**Universidade de São Paulo
Escola de Engenharia de São Carlos
Departamento de Engenharia Elétrica**

Luis Fernando Granato

Teoria Bayesiana na avaliação das condições da Laringe

DEDALUS - Acervo - EESC

31100052125

**São Carlos – SP
2005**

FOLHA DE JULGAMENTO

Candidato: Licenciado **LUIS FERNANDO GRANATO**

Tese defendida e julgada em 23-03-2005 perante a Comissão Julgadora:

Prof. Titular **JOSE CARLOS PEREIRA (Orientador)**
(Escola de Engenharia de São Carlos/USP)

APROVADO

Prof. Dr. **NELSON DELFINO D'ÁVILA MASCARENHAS**
(Universidade Federal de São Carlos/UFSCar)

APROVADO

Prof. Dr. **BAKER JEFFERSON MASS**
(Escola de Engenharia de São Carlos/USP)

APROVADO

Dra. **JENIFFER DE CÁSSIA RILLO DUTKA SOUZA**
(Hospital de Reabilitação de Anomalias Craniofaciais/USP)

APROVADO

Prof. Titular **FABIO VIOLARO**
(Universidade Estadual de Campinas/UNICAMP)

APROVADO

Prof. Associado **HOMERO SCHIABEL**
Vice-Coordenador em Exercício do Programa de Pós-Graduação
em Engenharia Elétrica

Profa. Titular **MARIA DO CARMO CALIJURI**
Presidente da Comissão de Pós-Graduação

Dedicatória Especial

**Dedico essa tese
a meu Pai Antonio,
minha Mãe Gracia,
minha irmã Claudia e
principalmente aos
amores da minha vida,
minha esposa Joanita e aos
nossos bebês que em breve
estarão alegrando nossas vidas.**

Agradecimentos

À Deus por mais uma oportunidade de vida.

Ao meu orientador professor José Carlos Pereira pela oportunidade de orientação e amizade auxiliadora que tanto acrescentaram para meu amadurecimento pessoal e profissional.

Aos amigos de laboratório, em especial Mori, Marcelo e Parê pela colaboração e grande ajuda na elaboração dessa tese.

Aos professores Nelson Mascarenhas e Baker pelas sugestões importantes.

À todos os funcionários do Departamento de Engenharia Elétrica pelo profissionalismo e compreensão.

À CAPES pelo apoio financeiro.

Enfim, agradeço a todos que de alguma forma auxiliaram na elaboração dessa tese.

Resumo

O uso de técnicas não invasivas está sendo utilizado na avaliação do trato vocal e da laringe como uma ferramenta de auxílio ao especialista na prevenção, diagnóstico e tratamento de alterações vocais e laríngeas. Dessa forma, esse trabalho apresenta o uso da teoria de decisão, mais precisamente da teoria Bayesiana na avaliação das condições da laringe.

Utilizando-se de sinais de voz diagnosticados, foram extraídas 38 medidas acústicas que representam informações numéricas do sinal. A partir dessas informações, em etapas univariada e multivariada (combinação das medidas acústicas), dentro de uma população de 4 patologias (edema, fenda, nódulo e pólipos) mais a voz normal, realizaram-se as discriminações individualmente e com a combinação das 38 medidas acústicas. Os resultados retornam percentuais de acerto das medidas acústicas na discriminação das 4 patologias mais a voz normal nas etapas citadas.

Os resultados mostraram o potencial do sistema implementado na discriminação das 4 patologias da laringe mais a voz normal.

Na etapa univariada, obteve-se 100% de acerto na discriminação apenas para a patologia pólipos e 90% para as outras 3 patologias e voz normal.

