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade de São Paulo – USP, São Carlos, 2011.

ARIAS-LONDOÑO, J. D. et al. Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients. **IEEE Transactions on Biomedical Engineering**, 58, 2, p.370–379, 2011.

BAO, P.; SIM, A. A hybrid Speech Recognition Model based on HMM and Fuzzy PPM. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS, 5, 1998. **Anais...** p.4148-4153.

BARUFALDI, B. et al. Text Classification by Literary Period Using PPM-C Data Compression. In: SEVENTH BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 7, 2009. **Anais...** p.125–133.

BEHLAU, M. **Voz, O Livro do Especialista**. Revinter, 2001, vol. I

BEHROOZMAND, R.; ALMASGANJ, F., Comparison of Neural Networks and Support Vector Machines Applied to Optimized Features Extracted from Patients' Speech Signal for Classification of Vocal Fold Inflammation. In: INTERNATIONAL SYMPOSIUM ON SIGNAL PROCESSING AND INFORMATION TECHNOLOGY, 5, 2005. **Anais...** p.844–849.

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

BENNETT S; BISHOP S; LUMPKIN S. M. Phonatory characteristics associated with bilateral diffuse polypoid degeneration. **The Laryngoscope**, 97, 4, p.446-50, 1987.

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

BRANDÃO, L. Análise Perceptiva Auditiva para o Diagnóstico de Patologias da Fala (disfonias vocais). 28 de dezembro de 2012. Campina Grande. Entrevista concedida a Sérgio Espínola.

BRANDT, R. R. **Classificação de Vozes Patológicas Utilizando Análise Paramétrica e Não Paramétrica**. 2012. 173f. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande – UFCG, Campina Grande, 2012.

BURBEY, I.; MARTIN, T. L.; Predicting Future Locations Using Prediction-by-Partial-Match. In: ACM INTERNATIONAL WORKSHOP ON MOBILE ENTITY LOCALIZATION AND TRACKING IN GPS-LESS ENVIRONMENTS, 1, 2008, New York. **Anais...** p.1-6.

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

CHAIWANAROM, P.; LURSINSAP, C. Link Completion using Prediction by Partial Matching. In: INTERNATIONAL SYMPOSIUM ON COMMUNICATIONS AND INFORMATION TECHNOLOGIES, 2008. **Anais...** p.675–680.

CLEARY, J. G.; WITTEN, I. H. Data Compression Using Adaptive Coding and Partial String Matching. **IEEE Transactions on Communications**, 32, 4, p.396-402, abr. 1984.

COLTON, R. H.; CASPER, J. K. **Compreendendo os problemas de voz**. Artes Médicas, Porto Alegre, 1996

COSTA, W. C. de A. **Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas**. 2012. 176f. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande – UFCG, Campina Grande, 2012.

COSTA, W. C. de A. et al. Pathological Voice Assessment By Recurrence Quantification Analysis. In: BIOSIGNALS AND BIOROBOTICS CONFERENCE (BRC), 3, 2012. **Anais...** p.1-6.

COSTA, S. L. N. C. **Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas**. 2008. 161f. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande – UFCG, Campina Grande, 2008.

COSTA, S. et al. O Uso da Entropia na Discriminação de Vozes Patológicas. In: CONGRESSO DE PESQUISA E INOVAÇÃO DA REDE NORTE NORDESTE DE EDUCAÇÃO TECNOLÓGICA, 2, 2007. **Anais...**

COUTINHO, B. C. et al. Atribuição de Autoria usando PPM. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 25, 2005. **Anais...** p.2208–2217.

DANIEL, R.; BOONE S.; McFARLANE C. **A voz e a terapia vocal**. Artes Médicas, Porto Alegre, 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. **Anais...** p.271–314.

DELLER Jr., R.; PROAKIS, J. G.; HANSEN, J. H. L. **Discrete-time Processing of Speech Signals**. Macmillan Publishing Co., 1993.

DRINIC, M.; KIROVSKI, D.; POTKONJAK, M. PPM Model Cleaning. In: DATA COMPRESSION CONFERENCE, 2003. **Anais...** p.163–172.

DRINIC, M.; KIROVSKI, D. PPMexe: PPM for Compressing Software. In: DATA COMPRESSION CONFERENCE, 2002. **Anais...** p.192-201.

ECKMANN, J. P.; KAMPHORST S. O.; RUELLE, D. Recurrence plots of dynamical systems. **Europhysics Letters**, 56, 5, pp. 973–977, 1987.

FARIAS, T. M. T. **Sistema Embarcado para um Monitor Holter que utiliza o Modelo PPM na Compressão de Sinais ECG**. 2010. 128f. Dissertação (Mestrado em Informática) – Universidade Federal da Paraíba – UFPB, João Pessoa, 2010.

FARRÚS, M.; HERNANDO, J. Using Jitter and Shimmer in speaker verification. **IET Signal Processing**, Stevenage, 3, 4, pp. 247–257, 2008.

FECHINE, J. M. **Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística**. 2000. 237f. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande – UFCG, Campina Grande, 2000.

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

GALVÃO, C. M.; SAWADA, N. O.; TREVIZAN, M. A. Revisão Sistemática: Recurso que proporciona incorporação das evidências na prática da Enfermagem. **Revista Latino-Americana de Enfermagem**, 12, 3, p.549-556, 2004.

GODINO-LLORENTE, J. **Estrategias para la detección automática de patología laríngea a partir del registro de la voz**. PhD thesis, Universidad Politécnica de Madrid, Madrid, 2002.

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**, 53, 10, p.1943-1953, 2006.

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

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

HAMMARBERG, B. Perception and acoustics of voice disorders: a combined approach. In: SYMPOSIUM ON DATABASES IN VOICE QUALITY RESEARCH AND EDUCATION. Utrecht: Utrecht Institute of Linguistics OTS, 1998. **Anais...** p. 1-6.

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

HIRANO, M. Structure of the vocal folds in normal and disease states: anatomical and physical studies. In: CONFERENCE ON THE ASSESSMENT OF VOCAL PATHOLOGY. Ludlow: ASHA Report 17, 1981. **Anais...** 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.

HONÓRIO, T. C. de S.; BATISTA, L. V.; DUARTE, R. C. M. Texture Classification Using Prediction by Partial Matching Models. In: WORKSHOP DE VISÃO COMPUTACIONAL, 5, 2009. **Anais...**

HU, Y.; LOIZOU, P. C. Evaluation of objective quality measures for speech enhancement. **IEEE Transactions on Audio, Speech, and Language Processing**, 16, 1, p.229-238, 2008.

ISSHIKI, N. Recent advances in phonosurgery. **Folia Phoniatrica et Logopaedica**, 32, 2, p.119-154, 1980.

JAIN, R. **The Art of Computer Systems Performance Analysis:** Techniques for Experimental Design, Measurement, Simulation, and Modeling. New York: Wiley-Interscience, 1991.

JOLLIFFE, I. T. **Principal Component Analysis**. 2.ed. Aberdeen: Springer, 2002.

KAY ELEMETRICS, Kay Elemetrics Corp. Disordered Voice Database1.03 ed. 1994.

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

KOHLER, M. R. **Redes Neurais Artificiais para classificação de patologias vocais**. 2011. 41f. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) - Pontifícia Universidade Católica do Rio de Janeiro - PUC-RJ, Rio de Janeiro, 2011.

LIEBERMAN, P. Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges. **Journal of the Acoustical Society of America**, Melville, 35, 3, p.344-353, 1963.

LIMA, J. S. et al. Classificação de Sinais Vozes Patológicas por meio do Parâmetro de Hurst e LDA. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, 30, 2012. **Anais...**

LONDOÑO, J. D. A. **Stochastic characterization of nonlinear dynamics for the automatic evaluation of voice quality**. 2010. 286f. Tese (Doutorado de Filosofia) - Universidad Politécnica de Madrid and Universidad Nacional de Colombia, Madrid, 2010.

LOPES J. et al. A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa. In: JORNADAS SOBRE TECNOLOGIA E SAÚDE, 1, 2008. **Anais....**

MACIEL, C. D.; PEREIRA, J. C.; STWEART, D. Identifying Healthy and Pathologically Affected Voice Signals. **IEEE Signal Processing Magazine**, v. 27, n. 1, p.120 - 123, jan. 2010.

MAHOUI, M. et al. Identification of Gene Functions Using Prediction by Partial Matching (PPM) Language Models. In: ACM CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 17, 2008. **Anais...** p.779–786.

MALHEIROS, Y. et al. A Source Code Recommender System to Support Newcomers. In: INTERNATIONAL CONFERENCE ON COMPUTER SOFTWARE AND APPLICATIONS, 36, 2012. **Anais...** p.19-24.

MARINUS, J. V. M. L. **Estudo de Técnicas para Classificação de Vozes Afetadas por Patologias**. 2008. 138f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Campina Grande - UFCG, Campina Grande, 2010.

MARINUS, J. V. M. L. et al. On the Use of Cepstral Coefficients and Multilayer Perceptron Networks for Vocal Fold Edema Diagnosis. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND APPLICATIONS IN BIOMEDICINE, 9, 2009, Larnaca. **Anais...** p.1-4.

MARQUES, J. R. T. et al. Compressão de Imagens Mamográficas utilizando Segmentação e o Algoritmo PPM. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA EM SAÚDE, 13, 2006. **Anais...**

MARTINEZ, C. E., RUFINER, H. L., Acoustic Analysis of Speech for Detection of laryngeal Pathologies. In: ANNUAL EMBS CONFERENCE, 22, 2000, Chicago. **Anais...** p.2369-2372.

MEDEIROS, T. F. L. et al. Heart arrhythmia classification using the PPM algorithm. In: BIOSIGNALS AND BIOROBOTICS CONFERENCE, 2011. **Anais...** p.1–5.

MOFFAT, A. Implementing the PPM Data Compression Scheme. **IEEE Transactions on Communications**, 38, 11, p.1917-1921, nov. 1990.

MONDAY, L. A. et al. Epidermoid cysts of the vocal cords. In: OTOLOGY RHINOLOGY & LARYNGOLOGY. [S.l.]: American Broncho-Esophagological Association, 92, 1983. **Anais...** p.124–127.

MONTEIRO, N. A. B. et al. Técnicas de Processamento Digital de Sinais de Voz para Detecção de Paralisia nas Dobras Vocais. In: SEMANA DE CIÊNCIA E TECNOLOGIA DO IFPB, 7, 2011. **Anais...**

MOORE, C. et al. Spectral pattern complexity analysis and the quantification of voice normality in healthy and radiotherapy patient groups. **Medical Engineering & Physics**, 26, p.291–301, 2004.

MOORE, C.; MANICKAM, K.; SLEVIN, N. Collective spectral pattern complexity analysis of voicing in normal males and larynx cancer patients following radiotherapy. **Biomedical Signal Processing and Control**, 1, p.113–119, 2006.

MOORE, D. S.; McCABE, G. P. **Introduction to the Practice of Statistics**. New York: W.H. Freeman & Company, 1998.

MOSTELLER F.; TUKEY J. W. **Data analysis, including statistics**. In Handbook of Social Psychology. Addison-Wesley, Reading, MA, 1968.

OATES, J. Auditory-perceptual evaluation of disordered voice quality. **Folia Phoniatrica et Logopaedica**, 61, 1, p.49--56, 2009.

OROZCO, J. R. et al. Voice Pathology Detection in Continuous Speech using Nonlinear Dynamics. In: INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE, SIGNAL PROCESSING AND THEIR APPLICATIONS, 11, 2012. **Anais...** p.1030-1033.

PAPARELLA, M. M.; SHUMRICK, D. A. **Otorrinolaringologia - Cabeza y Cuello**. Buenos Aires: 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.

PARSA, V., JAMIESON, D. Identification of pathological voices using glottal measures. **Journal of Speech and Hearing Research**, 43, p.469-485, 2000.

PATIL, H. A.; BALJEKAR, P. N. Classification of Normal and Pathological Voices using TEO Phase and Mel Cepstral Features. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATIONS, 2012. **Anais...** p.1-5.

PATIL, H. A.; PATEL, T. B. Novel Chaotic Titration Method for Analysis of Normal and Pathological Voices. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATIONS, 2012. **Anais...** p.1-5.

PAVELEC, D. et al. Author Identification using Compression Models. In: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 10, 2009. **Anais...** p.936–940.

QUEK, F. et al. Speech pauses and gestural holds in parkinson's disease. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. Denver: Speech Research Lab, 2002. **Anais...** p.2485–2488.

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

RAJU, N. et al. Normal versus pathology voice-an analysis. In: INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND APPLICATIONS, 2012. **Anais...** p.1-4.

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

ROADS, C. **Computer Music Tutorial**. Massachusetts: MIT Press, 1995.

RUSSO, I; BEHLAU,M. Percepção da Fala: Análise Acústica. Lovise, 1993.

SÁENS-LECHÓN, N., et al. Automatic assessment of voice quality acoording to the GRBAS scale. In: IEEE EMBS ANNUAL INTERNATIONAL CONFERENCE, 2006, New York. **Anais...** p. 2478–2481.

SALHI, L.; TALBI, M.; CHERIF, A. Voice Disorders Identification Using Hybrid Approach: Wavelet Analysis and Multilayer Neural Networks. In: WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY, 21, 2008. **Anais...** p.330-339.

SALOMON, D. **Data Compression:** The Complete Reference. 3.ed. New York: Springer, 2004.

SAMPAIO, R, F.; MANCINI, M, C. Estudos de Revisão Sistemática: um guia para síntese criteriosa da evidência científica. Revista Brasileira de Fisioterapia, 11, 1, p-83-89, 2007.

SCALASSARA, P. R. et al. Voice signals characterization through entropy measures. In: INTERNATIONAL CONFERENCE ON BIO-INSPIRED SYSTEMS AND SIGNAL PROCESSING, 2, 2008, Madeira, Portugal. **Anais...** p.163–170.

SCALASSARA, P. R. **Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias da Laringe**. 2009. 265f. Tese (Doutorado em Engenharia Elétrica) – Universidade de São Paulo – USP, São Carlos, 2009a.

SCALASSARA, P. R. et al. Analysis of Voice Pathology Evolution Using Entropy Rate. In: IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA, 10, 2009b. **Anais...** p.580–585.

SCALASSARA, P. R. et al. Relative entropy measures applied to healthy and pathological voice characterization. **Applied Mathematics and Computation**, 207, 1, p.95-108, 2009c.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, 52, 3 and 4. 1965. p591-611.

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

SOUZA, A. R. C. **Desenvolvimento e Implementação em FPGA de um Sistema portátil para Aquisição e Compressão sem perdas de**

Eletrocardiogramas. 2008. 180f. Dissertação (Mestrado em Informática) – Universidade Federal da Paraíba – UFPB, João Pessoa, 2008.

STEMPLE, J. C., GLASE, L., KLABEN, B. **Clinical Voice Pathology**: Theory and Management. 4.ed. San Diego: Plural Publishing, 2010.

TAVARES, R. et al. Combining Entropy Measurements and Cepstral Analysis for Pathological Voice Assessment. In: BIOSIGNALS AND BIOROBOTICS CONFERENCE, 2011. **Anais...** p.1–5.

TEAHAN, W. J. **Modelling english text**. 1997. 243f. Tese (Doutorado em Ciência da Computação) - University of Waikato, Hamilton, 1997.

TEIXEIRA, J. P.; FERREIRA, D; CARNEIRO, S. Análise acústica vocal - determinação do Jitter e Shimmer para diagnóstico de patologias da fala. In: CONGRESSO LUSO-MOÇAMBICANO DE ENGENHARIA, 6, 2011. **Anais...** Maputo. p.139-140.

Mathworks Inc. (2009) Versão 7.9.0.529. Disponível em http://www.mathworks.com

TORTORA, G. J.; GRABOWSKI, S. R. (2002). **Princípios de Anatomia e Fisiologia.** Guanabara Koogan, Rio de Janeiro, 9a ed.

VIEIRA, V. J.; COSTA, S. C.; COSTA, W. C. Análise de Quantificação de Recorrência e Análise Discriminante Aplicadas à Classificação de Sinais de Vozes Saudáveis e Sinaisde Vozes Patológicas. In: CONGRESSO NORTE NORDESTE DE PESQUISA E INOVAÇÃO, 7, 2012. **Anais...** Palmas.

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

YUMOTO, E.; GOULD, W.; BAER, T. Harmonics-to-noise ratio as an index of the degree of hoarseness. **Journal of the Acoustical Society of America**, Melville, 71, p.1544–1549, 1982.

ZHANG, Y.; ADJEROH, D. A. Prediction by Partial Approximate Matching for Lossless Image Compression. **IEEE Transactions on Image Processing**, 17, 6, p.924-935, 2008.

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

Apêndice A

O desempenho do PPM foi comparado ao de compressores conhecidos, usando como parâmetros a Razão de Compressão e o tempo de execução. Foi feita comparação com o compressor de Huffman, o algoritmo Deflate e o algoritmo que cria arquivos em formato RAR. Para os dois últimos, foi feito uso do programa WinRAR, disponível em licença gratuita temporária na Internet. São relatados resultados obtidos pela compressão de 5 arquivos: uma letra de música brasileira que não contém refrão (de modo que não há repetição neste fluxo, o que aumenta a entropia do fluxo), dois livros em língua inglesa e dois sinais de áudio da base de dados utilizada nesta pesquisa. Os resultados são sumariados no Quadro 9.

Em todas as classificações, o PPM, o Deflate e o algoritmo que origina o formato RAR executaram quase que instantaneamente (menos de 1s). O algoritmo de Huffman, porém, apresentou tempos de execução entre 100 s e mais de 90 min, dependendo do tamanho do arquivo usado na compressão.

É possível perceber, pelos tamanhos dos arquivos comprimidos obtidos pelos diferentes compressores, que o PPM apresenta resultados bastante superiores aos demais para os arquivos de texto, porém, o mesmo não acontece para os sinais de voz, para os quais o formato RAR gera sempre os menores arquivos.

Além disso, para os arquivos de texto, os melhores resultados obtidos a partir da utilização do PPM ocorreram com tamanhos de contexto 3, 4 e 5, respectivamente. Porém, para os sinais de voz, os melhores resultados são obtidos apenas com o contexto 0. Isso reflete pouca repetição de sequências nos sinais de voz e mostra porque a alimentação do modelo diretamente com os bytes dos sinais de voz não retorna bons resultados.

Quadro 9 - Benchmark entre diferentes compressores e diferentes tipos de arquivo

	Tamanho original	*Huffman*	*PPM*	*ZIP*	*RAR*
Faroeste Caboclo [29]	6707 b	3915 b	2661 b	3206 b	3186 b
The Little Book of Humility & Patience (Archbishop Ullathorne)	116983 b	64806 b	31412 b	40202 b	38751 b
Pride and Prejudice (Jane Austen)	708428 b	[30]	167117 b	244193 b	211882 b
Arquivo de voz normal	300046 b	[18]	284344 b	284510 b	214355 b
Arquivo de voz patológica	50402 b	47653 b	47495 b	47691 b	44822 b

[29] Música da banda brasileira de rock Legião Urbana que não apresenta refrão. Isso significa menor repetição, que implica em mais informação distinta e, consequentemente, maior entropia.

[30] A execução se mostrou demasiadamente longa, a ponto de não haver a espera até o final para serem obtidos os dados.

Apêndice B

No Quadro 10 estão listados os arquivos utilizados na pesquisa ora documentada.

Quadro 10 - Lista dos arquivos da base utilizados nesta pesquisa

Nome	Diagnóstico	Gênero do paciente
axh1nal	Normal	Feminino
bjv1nal	Normal	Feminino
cad1nal	Normal	Feminino
ceb1nal	Normal	Feminino
daj1nal	Normal	Feminino
dfp1nal	Normal	Feminino
dma1nal	Normal	Feminino
edc1nal	Normal	Feminino
hbl1nal	Normal	Feminino
jaf1nal	Normal	Feminino
jan1nal	Normal	Feminino
jap1nal	Normal	Feminino
jeg1nal	Normal	Feminino
jkr1nal	Normal	Feminino
jth1nal	Normal	Feminino
jxc1nal	Normal	Feminino
lad1nal	Normal	Feminino
ldp1nal	Normal	Feminino
lla1nal	Normal	Feminino
lmv1nal	Normal	Feminino
lmw1nal	Normal	Feminino
mam1nal	Normal	Feminino
mcb1nal	Normal	Feminino
mxb1nal	Normal	Feminino
mxz1nal	Normal	Feminino
njs1nal	Normal	Feminino

pbd1nal	Normal	Feminino
sck1nal	Normal	Feminino
sct1nal	Normal	Feminino
seb1nal	Normal	Feminino
slc1nal	Normal	Feminino
vmc1nal	Normal	Feminino
bjb1nal	Normal	Masculino
djg1nal	Normal	Masculino
dws1nal	Normal	Masculino
ejc1nal	Normal	Masculino
fmb1nal	Normal	Masculino
gpc1nal	Normal	Masculino
gzz1nal	Normal	Masculino
jmc1nal	Normal	Masculino
kan1nal	Normal	Masculino
mas1nal	Normal	Masculino
mfm1nal	Normal	Masculino
mju1nal	Normal	Masculino
ovk1nal	Normal	Masculino
pca1nal	Normal	Masculino
rhg1nal	Normal	Masculino
rhm1nal	Normal	Masculino
rjs1nal	Normal	Masculino
sis1nal	Normal	Masculino
sxv1nal	Normal	Masculino
txn1nal	Normal	Masculino
wdk1nal	Normal	Masculino
ana15an	Edema	Feminino
cac10an	Edema	Feminino
cak25an	Edema	Feminino
cer16an	Edema	Feminino
dbf18an	Edema	Feminino
djf23an	Edema	Feminino
exe06an	Edema	Feminino
hlm24an	Edema	Feminino
jaj31an	Edema	Feminino
jmc18an	Edema	Feminino

jxc21an	Edema	Feminino
jxf11an	Edema	Feminino
kab03an	Edema	Feminino
klc09an	Edema	Feminino
lad12an	Edema	Feminino
lgm01an	Edema	Feminino
lxd22an	Edema	Feminino
mca07an	Edema	Feminino
mcw21an	Edema	Feminino
nfg08an	Edema	Feminino
nlc08an	Edema	Feminino
pmf03an	Edema	Feminino
rcc11an	Edema	Feminino
sxg23an	Edema	Feminino
vaw07an	Edema	Feminino
ctb30an	Edema	Masculino
dmg07an	Edema	Masculino
dxc22an	Edema	Masculino
jjd29an	Edema	Masculino
jxb16an	Edema	Masculino
pat10an	Edema	Masculino
rjl28an	Edema	Masculino
rtl17an	Edema	Masculino
wst20an	Edema	Masculino
abb09an	Paralisia	Feminino
car10an	Paralisia	Feminino
dac26an	Paralisia	Feminino
edg19an	Paralisia	Feminino
esl28an	Paralisia	Feminino
hjh07an	Paralisia	Feminino
igd16an	Paralisia	Feminino
jpp27an	Paralisia	Feminino
jtg18an	Paralisia	Feminino
klc06an	Paralisia	Feminino
kmc19an	Paralisia	Feminino
kmc27an	Paralisia	Feminino
kms29an	Paralisia	Feminino