Na etapa multivariada, para a combinação 2a2 das medidas acústicas, apenas para a patologia edema obteve-se 90% de acerto na sua discriminação e para as outras 3 patologias e voz normal obteve-se 100% de acerto. Na combinação 3a3 das medidas acústicas, obteve-se 100% de acerto para as 4 patologias e voz normal, embora para a patologia edema apenas uma única combinação resultou em 100% de acerto. Dessa forma, realizou-se a combinação 4a4 das medidas acústicas na tentativa de se obter um número maior de combinações que resultassem em 100% de acerto na discriminação da patologia edema e o resultado foi positivo.

Na etapa multivariada, o processo de combinação se deu até 4a4 pelo fato de conseguirmos obter 100% para todas as 4 patologias mais a voz normal envolvidas na análise e dessa forma, não continuamos as combinações por questões como tempo computacional e volume de dados excessivo para análise. Em suma, essa etapa contribuiu de forma incisiva para listar as medidas acústicas que melhor discriminaram as 4 patologias mais a voz normal em análise, portanto, esse trabalho obtém resultados inéditos até o momento na literatura.

Com isso, a conclusão é que o sistema implementado mostrou-se potencial como ferramenta de auxílio ao especialista em caráter de triagem e consulta.

Abstract

Non invasive techniques have been used in the evaluation of the vocal tract and larynx as an additional tool of aid in the prevention, diagnosis and treatment of vocal and larynx alterations. This work presents the use of decision theory, more precisely the Bayesian theory in the evaluation of the conditions of the larynx. From the diagnosed voice signals, 38 acoustic parameters were extracted which represent numerical information of the signals. This information in univariated and multivariated stages enabled to perform individual and combined discriminations of these parameters by using Bayesian theory as discriminator. By using the procedures, percentages of success in the discrimination of the analyzed pathologies were achieved.

These results have shown the potential of the implemented system in the discrimination of larynx pathologies, where 4 pathologies were analyzed (edema, fend, nodule and polyp) and also normal voice. In the univariated step, it was achieved 100% of success only for Polyp and 90% for the rest of the pathologies.

In the multivariated step, for 2 for 2 combinations only for edema it was achieved 90% of success in the discrimination, and the others 100%. In 3 for 3 combinations, 100% of success was reached for all cases analyzed, although for edema only a single combinations resulted in 100% of success. In 4 for 4 combinations was realized so when could get a larger number of combination with 100% of success in the discrimination.

In the multivariated step, the combination process was done until 4 for 4, where we could reach 100% of success for the 4 pathologies e normal voice, therefore, we have not continued the combinations due to high computational time for the process, as well as too many data for analysis. That step, was primordial to list the measure that best discrimination the 4 pathologies and normal voice in analysis, whose results have never published up to now.

With this work we can conclude that the system proved to be a potential tool in medical aid and for selection and assistance.

Sumário

1.1. Caracterização do problema	1
1.2. Revisão da literatura	6
1.3. Motivação e objetivos	5
1.4. Descrição técnica	7
2. Anatomia, Fisiologia e Patologia da laringe	18
2.1. Anatomia básica da laringe	18
2.1.1. Músculos da laringe	20
2.1.2. Mecanismos de produção de voz	21
2.1.3. Vibração das pregas vocais	22
2.1.4. Onda mucosa	24
2.1.5. Articulação	24
2.2. Fisiologia da laringe	27
2.2.1. Proteção	28
2.2.2. Adução laríngea	28
2.2.3. A elevação laríngea	29
2.2.4. Respiração	30
2.2.5. Fonação	30
2.2.6. Abdução e tensão da prega vocal	30
2.2.7. A teoria mioelástica da fonação	32
2.3. Examinando a laringe	33
2.3.1. Laringoscopia indireta	33
2.3.2. Endoscopia flexível e rígida	33
2.3.3. O exame estroboscópico	34
2.3.4. Laringoscopia direta	35
2.4. Patologias da laringe e vocais	36
3. Produção , Modelo, Fonética e Fonologia	44
3.1. Processo de produção de fala : teoria acústica	44
3.2. Modelo matemático do trato vocal	50
3.3. Fonética e Fonologia	52
3.3.1. Consideração de natureza linguística	52
3.3.2. Fonologia	53
3.3.3. Fonética	54
3.3.4. Fonética articulatória	54
3.3.4.1. Modo de articulação	54
3.3.4.1. Ponto de articulação	56
3.3.5. Fonética acústica	57
4. Medidas Acústicas	59
4.1. Análise acústica da voz	59
4.1.1. Análise do sinal sonoro	60
4.2. Medidas acústicas : visão geral	61
4.2.1. Medidas de perturbação	61
4.2.2. Aquisição do sinal de voz	61

4.2.3. Período e amplitude da frequência fundamental F_0	62
4.2.4. <i>Jitter</i> e <i>Shimmer</i>	64
4.2.5. Medidas de ruído espectral	71
4.2.6. Avaliação por resíduos	81
4.2.7. Medidas realizadas sobre o pulso glotal	83
Quociente de abertura e de fechamento	83
Quociente de velocidade	83
Índice de velocidade	84
Relação excitação glotal/ruído – GNE	84
Índice de ruído turbulento - TNI	85
4.2.8. Outras técnicas acústicas	86
4.3. Medidas acústicas : descrição	87
4.3.1. Aquisição do sinal de voz	87
4.3.2. Cálculos preliminares sobre o sinal de voz	91
4.3.3. Medidas de perturbação	92
Estimador síncrono para período e amplitude de F_0	93
PFF (<i>Perturbation Frequency Factor</i>)	96
DPF (<i>Direction Perturbation Factor</i>)	97
JR (<i>Jitter Ratio</i>)	97
JF (<i>Jitter Factor</i>)	98
Séries temporais	98
AVI/PVI (<i>Amplitude/Perturbation Variation Index</i>)	99
Médias Móveis	100
Perturbação de ordem superior	100
4.3.4. Medidas de ruído	104
HPFR	104
Medidas de energia espectral	105
Br	105
SNL	106
HNR	106
Hr	108
Ra	109
NNE	110
N/S	112
HNR Cepstral	113
Primeiro e segundo harmônicos	114
CPP	115
P/A	115
RPK	115
4.3.5. Medidas no sinal residual	116
Estimação paramétrica na análise de voz	116
Filtros adaptativos	120
Estimativa usando gradiente descendentes	120
Estimativa recursiva por mínimos quadráticos – RLS	122
Estimativa por filtro de Kalman	123
Considerações sobre convergência	125
4.3.6. Medidas do sinal residual	127

5. Teoria de Decisão e Resultados	131
5.1. Teoria de decisão	131
5.2. Resultados	135
5.2.1. Etapa univariada	135
5.2.2. Etapa multivariada	139
Combinação 2 a 2	139
Combinação 3 a 3	145
Combinação 4 a 4	145
5.3. Conclusões	146
5.4. Trabalhos futuros	147
6. Bibliografia	148

Capítulo 1

1.1. Caracterização do problema

Certamente é difícil encontrar alguém que nunca tenha tido algum problema de voz. A maioria das pessoas em algum momento ficou rouca durante uma infecção respiratória ou perdeu temporariamente a voz quando conversava (ou cantava) com amigos em um ambiente ruidoso. Na maioria dos casos este tipo de problema deve desaparecer com o alívio da infecção ou após um descanso da voz, respectivamente. No entanto, existem situações, onde um tipo de respiração e a voz fraca permanecem mesmo após uma doença ter sido curada. Além disso, algumas pessoas parecem inclinadas a perder repetidamente suas vozes ou ter outras doenças relacionadas.

A voz é resultado de fenômenos fisiológicos, que determinam sua emissão acústica. Basicamente os sons pode ser oriundo de três formas :

- ✓ Sons vocálicos, que são sons produzidos por pulsos de ar quase periódicos gerados na laringe pelo movimento cíclico das pregas vocais que excitam o trato vocal, exemplo, vogais /a/, /e/, /i/, /o/ e /u/.
- ✓ Sons fricativos, que são sons produzidos pela passagem turbulenta de ar através de alguma constrição formada no trato vocal, exemplo, /r/ e /s/.
- ✓ Sons plosivos, que são sons produzidos pelo fechamento completo do trato vocal com consequente aumento da pressão anterior à obstrução e liberação abrupta, exemplo, /d/, /t/, /p/ e /b/.