lba24an	Paralisia	Feminino
ljs31an	Paralisia	Feminino
mec06an	Paralisia	Feminino
mec28an	Paralisia	Feminino
mnh04an	Paralisia	Feminino
mnh14an	Paralisia	Feminino
mps09an	Paralisia	Feminino
rab08an	Paralisia	Feminino
rab22an	Paralisia	Feminino
rec19an	Paralisia	Feminino
tac22an	Paralisia	Feminino
ajm05an	Paralisia	Masculino
bsa08an	Paralisia	Masculino
cty03an	Paralisia	Masculino
cty09an	Paralisia	Masculino
djp04an	Paralisia	Masculino
eec04an	Paralisia	Masculino
ejh24an	Paralisia	Masculino
fxc12an	Paralisia	Masculino
gsb11an	Paralisia	Masculino
jfg26an	Paralisia	Masculino
jfn11an	Paralisia	Masculino
jfn21an	Paralisia	Masculino
jxs01an	Paralisia	Masculino
jxs09an	Paralisia	Masculino
jxs23an	Paralisia	Masculino
kjb19an	Paralisia	Masculino
ran30an	Paralisia	Masculino
rpj15an	Paralisia	Masculino
swb14an	Paralisia	Masculino
tdh12an	Paralisia	Masculino
tps16an	Paralisia	Masculino
wdk47an	Paralisia	Masculino
amd07an	Outras	Feminino
anb28an	Outras	Feminino
axl04an	Outras	Feminino
brt18an	Outras	Feminino

bsd30an	Outras	Feminino
bsg13an	Outras	Feminino
cls31an	Outras	Feminino
dmc03an	Outras	Feminino
dmp04an	Outras	Feminino
drc15an	Outras	Feminino
dsc25an	Outras	Feminino
eab27an	Outras	Feminino
eeb24an	Outras	Feminino
ell04an	Outras	Feminino
hmg03an	Outras	Feminino
igd08an	Outras	Feminino
jcc10an	Outras	Feminino
jeg29an	Outras	Feminino
jmh22an	Outras	Feminino
jwe23an	Outras	Feminino
kas09an	Outras	Feminino
kcg23an	Outras	Feminino
kmw05an	Outras	Feminino
lac02an	Outras	Feminino
lai04an	Outras	Feminino
lap05an	Outras	Feminino
lba15an	Outras	Feminino
mcb20an	Outras	Feminino
mlf13an	Outras	Feminino
mlg10an	Outras	Feminino
mms29an	Outras	Feminino
mpb23an	Outras	Feminino
mpc21an	Outras	Feminino
mrc20an	Outras	Feminino
njs06an	Outras	Feminino
nmc22an	Outras	Feminino
pmd25an	Outras	Feminino
rjz16an	Outras	Feminino
rmb07an	Outras	Feminino
tlp13an	Outras	Feminino
tls09an	Outras	Feminino

amb22an	Outras	Masculino
amc14an	Outras	Masculino
cjb27an	Outras	Masculino
dmg24an	Outras	Masculino
eas11an	Outras	Masculino
ejb01an	Outras	Masculino
jbs17an	Outras	Masculino
jtm05an	Outras	Masculino
mpf25an	Outras	Masculino
oab28an	Outras	Masculino
rjc24an	Outras	Masculino
rjr15an	Outras	Masculino
rpc14an	Outras	Masculino
wjb12an	Outras	Masculino

No Quadro 11 estão listados os totais (quantidades de arquivos) contidos em cada classe.

Quadro 11 – Quantidades das classes de arquivos utilizadas nas classificações

Classificação	Quantidade de arquivos	Arquivos com vozes femininas	Arquivos com vozes masculinas
Normal	53	32	21
Edema	34	25	9
Paralisia	46	24	22
Outras	56	42	14
Patológico	136	91	45

A classe de arquivos denominada *Patológico* (que não consta no Quadro 10) consiste nos sinais correspondentes a Edema, Paralisia e Outras patologias reunidos, sendo utilizada na classificação Normal x Patológico, realizada com o intuito de comparar diretamente o desempenho do PPM nesta classificação com outros classificadores (e verificar seu potencial), além de servir de base para o estudo de similaridades entre as diferentes patologias.

É possível perceber a pouca representatividade de sinais de vozes masculinas na base de dados utilizada, de modo que um estudo aprofundado que inclua a distinção de gêneros, como o de Brandt (2012), deve levar isso em consideração.

Anexo A

Projeto Experimental

O conceito de Projeto Experimental é derivado da Regressão Linear. O intuito de construir modelos de regressão, os mais comuns utilizados por analistas estatísticos, é estimar/prever uma *variável de resposta* a partir de uma ou mais *variáveis preditoras* ou *fatores*. Cada variável preditora pode ter um ou mais *níveis*. A ideia é construir um modelo que permita a estimação dos valores obtidos com o mínimo possível de erros.

A execução de um projeto experimental se utiliza dos modelos de regressão, isolando os efeitos de cada fator dos efeitos de outros fatores de modo a inferir conhecimento significativo sobre seus diferentes níveis. O intuito principal é comparar diferentes alternativas, obtendo o máximo de informação com o mínimo de experimentos. Por exemplo, suponha que se quer comparar o tempo de resposta da execução de uma mesma carga de trabalho em três sistemas diferentes. Primeiramente, os tempos de cada execução em cada sistema são coletados e, utilizando projeto experimental, é possível identificar o sistema no qual esta carga de trabalho executa mais eficientemente. Nesse caso, o sistema no qual a carga executa é o fator, os diferentes sistemas estudados são os níveis deste fator e os tempos de respostas são as variáveis de resposta.

Existem vários tipos de projetos experimentais. Como nesta pesquisa foi utilizado um projeto experimental de fator único, no qual o fator era a entrada que alimentava o modelo, os diferentes tipos de medidas extraídas e combinações eram os níveis deste fator e os percentuais de acerto eram as variáveis de resposta, esse anexo se restringirá a descrever um projeto experimental de fator único. Informações sobre os outros tipos de projetos experimentais podem ser encontradas em Jain (1991).

Não há limites para a quantidade de níveis em projetos experimentais de fator único e para cada nível devem ser executadas diversas replicações. Mais

especificamente, são executadas r replicações utilizando a níveis do fator estudado. O modelo utilizado é $y_{ij} = \mu + \alpha_j + e_{ij}$, em que y_{ij} é a i-ésima resposta (replicação) obtida com o fator no j-ésimo nível (ou alternativa), μ é a média das respostas (resposta média), α_j é o efeito do nível j e e_{ij} é o erro desta estimativa. Os efeitos devem ser computados de modo que as somas resultem em 0, ou seja, $\sum \alpha_j = 0$.

As diversas estimativas podem ser organizadas em uma matriz r x a, de modo que a coluna j representa as diversas respostas obtidas utilizando o nível j do fator. Mais intuitivamente,

$$\begin{bmatrix} y_{11} = \mu + \alpha_1 + e_{11} & \cdots & y_{1a} = \mu + \alpha_a + e_{1a} \\ \vdots & \ddots & \vdots \\ y_{r1} = \mu + \alpha_1 + e_{r1} & \cdots & y_{ra} = \mu + \alpha_a + e_{ra} \end{bmatrix}.$$

Pelo fato de os efeitos serem computados de modo que a soma resulte em 0 e se desejar também que a soma dos erros seja 0, $\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij} = ar\mu + 0 + 0$. Sendo assim, $\mu = \frac{1}{ar}\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij}$. O lado direito desta equação é denotado também por $\bar{y}_{..}$. Os dois pontos denotam que essa é a média tanto das linhas quanto das colunas da matriz. O símbolo $\bar{y}_{.j}$, por exemplo, denota a média apenas da j-ésima coluna, que pode ser encontrada por $\bar{y}_{.j} = \frac{1}{r}\sum_{i=1}^{r} y_{ij}$. Substituindo y_{ij} por $\mu + \alpha_j + e_{ij}$, obtém-se

$$\begin{aligned} \bar{y}_{.j} &= \frac{1}{r}\sum_{i=1}^{r}(\mu + \alpha_j + e_{ij}) \\ &= \frac{1}{r}\left(r\mu + r\alpha_j + \sum_{i=1}^{r} e_{ij}\right) \\ &= \mu + \alpha_j . \end{aligned} \qquad (15)$$

Sendo assim, $\alpha_j = \bar{y}_{.j} - \bar{y}_{..}$, que é a quantificação do efeito do nível j do fator analisado sobre os resultados.

O seguinte exemplo, dado por Jain (1991), ajuda a compreender esta ferramenta estatística, sua utilidade e o significado dos resultados que se obtém. Suponha que se queira analisar o tamanho em bytes requeridos para codificar uma mesma carga de trabalho em três arquiteturas diferentes, R, V e Z. O fator, nesse caso, é a arquitetura computacional e os níveis são as arquiteturas R, V e Z. A variável de resposta é o tamanho do código gerado. Foram executadas cinco replicações de cada experimento, cada uma gerando resultados distintos, os quais são mostrados no Quadro 12.

Quadro 12 - Relação de arquiteturas hipotéticas e números de bytes para codificar uma carga de trabalho

R	V	Z
144	101	130
120	144	180
176	211	141
288	288	374
144	72	302

A média da primeira coluna é 174,4, a da segunda 163,2 e a da terceira, 225,4. A média total é 187,7. Sendo assim, com base na Equação 15, os efeitos das arquiteturas R, V e Z são, respectivamente, -13,3, -24,5 e 37,7. Isso significa que a arquitetura R codifica essa carga de trabalho utilizando, em média, -13,3 bytes que o processador médio. É fácil perceber, portanto, que V é a arquitetura mais eficiente e Z é a menos eficiente.

Anexo B

Intervalos de Confiança

Seja A um conjunto de números de tamanho n e média μ e desvio padrão σ. Suponha que se deseja extrair informações deste conjunto. Se n for muito grande (por exemplo, a população de todo o planeta ou a quantidade de moléculas de uma substância), pode ser inviável ou até mesmo impossível obter essas informações utilizando todo o conjunto, chamado deste ponto em diante de *população*.

Por essa razão, é comum a utilização de uma *amostra* da população, que consiste basicamente em um subconjunto dessa população. Seja $\bar{x}$ a média amostral da amostra extraída e s o desvio padrão amostral. A estatística constantemente procura inferir conhecimento de uma população utilizando amostras dessa população. Não é difícil perceber que o conhecimento inferido pode não ser realístico, devido à possibilidade de a amostra extraída não ser representativa da população. Por exemplo, é possível que $\bar{x}$ seja diferente de μ, caso a amostra extraída se desvie muito da população A. É possível até mesmo que duas amostras extraídas da mesma população sejam distintas entre si.

Por isso, é comum a utilização de intervalos de confiança, que permitem a estimação de um intervalo em que há alta probabilidade de as informações da população estarem contidas. A probabilidade é dada por 1 - α, em que α é o nível de significância do intervalo e 1 - α é o coeficiente de confiança. Os coeficientes mais comumente utilizados são 90 e 95%. Nesta pesquisa foi utilizado sempre o coeficiente 95%.

O intervalo de confiança de uma amostra é dado por

$$\left(\bar{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}} \right) \qquad (16)$$

em que $\bar{x}$ é a média amostral, $z_{1-\alpha/2}$ é o $1-\alpha/2$-quantil de uma variável normal (livros de estatística normalmente disponibilizam uma tabela com os valores desse quantil), s é o desvio padrão da amostra e n é o tamanho da amostra. Por exemplo, dada uma amostra com 32 números, média amostral 3,9 e desvio padrão amostral 0,95, o intervalo de confiança dessa amostra a 90% de confiança é (3,62; 4,17), o que significa que há 90% de chances de a média da população do exemplo estar dentro deste intervalo. Em outras palavras, se extrairmos 100 amostras da população, a média de cerca de 90 delas estará dentro deste intervalo.