Dos três modos de geração dos sons citados, os sons vocálicos são os que apresentam vibração das pregas vocais além das ressonâncias do trato vocal completo. Dessa forma, visando a análise do trato vocal, esse modo de geração de voz contém toda informação fisiológica deste.

O fluxo de ar é modulado pelas vibrações das pregas, gerando pulsos de ar que emergem da laringe propagando-se pelas cavidades do trato vocal e irradiados pela boca e nariz, gerando um sinal acústico. Em havendo distúrbios fisiológicos na laringe, esse sinal acústico é afetado e constitui-se na fonte principal de informações a respeito desse distúrbio.

Os primeiros sintomas acústicos de patologia da laringe são as mudanças na frequência fundamental (F_0), intensidade e qualidade de voz. Esses sintomas são indicativos de possíveis disfunções orgânicas ou funcionais, onde a natureza desses sintomas é dependente da individualidade do paciente e do estágio da patologia.

A F_0 de um som vocálico é função da massa, elasticidade e comprimento das pregas vocais. Depende ainda da pressão sub-glótica e da configuração do trato vocal. Constatar que um paciente tem uma regulação de frequência adequada é feito na maioria pelo julgamento de um fonoaudiólogo, envolvendo a comparação com outros pacientes que apresentem tamanho, sexo e idade similares, ou seja, depende da experiência adquirida pelo especialista.

A frequência fundamental corresponde ao número de ciclos com que a mucosa das pregas vocais vibra em um segundo. Na voz falada, oscila entre 100 Hz a 150 Hz no sexo masculino e entre 200 Hz a 300 Hz no feminino (NISHIZAWA et al: 1988).

A visualização dos rápidos ciclos vocais não é possível a olho nu ou por meio de câmeras convencionais, as quais fornecem de 30 a 60 imagens por segundo. Exige a utilização de equipamento específico e de alto custo, como a fotografia e a filmagem ultra-rápida que geram cerca de 5.000 imagens por segundo. O recurso da estroboscopia, o método mais utilizado na prática clínica, somente possibilita a observação de amostras de vários ciclos compostos, em um efeito de câmara lenta. O videoquimógrafo é um equipamento simples e efetivo, capaz de fornecer 7.812,5 imagens por segundo, embora somente de uma das linhas da imagem convencional (SVEC; SCHUTTE, 1996). Este possibilita reconstruir todos os movimentos ocorridos na linha selecionada sequencialmente, de tal forma que a imagem final mostra o movimento da onda mucosa (bordo superior e inferior) em uma sucessão de ciclos. Oferece avaliação objetiva permitindo mensurar as fases aberta, fechada, de abertura e de fechamento do ciclo vibratório das pregas vocais, além da frequência fundamental e amplitude de vibração.

A diminuição da frequência fundamental (agravamento) da voz em pacientes do sexo feminino pode ocorrer por inúmeras causas, como por exemplo o uso crônico de esteróides por pacientes portadores de doenças reumatológicas e endócrinas, e na síndrome adrenogenital.

Outra doença que também pode levar à diminuição da frequência fundamental é a paralisia do nervo laríngeo superior, cujo ramo externo enerva o músculo cricotireóideo, principal tensor das pregas vocais. Sua contração aproxima anteriormente as cartilagens tireóide e cricóide, afastando as aritenóides da cartilagem tireóide e distendendo as pregas vocais (ISSHIKI, 1989). Esse mecanismo é também muito utilizado no canto, especialmente na emissão da voz aguda (falsete).