A Equação 16 somente deve ser utilizada caso a amostra extraída tenha mais que 30 números. Para amostras menores, os intervalos de confiança somente devem ser utilizados caso a distribuição amostral seja normal e, neste caso, é usado o quantil de uma variável t com n-1 graus de liberdade. Nesse caso, deve ser usada a Equação 17.

$$\left(\bar{x} - t_{[1-\alpha/2;\,n-1]}\frac{s}{\sqrt{n}}, \bar{x} + t_{[1-\alpha/2;\,n-1]}\frac{s}{\sqrt{n}}\right) \qquad (17)$$

Intervalos de confiança podem ser usados para comparar duas alternativas. Para comparar mais de duas alternativas entre si, recomenda-se expressamente a utilização de Projeto Experimental. A utilização de intervalos de confiança na comparação de duas alternativas não-pareadas é referida por Jain (1991) como *t-test* e consiste no seguinte: (i) extraem-se as médias amostrais, $\bar{x}_1$ e $\bar{x}_2$, (ii) os desvios padrão amostrais s_1 e s_2, (iii) computa-se a diferença das médias amostrais e (iv) o desvio padrão da diferença média $s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. Em seguida, deve ser obtido o número real de graus de liberdade destes dados, que pode ser obtido pela Equação 18.

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1+1} \bullet \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2+1} \bullet \left(\frac{s_2^2}{n_2}\right)^2} - 2 \qquad (18)$$

O último passo é computar o intervalo de confiança, procedimento feito pela Equação 19.

$$\left(\overline{x_1} - \overline{x_2}\right) \mp t_{[1-\alpha/2;v]} s \qquad (19)$$

Caso esse intervalo contenha o 0, então não há diferença significativa (com confiança 1 - α) entre as alternativas comparadas. Caso não contenha, a melhor alternativa é identificada pelo sinal dos limites do intervalo: para limites positivos, a melhor alternativa é a primeira; caso contrário, a segunda.

Anexo C

CLASSIFYING VOCAL DIAGNOSTICS USING PPM

ABSTRACT

Speech organs are very susceptible to several types of pathologies, which may harm voice production. Several techniques have been traditionally used to detect these pathologies. However, they present drawbacks concerning the accuracy and the comfort of patients during application. Moreover, results obtained by computing techniques have not yet matured to a reliable tool for application in clinics. In this research, a classification approach based on a method not previously employed in classification of vocal tract diseases is proposed. It is based on Prediction by Partial Matching (PPM), which uses acoustical and temporal features to feed models. It obtained very promising results in the presence or absence of pathologies (at least 92%). With regard to pathology discrimination, preliminary results confirmed that PPM is a high potential technique for voice pathology classification, although its clinical application for diagnosis of voice pathologies still needs deeper investigations.

KEYWORDS

speech pathologies; prediction by partial matching; acoustical, temporal and statistical features.

1. INTRODUCTION

Voice is the most important and most natural means of communication of mankind. However, for a effective communication, it is necessary correct understanding of voice enunciated by interlocutor. If this doesn't occur, there will be more propensity to misunderstandings, what discourages communication and causes embarrassment in the speaker. This is named dysphonia and it is often caused by speech pathologies, to which the production system of speech is susceptible. It is common that a person be affected by up to 8 pathologies (KAY ELEMETRICS, 1994). They can be caused by either psychoemotional alterations, or neurodegenerative diseases, or misuse of voice or unhealthy social habits, such as smoking and drinking alcohol (COSTA et al., 2012; COSTA, 2008; MARINUS, 2010), which can explain their most frequent occurrence in smokers and professional categories that use voice as main work tool, e.g., teachers, singers, and journalists (MARINUS, 2010). Costa et al. (2012) stated that there is an estimate of that between 3 to 10% of the general population have production system of speech affected by some pathology.

Two types of mechanisms are traditionally used in detecting speech pathologies. The former consists of hearing patient's vocal utterance by a professional (a speech therapist or an otorhinolaryngologist). This method was the most used until few years ago (HU; LOIZOU, 2008). However, it isn't hard to be aware about your subjectivity and high propensity for inducing errors, mainly in initial stages of pathology, due to high dependency of accuracy, fatigue level and sensitivity of the auditory system of professional[31] (LOPES et al., 2008; OATES, 2009). This kind of examination should be performed only in the absence of other alternatives.

The second consists of evaluating patient's phonatory system by visual clinical procedures, such as videolaringoscopy and videoestroboscopy (COSTA, 2008). In spite of its accuracy, these examinations are very invasive and uncomfortable to patients, causing sometimes reflect during application, due to laryngeal sensitivity, what may result in false diagnosis (MARINUS, 2010). Moreover, the equipment required to execution is expensive and sophisticated, harming financially both sides due to recharge of costs to patients and restricting access to big part of the population.

Numberless studies about detection of speech pathologies by computer has been developed, in order to reduce need for visual examinations. Among used techniques, are included Hidden Markov Models (HMM) (COSTA, 2008), Vector Quantization, Neural Networks (MARINUS, 2010), Recurrence Plots (COSTA et al., 2012; VIEIRA et al., 2012), Gaussian Mixture Models (GMM) (MARINUS, 2010; ARIAS-LONDOÑO et al., 2011), TEO Phase (PATIL; BALJEKAR, 2012), Support Vector Machines (RAJU et al., 2012; ARIAS-LONDOÑO et al., 2011), Nonlinear Dynamics and complexity measures (OROZCO et al., 2012), Hurst

[31] Different diagnosis can be informed by different professionals or even by the same professional in different occasions.

parameter (LIMA et al., 2012), Entropy (TAVARES et al. 2011) and so on. However, it wasn't found reports in the literature review about using data compression methods. Although they have been initially designed to compress data, it was noticed that rich statistic model built by some of these algorithms can be also used in classification.

According to Medeiros et al. (2011), one of the most effective methods in data compression is the Prediction by Partial Matching (PPM). Its operation will be described in Section 2.2. However, it is important to be aware of that good results has been obtained from its use in data compression and classification of binary files, text, electrocardiogram data, image (COUTINHO et al., 2005; BARUFALDI et al., 2009; HONÓRIO et al., 2009; MEDEIROS et al., 2011) and other signal types.

In this paper, PPM was used together with temporal, acoustical and statistical (entropy) parameters, singly or combined, which composed feature vectors. PPM models maintained during process were fed with these feature vectors, what enabled maintaining smaller models and feeding with data already used successfully in other works that do processing of voice signals, such as Fechine (2000), Costa (2008), Tavares et al. (2011), Scalassara et al. (2009) and others.

2. THEORETICAL FUNDAMENTALS

The theoretical fundamentals of this paper are divided in two parts. Firstly, it will be presented measures extracted of voice signals, which were used as input to the models maintained during classifications. After that, the functioning of PPM will be briefly presented.

2.1 Parameters

It is important to mean that voice signals present statistical invariability in windows of 32 ms duration because, in spite of the voice be time invariant, vocal tract presents dynamical nature[32], what affects parameters that represent the voice and, consequently, its production. Thus, in order that extracted values are accurate, all considered measures were extracted into windows within that range.

Energy - sum of squares of amplitude values of the signal. To obtain this value in decibels, it's necessary multiply logarithm of the value found by 10. Formulas related to this parameter can be found in Fechine (2000).

Zero-Crossing Rate - number of times that waveform of the signal crosses horizontal axis (time), i.e, how many times a positive value of amplitude is succeeded by a negative value and vice versa.

Total Number of Peaks (TNP) - amount of peaks, positive and negative, existent in the waveform of signal.

Difference in Number of Peaks (DNP) - differs from the previous measure only in signal used in operation, since in obtaining TNP, the amounts of peaks are summed and in obtaining DNP, positive peaks are subtracted of negative peaks.

Fundamental Frequency - Sound is transmitted by mechanical vibrations that propagate itself through interaction with a physical medium. If there is a pattern in these vibrations, sound has a periodic waveform. If there isn't a pattern, it is classified as noise. Repetition of a periodic waveform is named cycle. The number of cycles in a second that occurs on transmission of a sound represents the Fundamental Frequency of this sound (ROADS, 1995).

In the context of processing voice signals, F_0 represents vibration frequency of vocal folds and influences tone of voice directly. Men present F_0 values between 100 and 137 Hz, women present values between 177 and 244 Hz and children present even higher values, between 206 and 281 Hz. However, these values only occur in healthy voices, because speech pathologies affect vibratory pattern of vocal folds, always reducing their vibration speed and, consequently, the F_0 value. Bennett et al. (1987) did a survey with a group composed by man and women that presented Edema. The authors noticed that women presented an average value of F_0 of 108 Hz, while men presented an average value of F_0 of 91 Hz.

Jitter - measure of variability of T_0 values of the signal. It is useful for evaluating stability of phonatory system, more specifically vibration of vocal folds, which is reduced on presence of pathologies (TEIXEIRA et al., 2011). Formulas related to this parameter can be found in Teixeira et al. (2011) and Farrús and Hernando (2008).

Shimmer - similar to Jitter, however, it is employed in analysis of existent perturbation in amplitude values of the signal. This measure is useful for verifying stability of intensity vocal, which is affected by subglotic pressure and, on the other hand, it is influenced by vibration amplitude and tension of vocal folds (FARRÚS; HERNANDO, 2008). Formulas related to this parameter can be found in Teixeira et al. (2011) and Farrús and Hernando (2008).

---

[32] Its configuration varies with time during production of speech

Harmonic-to-Noise Ratio (HNR) - it enables to evaluate amount of noise present in a signal, relative to its periodic component. It is appropriate in voice signal digital processing because periodic component is due to vibration of vocal folds and aperiodic component, from glottal noise (LOPES et al., 2008). Pathologies interfere directly on vibratory pattern of vocal folds, giving to voice a noisy aspect, so that a low HNR value represents a strong indication of presence of a pathology. Parsa and Jamieson (2000) stated that 83% of pathological voices present low HNR value and that the more noisy is a voice signal, the more advanced the stage of present pathology .

Linear Predictive Analysis - it is considered one of the most powerful techniques in speech analysis. It is intimately related with the idea that it is possible to estimate a sample of signal voice by linear combination of previous samples. It means that whole signal can be rebuilt from estimates. That estimation need to be done aiming to minimize the mean-squared prediction error between real samples and predicted samples, what enable to find an only set of p predictor coefficients. The bigger p, the more accurate the estimates and, consequently, the more fair the rebuilt signal. Its use is common to transmit speech, in order that to enable a transmission at low bitrate.

Linear Predictive Analysis is originated from the notion of speech production as a linear system time variant, excited by a pulse train near-periodic and random noise. Linear prediction provides robust methods and accurate to estimate parameters that characterize this system, among them Linear Predictive Coding (LPC), which was used in this research. This parameter was used by Costa (2008), Marinus (2010) and others authors for classifying vocal diagnostics.

Entropy - a measure of uncertainty of a random variable. It is fundamental to Theory Information, in which is employed as maximum compression threshold that can be obtained in compressing a message from a source information. This parameter was used by Tavares et al. (2011) and others authors for classifying vocal diagnostics.

2.2 Compression method PPM

PPM is adopted for compressing data streams. It is the most effective compressor, state of the art in area (SALOMON, 2004; BARUFALDI et al., 2009; HONÓRIO et al., 2009; MEDEIROS et al., 2011). Its use in commercial scale is still very limited, restricted to academic researches, because during its execution, it is created and stored a very accurate model of the data source being compressed, causing high memory consumption and relatively low execution speed, mainly when compared to the most used compressors.

Functioning of statistical methods of data compression, such as PPM, can be split in two stages: modeling and encoding (SALOMON, 2004). Modeling consists of storing probabilities of symbols from a data stream. Bearing in mind that PPM model is context-based, it can also store probabilities of sequences of symbols, in order to estimate, as accurately as possible, the probabilities distribution of the stationary abstract data source that generated the data stream. The choice of maximum size context stored is free, but it is noteworthy that memory consumption grows exponentially with this value. Moreover, models with big contexts store many information that occur infrequently and, according to Salomon (2004), learning curve of PPM model ceases its growth from a given context size, reducing compression rate. Thus, maximum size context used in any execution of PPM deserves attention.

Second stage of a statistical process of compression is encoding, that normally consists of assigning codes to each symbol, aiming to generate a smaller data stream that the original. Thereto, most frequent symbols and sequences (identified on modeling stage) need to receive the smallest codes, with lowest amount of bits, while less frequent symbols receive the largest codes. However, in obtaining a compressed code of ideal size, it is necessary assign fractionary amounts of bits to each symbol of original stream. Arithmetic Coding reaches that goal by making original message in a number with unbounded amount of decimal places, within a range defined during its operation. At beginning of process, range is [0, 1). To the extent that more probabilities are received, this range reduces, but amount of decimal places of bounds enlarges. In the end of process, it has a small range with boundaries containing a lot of decimal places. Compressed stream may be any number within that range. Thus, to each symbol of the original stream is assigned a fractionary amount of bits.

3. MATERIALS AND METHODS

The database used in this research was developed by Kay Elemetrics, recorded at Voice Speech Lab of Massachussets Eye and Ear Infirmary (MEEI) (KAY ELEMETRICS, 1994). It is composed by 1381 files in NSP format, containing 16 bits/sample, of which 666 contain pronunciation of the sustained vowel *a* (the set used in this research). Among these files, 53 contain healthy voice signals (each of them during 3s) and remainder contain pathological voice signals (duration of 1s). It is noteworthy that there isn't unanimity among sampling rates of files. All healthy voices signals were sampled at 50 kHz (50,000 samples per second).

Pathological voices signals, on other hand, were sampled at 25 kHz, although there are some signals that also was sampled at 50 kHz. Among the 666 that contain pronunciation of sustained vowel *a*, were used 189, which 53 contain Normal voices, 34 contain voices presenting Edema, 46 contain voices with Paralysis and 56 contain voices with Other pathologies.

In both training and tests, any processing of voice signals include stages such as segmentation (with overlap of 50%) and windowing, that consists of, respectively, to extract only a part (window) of the signal, softening effect of the extremities of each window and maintaining the effect of center. Also in both training and tests, the measures presented in Section 2 are extracted of each window, in order to compose a feature vector representative of the window. A whole voice signal is represented by a sequence of these feature vectors. PPM is fed with these vectors instead of the original voice signal, as usually happens.

The difference between training and test is that in training, the model is modified during process, that is, it grows by inclusion of new entries (new contexts or symbols that succeed existing contexts) and it is updated by increment of internal counters. Moreover, in training, there isn't encoding (converting input stream in a compressed stream). In tests, on other hand, there is encoding and the model becomes static, so that internal data aren't updated. This is done to avoid that model learn about voice signals used in tests, what characterizes bias in the process.

First stage of this research consisted of identifying best input type for each case of classification, among temporal parameters, acoustical parameters, entropy, LPC coefficients and combinations among them. It was done using one-factor Experimental Design, presented by Jain (1991), in that the factor is the input type that feeds models and the levels of this factor are the several input types and combinations among them. This tool provides as a result the quantification of impact of each level of the factor analyzed (the effect of each factor) in the results, that is, how much it raises them or lowers them, what enable identifying the best level of the factor: the input type that returns best results. In this stage, 60% of files were used in training and remaining in tests. Selection of files for training was random, in order to avoid bias in results and obtaining a good error estimate.

Then, after identifying the best input type to each case, it obtained percentages characterizing the effectiveness of PPM in diagnostic of speech pathologies by 4-fold Cross Validation. At this stage, differently of previous stages, it isn't possible to insert randomness. It was employed 4 parties in Cross Validation because it was realized that this is the minimal amount of parties that maintains a enough amount (statistically significant) of voice signals to be used in tests. That's because are used 34 voice signals presenting Edema (the smallest amount among considered classes). By dividing in 4 parties, each part contains 8 files. If it were used a bigger amount of parties, each subset used in tests would hold less than 8 files, what doesn't enable a good error estimate.

4. RESULTS

At this section, the obtained results in all stages of research will be shown and discussed.

4.1 Identification of the best input type

The objective of this stage of the investigation was to identify the best input type for each case, among temporal parameters, acoustical parameters, entropy, LPC coefficients and combinations among them. As shown at Table 1, it discovered that there isn't an input type that returns highest percentages in all cases. The best input type varies with accomplished classification.

Table 1. Obtained results in each case and input types used in obtaining them

Classification	**Median in % of first class and input type**	**Median in % of second class and input type**
Normal x Pathological	100,0 Only Temporal; Only Entropy; Temporal and Entropy.	96,3 Temporal and Entropy.
Normal x Edema	100,0 Only Temporal; Temporal and Entropy.	100,0 Temporal and Entropy.

Normal x Paralysis	100,0 Only Temporal; Temporal and Entropy.	94,7 Only Temporal.
Normal x Others	100,0 Only Temporal; Temporal and Entropy.	95,6 Only Temporal.
Edema x Paralysis	64,2 Only LPC.	73,7 Temporal and Acoustical.
Edema x Others	57,1 Only LPC.	65,2 Only Acoustical.
Paralysis x Others	68,4 Only Temporal.	65,2 Temporal, Acoustical and Entropy.

4.2 Percentages by Cross Validation

After finding the best and most feasible configurations for each case of classification, it used 4-fold Cross Validation aiming to obtain percentages of correct answers for each of them, in order to characterize PPM as classifier of speech pathologies regarding effectiveness. Results are shown at Table 2.

Table 2. Percentages obtained by Cross Validation using best configurations

Classification	**Obtained results of first class (median %)**	**Obtained results of second class (median %)**
Normal x Pathological	100,0	94,1
Normal x Edema	100,0	95,0
Normal x Paralysis	100,0	92,3
Normal x Others	100,0	96,4
Edema x Paralysis	50,0	84,6
Edema x Others	40,0	75,0
Paralysis x Others	57,6	64,2

4.3 Considerations about efficiency

Efficiency of PPM on classifying speech pathologies can be analyzed regarding performance and computational resources consumption, particularly memory.

Regarding speed, were analyzed execution times from 10 classifications using the most well represented classes of database: Normal and Pathological (which includes all pathologies considered in this paper), that contains 53 and 136 voice signals, respectively. These classes were chosen because they demand more files than others, both in training and tests, so that times of other classifications certainly are smaller than times exhibited at Table 3. The intervals presented at Table 3 correspond to 95% confidence interval of times of training and individual tests.

Table 3. Execution times of Normal x Pathological classifications

Stage	Confidence Interval (ms)
Training	[405,3; 436,6]
Individual tests	[10,8; 12,2]

Regarding memory consumption, 10 executions of same classification were analyzed and maximal memory consumption in training was recorded, because in that stage occurs the maximal memory consumption, since tests don't update model. The 95% confidence interval of these records is [15.02; 15.77] MB. VisualVM, a tool provided by the distributor of Java platform, was used for monitoring.

For obtaining these values, it was used a machine with 6 GB of main memory and processor Intel Core i5, performing on Windows and Linux (Ubuntu). It isn't possible compare these values to values from other studies reported in literature review because they haven't done this kind of analysis.

4.4 Discussion

It is possible to notice on Table 1 that there are cases in that more than one input type provide similar results. In this cases, it is advisable using the input type that consume the smallest amount of memory. For example, in classification of normal voices faced with voices with Edema, it is advisable using only temporal parameters, because they consume less memory - the growth of model is smaller than by using temporal parameters and entropy.

It should also be highlighted the strong participation of temporal parameters in the most of cases exhibited at Table 1, either singly or in combination with others parameters, what indicates the big potential of this type of data on classifying speech pathologies.

Considering Table 2, it is possible to realize that the best results are from classifications that involved normal voices signals (first four lines), which were executed in all reports cited in literature review briefly presented in Section 1. Nevertheless, in some of them, only one pathology was considered, especially Edema. In this paper, such as in Costa (2008) and Lima et al. (2012), classifications were performed considering one or more pathologies. Through them, it was obtained results between 92,3 and 100%. All normal voices signals were correctly classified, a result only reached by Costa (2008), Marinus (2010), Tavares et al. (2011) (in some of cases presented by them) and Lima et al. (2012).

On classifying pathological voices signals against normal voices signals, it were obtained results between 92 and 96%, what may be considered very good, but equal or lesser than best results obtained by Costa (2008), Marinus (2010) and Arias-Londoño et al. (2011), which some of obtained percentages were bigger than 96%, reaching up 100%.

These results only confirm potential of PPM to classify vocal patterns. However, an important result consists of discriminating distinct pathologies accurately. This type of classification is considered more difficult and wasn't performed by almost none of authors previously cited. The results obtained in this paper can be considered good or median, depending of classification. The best results are originating from classifications aiming to discriminate other pathologies faced with Edema, within range 75 to 84,6%. The worst results are from classifications in which the objective was to diagnose voice signals presenting Edema. It is due to little representativity of this pathology in used database, since among 43 files, only 3 contain voices that present only Edema. All other files present other pathologies. Some of them present 5 (five) other pathologies. The more pathologies presents a voice record, the less this record represents its main pathology (BRANDT, 2012), what may to confuse classifiers, decreasing results.

These results are near to results obtained by Marinus (2010), but smaller than results obtained by Costa (2008), that obtained percentages bigger than 90%, and can be considered unsatisfactory for real clinical environment.

5. CONCLUSION

This work aimed to present a new approach for using an already known method in data compression. It is possible to realize, through results presented in Section 4, that this methodological approach is promising, since big part of obtained results was above 90%. However, in discriminating pathologies, additional research is needed, since results can be considered unsatisfactory for use in real clinical environments.

As future works, it is recommended to search input types and/or combinations that enable better distinction among pathologies, that is, that be affected in different ways by different pathologies to which the phonatory system is susceptible.

ACKNOWLEDGEMENT

We thank to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support during period of research, and to Universidade Federal de Campina Grande (UFCG), especially to prof. Benedito, for making database available.

REFERENCES

Arias-Londoño, J. D. et al., 2011. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *In IEEE Transactions on Biomedical Engineering*, Vol. 58, No. 2, pp. 370-379.

Barufaldi, B. et al., 2009. Text classification by literary period using PPM-C data compression. *Proceedings of Seventh Brazilian Symposium in Information and Human Language Technology*. São Carlos, Brazil, pp. 125-133.

Bennett, S. et al., 1987. Phonatory characteristcs associated with bilateral diffuse polypoid degeneration. *In The Laryngoscope*, Vol. 97, No. 4, pp. 446-450.

Brandt, R. R., 2012. *Classificação de vozes patológicas utilizando análise paramétrica e não-paramétrica*. Doctoral Thesis, Universidade Federal de Campina Grande, Campina Grande, Brazil.

Costa, S. L., 2008. *Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas*. Doctoral Thesis, Universidade Federal de Campina Grande, Campina Grande, Brazil.

Costa, W. C. et al., 2012. Pathological voice assessment by recurrence quantification analysis. *Proceedings of Biosignals and Biorobotics Conference*. Manaus, Brazil, pp. 1-6.

Coutinho, B. C. et al., 2005. Atribuição de autoria usando PPM. *Proceedings of XXV Congresso da Sociedade Brasileira de Computação*. São Leopoldo, Brazil, pp. 2208-2217.

Farrús, M. and Hernando, J., 2008. Using jitter and shimmer in speaker verification. *In IET Signal Processing*, Vol. 3, No. 4, pp. 247-257.

Fechine, J. M., 2000. *Reconhecimento automático de identidade vocal utilizando modelagem híbrida: paramétrica e estatística*. Doctoral Thesis, Universidade Federal de Campina Grande, Campina Grande, Brazil.

Honório, T. C. et al., 2009. Texture classification using prediction by partial matching models. *Proceedings of Workshop de Visão Computacional*. São Paulo, Brazil.

Hu, Y. and Loizou, P. C., 2008. Evaluation of objective quality measures for speech enhancement. *In IEEE Transactions on Audio, Speech and Language Processing*, Vol. 16, No. 1, pp. 229-238.

Jain, R., 1991. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley-Interscience, New York.

Kay Elemetrics, Kay elemetrics corporation disordered voice database, Model 4337, 03 Ed., 1994.

Lima, J. S. et al., 2012. Classificação de sinais vozes patológicas por meio do parâmetro de Hurst e LDA. Proceedings of XXX Simpósio Brasileiro de Telecomunicações. Brasília, Brazil.

Lopes, J. et al., 2008. A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa. *Proceedings of I Jornadas sobre Tecnologia e Saúde*. Guarda, Portugal.

Marinus, J. V., 2010. *Estudo de técnicas para classificação de vozes afetadas por patologias*. Master Thesis, Universidade Federal de Campina Grande, Campina Grande, Brazil.

Medeiros, T. F. L. et al., 2011. Heart arrhythmia classification using the PPM algorithm. *Proceedings of Biosignals and Biorobotics Conference*. Vitória, Brazil, pp. 1-5.

Oates, J., 2009. Auditory-perceptual evaluation of disordered voice quality. *In Folia Phoniatrica et Logopaedica*, Vol. 61, No. 1, pp. 49-56.

Orozco, J. R. et al., 2012. Voice pathology detection in continuous speech using nonlinear dynamics. *Proceedings of International Conference on Information Science, Signal Processing and their Applications*. Montreal, Canada, pp. 1030-1033.

Parsa, V. and Jamieson, D., 2000. Identification of pathological voices using glottal measures. *In Journal of Speech and Hearing Research*, Vol. 43, pp. 469-485.

Patil, H. A. and Baljekar, P. N., 2012. Classification of normal and pathological voices using TEO phase and mel cepstral features. *Proceedings of International Conference on Signal Processing and Communications*. Bangalore, India, pp. 1-5.

Raju, N. et al., 2012. Normal versus pathology voice-an analysis. *Proceedings of International Conference on Computing, Communication and Applications*. Dindigul, Tamilnadu, India, pp. 1-4.

Roads, C. 1995. *Computer Music Tutorial*. MIT Press, Massachusetts.

Salomon, D., 2004. *Data Compression: The Complete Reference*. Springer, New York.

Scalassara, P. R. et al., 2009, Predictability analysis of voice signals: analyzing healthy and pathologic samples. *In IEEE Engineering in Medicine and Biology Magazine*, Vol. 28, pp. 30-34.

Tavares, R. et al., 2011. Combining entropy measurements and cepstral analysis for pathological voice assessment. *Proceedings of Biosignals and Biorobotics Conference*. Vitória, Brazil, pp. 1-5.

Teixeira, J. P. et al., 2011. Análise acústica vocal - determinação do jitter e shimmer para diagnóstico de patologias da fala. *Proceedings of Congresso Luso-Moçambicano de Engenharia*. Maputo, Moçambique, pp. 139-140.

Vieira, J. V. et al., 2012. Análise de quantificação de recorrência e análise discriminante aplicadas à classificação de sinais de vozes saudáveis e sinais de vozes patológicas. *Proceedings of VII Congresso Norte Nordeste de Pesquisa e Inovação*. Palmas, Brazil.

Anexo D

Classificação de Patologias da Fala a partir do PPM

Hildegard Paulino Barbosa, Joseana Macêdo Fechine, José Eustáquio Rangel
Centro de Engenharia Elétrica e Computação
Universidade Federal de Campina Grande (UFCG)
Campina Grande, Brasil
hildegardpaulino@gmail.com, joseana@dsc.ufcg.edu.br, rangeldequeiroz@gmail.com

***Abstract*—Speech organs are very susceptible to several types of pathologies, which may harm voice production. Several techniques have been traditionally used to detect these pathologies. However, they present drawbacks concerning the accuracy and the comfort of patients during application. Moreover, results obtained by computing techniques have not yet matured to a reliable tool for application in clinics. In this research, a classification approach based on a method not previously employed in classification of vocal tract diseases is proposed. It is based on Prediction by Partial Matching (PPM), which uses acoustical and temporal features to feed models. It were obtained very promising results in the presence or absence of pathologies (at least 92%). With regard to pathology discrimination, preliminary results confirmed that PPM is a high potential technique for voice pathology classification, although its clinical application for the diagnosis of voice pathologies still needs deeper investigation.**

Keywords—speech pathologies; prediction by partial matching; acoustical and temporal features;

I. INTRODUÇÃO

A voz é o meio de comunicação mais importante e mais natural do ser humano, a partir da qual são expressas vontades, pensamentos, ordens e informações. Entretanto, para que a comunicação seja efetiva, é necessário o entendimento correto da voz enunciada por parte do interlocutor do processo. Se isto não ocorrer, haverá maior propensão a equívocos, o que desestimulará a comunicação causando, até mesmo, o constrangimento do locutor. Tal problema, denominado *disfonia*, é causado muitas vezes por patologias da fala, às quais o sistema fonador humano é muito suscetível. Estima-se que entre 3 e 10% da população geral tenha o sistema fonador comprometido por alguma patologia [1], além do que é comum que o mesmo indivíduo possa ser acometido por até 8 patologias [2], as quais podem ser causadas por alterações psicoemocionais, doenças neurodegenerativas, mau uso da voz ou hábitos sociais não saudáveis, tais como o tabagismo e a ingestão de álcool [1] [3] [4]. Algumas destas razões explicam a ocorrência mais freqüente de patologias da fala em fumantes e em categorias de profissionais que utilizam a voz como seu principal instrumento de trabalho, e.g., professores, cantores, radialistas, jornalistas [4]. Atualmente, há conhecimento de mais de 120 patologias [5], mas as mais conhecidas são *Nódulo*, *Edema*, *Paralisia* e *Pólipo*.

Na detecção de patologias da fala, são usados, tradicionalmente, dois tipos de mecanismos. O primeiro, consiste da escuta da elocução vocal do paciente por um profissional (normalmente, um fonoaudiólogo ou um otorrinolaringologista), visando a decidir sobre a presença ou ausência de uma patologia. Até poucos anos, este era o método mais empregado [6]. Contudo, não é difícil perceber seu caráter altamente subjetivo e propenso à indução de erros, principalmente nos casos em que a patologia se encontra em estágios iniciais, devido à alta dependência da experiência, da acurácia, do nível de fadiga e da sensibilidade do sistema auditivo do profissional[a] [7] [8].

O segundo mecanismo consiste de procedimentos clínicos nos quais a voz do paciente é avaliada por meio de recursos visuais. Dentre os exames mais comuns desta natureza estão a *videolaringoscopia* e a *videoestroboscopia* [3]. Embora precisos, estes exames são bastante invasivos e desconfortáveis para o paciente, causando, em alguns casos, a ação de reflexo durante a aplicação, em função de sua sensibilidade laríngea, o que pode acarretar falsos diagnósticos [4]. Além disto, comprometem financeiramente ambas as partes, já que os equipamentos requeridos para executá-los são caros e sofisticados, obrigando o repasse dos custos ao paciente e restringindo seu acesso a grande parte da população.

Inúmeras pesquisas sobre a detecção de patologias da fala por computador têm sido desenvolvidas, com o intuito de reduzir significativamente a necessidade e a frequência de exames visuais. Dentre as técnicas utilizadas, incluem-se Modelos de Markov Escondidos (*Hidden Markov Models - HMM*) [3], Quantização Vetorial, Redes Neurais [4], Gráficos de Recorrência [1] [9], Modelos de Misturas Gaussianas (*Gaussian Mixture Models - GMM*) [4] [10], fase TEO [11], Máquinas de Suporte Vetorial [12], Dinâmica Não Linear com medidas de complexidade [13], Expoente de Hurst [14] dentre outras. Porém, um tipo de abordagem sobre a qual não foi

[a] Diferentes diagnósticos podem ser dados por diferentes profissionais ou, até mesmo, pelo mesmo profissional, em ocasiões diferentes.

encontrado registro na revisão de literatura foi o uso de métodos de compressão de dados. Embora eles tenham sido projetados inicialmente para comprimir dados, percebeu-se que o rico modelo estatístico gerado por alguns destes algoritmos pode ser empregado também em atividades de classificação. Por esta razão, se afigura importante o estudo da eficácia deste tipo de método na discriminação de patologias.

Um dos métodos mais eficazes de compressão de dados é a Predição por Casamento Parcial (*Prediction by Partial Matching* - PPM) [18]. Seu princípio de funcionamento será descrito na Seção II.B. Contudo, deve-se considerar que bons resultados têm sido obtidos a partir do seu uso em atividades de compressão e classificação de arquivos binários, textos, sinais de eletrocardiograma e imagens, dentre outros tipos de sinais [15][16][17][18].

II. Fundamentação Teórica

Nesta seção, serão inicialmente apresentadas medidas extraídas dos sinais de voz, as quais serão utilizadas como entradas para os modelos utilizados na etapa de classificação. Em seguida, será apresentado brevemente o funcionamento do método proposto.

A. Parâmetros Utilizados

É importante mencionar de antemão que os sinais de voz apresentam invariabilidade estatística em janelas de até 32 ms porque, mesmo sendo a voz invariante no tempo, o trato vocal apresenta natureza dinâmica[b], o que afeta os parâmetros que representam a voz e, consequentemente, sua produção. Sendo assim, para que os valores extraídos representem a realidade acerca do sinal de voz manipulado, todas as medidas consideradas foram extraídas em janelas dentro daquele intervalo, de modo que um arquivo era representado por um conjunto de valores de determinada medida.

Energia - a soma dos quadrados dos valores de amplitude. Para obter seu valor em decibéis, basta multiplicar o logaritmo do valor encontrado por 10.

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

$$E_{seg}(dB) = 10 \cdot \log[E_{seg}]$$

Em (1), N_A é o tamanho da janela e s(n) é a n-ésima amostra do sinal de voz (amplitude).

Taxa de Cruzamento por Zero - número de vezes em que a forma de onda do sinal cruza o eixo das abscissas (tempo), i.e, quantas vezes um valor positivo de amplitude é sucedido por um negativo e vice-versa.

$$\begin{aligned} TCZ &= N_A \cdot E\{\text{sgn}[s(n)] - \text{sgn}[s(n-1)]\} \\ &= \sum_{n=1}^{N_A-1} |\,\text{sgn}[s(n)] - \text{sgn}[s(n-1)]\,| \end{aligned} \qquad (2)$$

em que

$$\text{sgn}[s(n)] = \begin{cases} 1, & \text{se s(n)} \geq 0 \\ -1, & \text{se s(n)} < 0 \end{cases} \qquad (3)$$

Os elementos dessa equação apresentam o mesmo significado dos elementos de (1).

Número Total de Picos (NTP) - quantidade de picos, positivos e negativos, existente na forma de onda do sinal.

Diferença no Número de Picos (DNP) - difere da anterior no sinal usado na operação: enquanto no NTP somam-se as quantidades de picos, na DNP subtraem-se dos picos positivos os picos negativos

Frequência Fundamental - O som é decorrente de vibrações mecânicas que se propagam por meio da interação com um meio físico. Se há um padrão nessas vibrações, diz-se que o som tem uma forma de onda *periódica*. Se não há nenhum padrão, ele é classificado como ruído. À repetição de uma forma de onda periódica dá-se o nome de *ciclo*. O número de ciclos por segundo que ocorre na transmissão de um som representa a *Frequência Fundamental* desse som [19].

No contexto de sinais de voz, F_0 representa a frequência na qual as dobras vocais vibram (abrem-se e fecham-se a cada ciclo) e influi diretamente na tonalidade da voz. Homens apresentam valores de Frequência Fundamental mais baixos (entre 100 e 137 Hz) e mulheres, mais altos (entre 177 e 244 Hz), enquanto crianças apresentam valores ainda mais elevados (entre 206 e 281 Hz). Porém, esses valores ocorrem apenas em vozes saudáveis, pelo fato de as patologias da laringe afetarem o padrão vibratório das dobras vocais, sempre reduzindo sua velocidade de vibração e, consequentemente, o valor da Frequência Fundamental da voz. Em [20] foi feita uma pesquisa com um grupo de homens e mulheres que apresentavam Edema e percebeu-se que as mulheres tinham uma freqüência fundamental média de 108 Hz, enquanto o grupo de homens com a mesma patologia apresentou Freqüência Fundamental média de 91 Hz.

Jitter - consiste da perturbação dos valores de período fundamental (T_0) do sinal. Sendo assim, para obtê-lo, devem ser extraídos primeiramente os valores de T_0 (inverso de F_0) de pequenas janelas do sinal e em seguida verificar, grosso modo, quanto difere cada valor de seus vizinhos. O *Jitter* consiste no somatório dessas diferenças, sendo útil para verificar a estabilidade do sistema fonador, mais especificamente da vibração das dobras vocais, reduzida na presença de patologias [21]. Há 4 tipos de *Jitter* comumente usados. As equações (4) a (7) são utilizadas na obtenção de cada um.

$$jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |\,T_i - T_{i-1}\,| \qquad (4)$$

$$jitt = \frac{\dfrac{1}{N-1} \displaystyle\sum_{i=1}^{N-1} |\,T_i - T_{i-1}\,|}{\dfrac{1}{N} \displaystyle\sum_{i=1}^{N} T_i}$$

[b] Sua configuração varia com o tempo durante a produção da fala

$$rap = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - (\frac{1}{3}\sum_{n=i-1}^{i+1}T_n)|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \quad (6)$$

$$ppq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}|T_i - (\frac{1}{5}\sum_{n=i-2}^{i+2}T_n)|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \quad (7)$$

Nessas equações, T_i é o valor do período fundamental da i-ésima janela e N é o tamanho da janela.

Shimmer - similar ao *Jitter*, mas empregada na análise da perturbação existente nos valores de amplitude dos picos do sinal. A forma de obter os valores de amplitude para o cálculo é similar: a partir de segmentos do sinal (janelas), é obtida a distância entre os dois picos desse segmento (o mais alto e o mais baixo). Esta medida é útil para verificar a estabilidade da intensidade vocal, afetada pela pressão subglótica e, por sua vez, influenciada pela amplitude de vibração e pela tensão das dobras vocais [22].

Semelhante ao *Jitter*, existem vários tipos de *Shimmer*. As equações (8) a (10) são utilizadas na obtenção de cada um. Vale ressaltar que no cálculo do APQ, foram encontradas várias quantidades de amplitudes adjacentes consideradas, além da utilizada (10): 5, 11 e 55.

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \quad (8)$$

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}|20\cdot\log(A_{i+1}/A_i)| \quad (9)$$

$$apq3 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - (\frac{1}{3}\sum_{n=i-2}^{i+2}A_n)|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \quad (10)$$

Nessas equações, A_i é o valor do período fundamental da i-ésima janela e N é o tamanho da janela.

Relação Harmônico Ruído (*Harmonic-to-Noise Ratio - HNR*) - permite avaliar a quantidade de ruído presente em um sinal, em comparação a sua componente periódica. Se mostra adequada ao contexto do processamento digital de sinais de voz, pelo fato de a componente periódica ser decorrente da vibração das dobras vocais e a aperiódica, do ruído glótico [7]. As patologias interferem diretamente no padrão vibratório das dobras vocais, deixando a voz com um aspecto ruidoso, de modo que uma baixa HNR representa um forte indício da presença de uma patologia. Em [23] há a afirmação de que 83% das vozes patológicas apresentam baixa HNR, além de que quanto mais ruído um sinal de voz apresenta, mais avançado é o estágio da patologia.

Análise por Predição Linear - considerada uma das técnicas mais poderosas de análise da fala, parte da ideia de que é possível estimar uma amostra de um sinal de voz por combinação linear de amostras passadas. Isso significa que o sinal como um todo pode ser reconstruído, a partir da estimação das amostras. Essa estimativa deve ser feita de modo a minimizar a soma das diferenças quadradas entre as amostras reais e amostras preditas, o que permite encontrar um conjunto único de p coeficientes preditores. Quanto maior o valor de p, mais precisas serão as estimativas e, consequentemente, mais íntegro o sinal reconstruído. É comum seu uso na transmissão de fala, com o intuito de possibilitar a transmissão em uma baixa taxa de bits.

A Análise por Predição Linear é originária da concepção da produção da fala como um sistema linear variante no tempo, excitado por um trem de pulsos quase periódicos e ruído aleatório. A predição linear provê métodos robustos e precisos para estimar os parâmetros que caracterizam este sistema, dentre eles a Codificação por Predição Linear (*Linear Predictive Coding – LPC*), os quais foram usados neste trabalho.

Entropia - medida da incerteza de uma variável aleatória. È fundamental para a Teoria da Informação, sendo empregada como o limiar de compressão máximo possível de ser obtido na compressão de uma mensagem proveniente de uma fonte de informação. O cálculo da entropia é feito por (11).

$$H = -\sum_{i=0}^{N} p_i \log_2(p_i) \quad (11)$$

Em (11), p_i é a probabilidade de surgimento do i-ésimo símbolo da fonte de informação analisada e N é o tamanho do alfabeto dessa fonte.

B. Método de Compressão PPM

O PPM é um método adotado para comprimir fluxos de dados. Do ponto de vista do poder de compressão, é o compressor de dados mais eficaz, i.e., o estado da arte na área [18] [24]. Seu uso em escala comercial ainda é muito limitado, restringindo-se ao âmbito da pesquisa acadêmica, pelo fato de ser armazenado, durante sua execução, um modelo muito preciso da fonte de dados sendo comprimida, o que acarreta alto consumo de memória e velocidade de execução relativamente baixa, principalmente se comparado com os compressores mais utilizados.

Métodos estatísticos destinados à compressão de dados, tal como o PPM, podem ter sua operação dividida em duas etapas: *modelagem* e *codificação* [24]. A *modelagem* consiste no armazenamento das probabilidades de símbolos originários de um fluxo de dados. Tendo em vista que o modelo PPM é contextual, também podem ser armazenadas as probabilidades de sequências dos símbolos. A intenção é estimar, o mais precisamente possível, as probabilidades da fonte de dados estacionária abstrata que gerou o fluxo de dados. A escolha do tamanho máximo de contexto armazenado é livre, mas vale salientar que o consumo de memória cresce exponencialmente com esse valor. Além disso, contextos muito altos guardam muitas informações que aparecem com pouca freqüência e, segundo [24], a curva de aprendizado do modelo PPM cessa seu crescimento a partir de um dado tamanho de contexto, o que ocasiona redução na taxa de compressão. Sendo assim, merece atenção o tamanho máximo de contexto empregado em qualquer utilização do PPM.

A segunda etapa de um processo de compressão estatístico é a codificação, que normalmente consiste em atribuir códigos a cada símbolo, visando a gerar um fluxo de dados menor que o original. Para tanto, os símbolos e sequências mais frequentes (identificados na etapa de modelagem) devem receber os menores códigos, com o mínimo de bits possível, enquanto os menos frequentes recebem os maiores. Entretanto, para que seja obtido um código comprimido de tamanho ideal, é necessário atribuir quantidades fracionárias de bits a cada símbolo do fluxo original. O codificador Aritmético alcança este objetivo ao transformar a mensagem original em um número com quantidade ilimitada de casas decimais, dentro de um intervalo limitado durante sua operação. No início do processo, o intervalo é [0, 1). À medida que mais probabilidades são recebidas, este intervalo diminui, mas o número de casas decimais de cada limite aumenta. No fim, tem-se um intervalo pequeno, mas com limites contendo várias casas decimais. O fluxo comprimido pode ser qualquer número dentro deste intervalo. Dessa forma, são atribuídas quantidades fracionárias de bits para cada símbolo do fluxo original.

III. Materiais e Métodos

Nesta pesquisa, foi utilizada a base de dados da Kay Elemetrics, gravada no Voice Speech Lab da Massachussets Eye and Ear Infirmary (MEEI) [2]. Esta base é composta por 1.381 arquivos, em formato NSP, com 16 bits/amostra, dos quais 666 contêm o pronunciamento da vogal *a* sustentada (conjunto utilizado nesta pesquisa). Desses, 53 correspondem a sinais de vozes saudáveis (cada um dos quais com 3s de duração) e o restante a sinais de vozes patológicas (com duração de 1s). É importante salientar que não há unanimidade entre as frequências de amostragem (taxas de amostragem) dos arquivos. Todos os arquivos contendo sinais de vozes saudáveis têm frequência de amostragem de 50 kHz, ou seja, 50 mil amostras por segundo. Os arquivos contendo sinais de vozes patológicas, por sua vez, têm frequência de amostragem de 25 kHz, mas alguns também têm 50 kHz.

Dos 666 arquivos que contêm a elocução da vogal *a* sustentada, foram utilizados 189, sendo 53 de vozes Normais, 34 de vozes com Edema, 46 de vozes com Paralisia e 56 de vozes com Outras patologias. Em uma classificação, 60% dos arquivos são utilizados para treinamento e os 40% restantes, para testes. A seleção dos arquivos para treinamento é feita aleatoriamente, com o intuito de evitar vieses nos resultados e de ser obtida uma boa estimativa de erro. Todas as classificações incluem etapas como segmentação (com superposição de 50%) e janelamento, que consistem, respectivamente, de extrair apenas uma parte (janela) do sinal e diminuir o efeito das extremidades de cada janela, mantendo o do centro. Os modelos PPM são alimentados com medidas extraídas de cada janela.

A primeira etapa da investigação consistiu em identificar o melhor tipo de entrada para cada classificação executada, dentre parâmetros temporais, acústicos, entropia, coeficientes LPC e combinações entre eles. A identificação do melhor tipo de entrada foi levado a efeito utilizando Projeto Experimental de fator único, apresentado por [25].

A segunda etapa consistiu em analisar o impacto de atividades de pré-processamento como pré-ênfase, distinção de gênero (classificações utilizando sinais de vozes de um único gênero) e subamostragem (utilização de apenas metade das amostras dos sinais que contém 50 kHz de frequência de amostragem), sobre os resultados, isto é, se implicam ou não em ganho significativo dos percentuais. Paralelamente, foi investigada também a viabilidade do aumento do tamanho do contexto do classificador, já que ele implica em aumento da utilização de recursos computacionais. Em outras palavras, se o aumento desse tamanho implica em aumento significativo dos percentuais de acerto.

A investigação se deu mediante a comparação par a par entre dois casos, sendo um deles advindo da etapa anterior. Por exemplo, caso na etapa anterior tenha se verificado que os melhores resultados da classificação de sinais de vozes Normais quando confrontados com sinais de vozes com Edema foram obtidos utilizando apenas coeficientes LPC, nesta etapa, estes resultados foram comparados par a par com os resultados obtidos ao aplicar filtro de pré-ênfase, distinção de gênero e subamostragem e ao utilizar tamanhos de contexto entre 1 e 4.

As análises foram feitas utilizando Intervalos de Confiança (procedimento referido em [25] como *t-test*) ou teste de Mann-Whitney, a depender da distribuição amostral dos resultados comparados. A utilização destes testes é justificada pelo fato de que o intuito desta investigação era verificar se a aplicação de atividades de pré-processamento ou aumento do tamanho do modelo forneciam resultados significativamente maiores que os resultados obtidos nas classificações sem essas etapas, de modo a justificar sua inclusão.

Por fim, identificada a melhor configuração do classificador para cada caso, foram obtidos percentuais que caracterizassem a eficácia do PPM no contexto de diagnóstico de patologias da fala, por meio do procedimento conhecido como Validação Cruzada com 4 parcelas. Nesta etapa, diferentemente das anteriores, não é possível inserir aleatoriedade, pela própria natureza do processo.

IV. Resultados

Nesta seção serão apresentados e discutidos os resultados obtidos em todas as etapas da investigação.

A. *Identificação do melhor tipo de entrada*

Nesta etapa da investigação, objetivou-se identificar o melhor tipo de entrada para cada caso de classificação. Conforme mostrado na Tabela 1, descobriu-se que não há um tipo de entrada que retorne os melhores percentuais em todos os casos. O melhor tipo de entrada varia com a classificação executada.

Tabela 1. Os melhores tipos de entrada e os resultados obtidos

Classificação	Mediana em % da primeira classe e tipo de entrada	Mediana em % da segunda classe e tipo de entrada
Normal x Tudo	100,0 - Temporais; Entropia; Temporais e Entropia	96,3 - Temporais e Entropia
Normal x Edema	100,0 - Temporais; Temporais e Entropia	100,0 - Temporais e Entropia
Normal x Paralisia	100,0 - Temporais; Temporais e Entropia	94,7 - Temporais
Normal x Outras	100,0 - Temporais; Temporais e Entropia	95,6 - Temporais
Edema x Paralisia	64,2 - LPC	73,7 - Temporais e Acústicos
Edema x Outras	57,1 - LPC	65,2 - Acústicos
Paralisia x Outras	68,4 - Temporais	65,2 - Temporais, Acústicos e Entropia

É possível perceber que há casos em que mais de um tipo de entrada fornece resultados semelhantes. Para tanto, é recomendável a utilização daquele que utilize menos memória. Por exemplo, na classificação de arquivos Normais confrontados com arquivos de vozes com Edema, recomenda-se a utilização apenas de parâmetros temporais, por utilizarem menos memória.

É válido destacar também a forte participação dos parâmetros temporais na maioria dos casos, seja isoladamente ou em combinação com outros parâmetros, o que indica o potencial deste tipo de dado na classificação de patologias da fala.

B. *Influência de pré-processamentos nos resultados*

Nenhuma atividade de pré-processamento executada implicou o aumento significativo dos resultados. Ao invés disto, notou-se com frequência a redução significativa nos resultados devido ao pré-processamento dos sinais de entrada. Cenário semelhante foi observado na variação do tamanho do contexto do classificador. Sendo assim, é possível concluir que, em todos os casos, a melhor configuração consistiu da não inclusão de nenhuma etapa de pré-processamento e da adoção do contexto 0 em todo o processo.

C. *Percentuais via validação cruzada*

Tendo sido encontradas as melhores e mais viáveis configurações para cada caso de classificação, foi utilizado o procedimento de Validação Cruzada, a fim de obter os percentuais de acerto para cada um deles, com o intuito de caracterizar o PPM quanto à eficácia, na classificação de patologias da fala. Os resultados obtidos encontram-se na Tabela 2

Tabela 2. Percentuais obtidos com Validação cruzada utilizando as melhores configurações

Classificação	Resultados obtidos (mediana %)
Normal x Tudo - Normal	100,0
Normal x Tudo - Tudo	94,1
Normal x Edema - Normal	100,0
Normal x Edema - Edema	95,0
Normal x Paralisia - Normal	100,0
Normal x Paralisia - Paralisia	92,3
Normal x Outras - Normal	100,0
Normal x Outras - Outras	96,4
Edema x Paralisia - Edema	50,0
Edema x Paralisia - Paralisia	84,6
Edema x Outras - Edema	40,0
Edema x Outras - Outras	75,0
Paralisia x Outras - Paralisia	57,6
Paralisia x Outras - Outras	64,2

D. *Eficiência da abordagem metodológica*

A eficiência desta abordagem metodológica pode ser analisada quanto à velocidade de execução e à utilização de memória.

Quanto à velocidade de execução, foram analisados os tempos de execução de 10 classificações com as classes mais bem representadas da base de dados: Normal e Tudo, com 53 e 56 arquivos cada uma. Essas classes foram escolhidas por serem as mais bem representadas, de modo que os tempos das outras classificações serão certamente menores que os relatados na Tabela 3. Os intervalos apresentados correspondem aos intervalos de confiança a 95% de significância estatística dos tempos de execução do treinamento e de todo o processo, o que inclui treinamento e testes com vários sinais de voz.

Com relação ao uso de memória, também foi registrada a utilização máxima de memória em 10 execuções desta mesma classificação. O intervalo de confiança a 95% de significância destes registros foi extraído, tendo sido obtido o intervalo [15,02; 15,77] MB. A ferramenta de monitoramento utilizada

foi a VisualVM, fornecida pela distribuidora da plataforma Java.

TABELA 3. TEMPOS DE EXECUÇÃO DA CLASSIFICAÇÃO NORMAL X TUDO

Etapa	Intervalo de confiança (ms)
Treinamento	[405,3; 436,6]
Todo o processo	[804,0; 844,0]

E. Discussão

É possível constatar que os melhores resultados estão associados a classificações entre sinais de vozes Normais e sinais de vozes que apresentam alguma patologia, executadas com o intuito de detectar a presença de patologias. Nelas, obtiveram-se resultados entre 92,3 e 100%. Porém, estes resultados apenas confirmam o potencial do PPM no contexto da classificação de padrões vocais, haja vista que resultados semelhantes foram alcançados em pesquisas anteriormente conduzidas, tais como [1] [3] [4].

Um resultado importante, almejado por pesquisadores da área, consiste da discriminação precisa de patologias distintas. Na pesquisa ora descrita, foram obtidos resultados deste tipo que podem ser considerados bons e medianos, a depender da classificação. Os piores resultados são oriundos das classificações em que o intuito é diagnosticar sinais de voz com Edema. Isto se deve à pequena representatividade desta patologia na base de dados utilizada, haja vista que dos 43 registros de arquivos de vozes com Edema, apenas 3 são de vozes que apresentam apenas a patologia Edema. Todos os demais apresentam outras patologias, além desta. Alguns deles chegaram a apresentar outras 5 (cinco) patologias. Quanto mais patologias um registro de voz apresenta, menos tal registro representará a patologia principal que contém [5], o que pode confundir os classificadores utilizados, comprometendo os resultados.

Contudo, bons resultados de discriminação de patologias também foram obtidos, especialmente aqueles de classificações com o intuito de discriminar outras patologias em confronto com sinais de vozes com Edema, que se encontram entre 75 e 84,6%. Mesmo assim, eles podem ser considerados insuficientes para um ambiente clínico real.

V. CONSIDERAÇÕES FINAIS

O presente trabalho visou a apresentar uma nova abordagem de utilização para um método já conhecido no contexto da compressão de dados. É possível constatar, pelos resultados apresentados na Seção IV, que a abordagem metodológica proposta se mostra promissora, haja vista que boa parte dos resultados de classificação obtidos é superior a 90%. Porém, os casos de discriminação de patologias ainda carecem de investigação adicional, já que os resultados podem ser considerados inadequados para a diagnose em ambientes clínico reais.

Como sugestão de trabalhos futuros, recomenda-se a identificação de entradas e/ou combinações que permitam melhor distinção entre patologias, isto é, que sejam afetadas de forma diferente para as diferentes patologias às quais o sistema de produção da fala humano é suscetível.

AGRADECIMENTOS

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro durante o período da pesquisa, e à Universidade Federal de Campina Grande (UFCG), especialmente ao prof. Benedito, pela disponibilização da base de dados.

REFERÊNCIAS

[1] W. C. de A. Costa, F. M. Assis, B. G. Aguiar Neto, S. C. Costa e V. J. D. Vieira, "Pathological voice assessment by recurrence quantification analysis", BRC 2012, in press.

[2] K. Elemetrics, "Kay elemetrics corp. disordered voice database", Model 4337, 03 Ed., 1994.

[3] S. L. N. C. Costa, "Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas", Tese (Doutorado em Engenharia Elétrica) - UFCG, Campina Grande, 2008.

[4] J. V. M. L. Marinus, "Estudo de técnicas para classificação de vozes afetadas por patologias", Dissertação (Mestrado em Ciência da Computação) - UFCG, Campina Grande, 2010.

[5] R. R. Brandt, "Classificação de vozes patológicas utilizando análise paramétrica e não-paramétrica", Tese (Doutorado em Engenharia Elétrica) - UFCG, Campina Grande, 2012.

[6] Y. Hu e P. C. Loizou, "Evaluation of objective quality measures for speech enhancement", IEEE Trans. on Audio, Speech and Lang. Proc., vol. 16, pp. 229-238, 2008.

[7] J. Lopes et al., "A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa", I JTS, in press.

[8] J. Oates, "Auditory-perceptual evaluation of disordered voice quality", Fol. Phon. et Logop., vol. 61, pp. 49-56, 2009.

[9] J. V. Vieira, S. C. Costa e W. C. Costa, "Análise de quantificação de recorrência e análise discriminante aplicadas à classificação de sinais de vozes saudáveis e sinais de vozes patológicas", VII CONNEPI, in press.

[10] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz e G. Castellanos-Domínguez, "Automact detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients", IEEE Trans. on Biom. Engin., vol. 58, pp. 370-379, 2011.

[11] H. A. Patil e P. N. Baljekar, "Classification of normal and pathological voices using TEO phase and mel cepstral features", ICSPCS 2012, in press.

[12] N. Raju, T. L. Priya, S. Mathini e P. Preethi, "Normal versus pathology voice-an analysis", ICCCA 2012, in press.

[13] J. R. Orozco et al., "Voice pathology detection in continuous speech using nonlinear dynamics", ISSPA 2012, in press.

[14] J. dos S. Lima, T. T. C. Palitó, S. C. Costa e S. E. N. Correia, "Classificação de sinais vozes patológicas por meio do parâmetro de Hurst e LDA", XXX SBrT, in press.

[15] B. C. Coutinho, J. L. de M. Macêdo, A. Rique Júnior e L. V. Batista, "Atribuição de autoria usando PPM", XXV CSBC, in press.

[16] Barufaldi et al. "Text classification by literary period using PPM-C data compression", STIL 2009, in press.

[17] T. C. de S. Honório, L. V. Batista e R. C. M. Duarte, "Texture classificatoin using prediction by partial matching models", WVC 2009, in press.

[18] T. F. L. Medeiros et al. "Heart arrhythmia classification using the PPM algorithm", BRC 2011, in press.

[19] C. Roads, Computer Music Tutorial. Massachusetts: MIT Press, 1995.

[20] S. Bennett, S. Bishop e S. M. Lumpkin, "Phonatory characteristcs associated with bilateral diffuse polypoid degeneration", The Laryngoscope, vol. 97, pp. 446-450, 1987.

21] J. P. Teixeira, D. Ferreira e S. Carneiro, "Análise acústica vocal - determinação do jitter e shimmer para diagnóstico de patologias da fala", CLME 2011, in press.

22] M. Farrús e J. Hernando, "Using jitter and shimmer in speaker verification", IET Sig. Proc., vol. 3, pp. 247-257, 2008.

23] V. Parsa e D. Jamieson, "Identification of pathological voices using glottal measures", J. of Speech and Hear. Res., vol. 43, pp. 469-485, 2000.

24] D. Salomon, Data Compression: The Complete Reference, 3rd ed. New York: Springer, 2004.

25] R. Jain, The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling. New York: Wiley-Interscience, 1991.