Com o advento das fonocirurgias e a melhor compreensão da fisiologia da emissão vocal, têm sido propostos procedimentos específicos para a modulação da voz. Dentre os tratamentos cirúrgicos que objetivam a elevação da frequência fundamental, podemos destacar a aproximação cricotireóide (ISSHIKI et al., 1974,1983, ISSHIKI, 1977, 1980) que simula a contração do músculo cricotireóideo através de suturas. Essa técnica é fisiológica, reversível, preserva o músculo cricotireóideo e não manipula a comissura anterior .

Os efeitos desse procedimento em relação à vibração vocal ainda estão pouco estabelecidos. Estudos clínicos sofrem influência dos fenômenos compensatórios fisiológicos da laringe, os quais podem falsear o real efeito biomecânico da aproximação cricotireóidea. De outro modo, o estudo experimental em laringes humanas excitadas permite avaliar adequada e isoladamente os efeitos desse procedimento sobre a vibração vocal.

A análise da videoendoscopia (Hirano, 1993) é a principal técnica para a avaliação de distúrbios de voz. Nessa técnica, uma pequena câmera é acoplada a um telescópio oral (ou fibroscópio nasal), fornecendo imagens das pregas vocais e outras estruturas da laringe, sendo que a maioria das lesões nas pregas vocais podem ser identificadas nessas imagens. Combinando com técnicas estroboscópicas (videoestroboscopia), uma sequência de movimentos lentos fornece informações adicionais sobre o sutil movimento vibratório das pregas vocais. A videoestroboscopia é particularmente importante permitindo a detecção prematura de lesões malignas que, embora pequenas no estágio inicial, podem causar prejuízos sérios a vibração das pregas vocais.

As técnicas acústicas para análise da voz baseiam-se na fonte de sinal e nos métodos de análise. Uma fonte é o sinal direto, a pressão de som irradiada pela boca ou o sinal de contato com a garganta. Essa pressão é o sinal mais imediato para a análise acústica, porém sua utilidade para avaliar as condições da laringe

é limitada. A produção de uma vogal sustentada é realizada pela fonte glotal, que será afetada pela patologia da laringe, e pela estrutura supra-glotal, cujas características ressonantes possivelmente não serão afetadas pela patologia. As possíveis mudanças no pulso glotal devido a patologia alterariam o formato de onda da pressão irradiada, podendo atenuar ou reforçar componentes ou até mesmo eliminá-los.

Outra fonte é o sinal indireto, o sinal glotal ou residual são obtidos por técnicas de filtragem inversa. O som vocálico é modelado como uma fonte (glote) excitando um tubo acústico. A filtragem inversa retira os efeitos do tubo e o sinal resultante aproxima-se do sinal periódico da fonte. Caso a estrutura supra-glotal não é afetada pela patologia e a causa da mudança da voz esta na fonte, então a aproximação do sinal periódico da fonte contém informação suficiente para se analisar os efeitos acústicos da patologia.

O uso da instrumentação eletrônica na análise objetiva quantificar e fornecer medidas ou parâmetros acuradas das características do sinal de voz. Para cada uma dessas fontes, os sinais podem ser analisados no domínios do tempo ou da frequência, dependendo das vantagens e limitações que apresentarem.

• Medidas acústicos no domínio do tempo

A vibração das pregas vocais produz um pulso de ar glotal quase periódico. A utilização de medidas de perturbação do período de “pitch” do sinal acústico discriminam casos patológicos de normais. Baseado nesse fato, várias medidas na perturbação de “pitch” e amplitude são propostas na literatura. A primeira foi o fator de perturbação de frequência (PFF) determinado por Lieberman [Lieberman, 61], sendo definido como a frequência relativa das perturbações de período de “pitch” maiores que 0,5 mS ocorrendo em uma vogal sustentada. É definida portanto, como a diferença de tempo entre as durações de períodos sucessivos no sinal de voz.

Uma segunda medida de perturbação é a perturbação média relativa (RAP), proposto por Koike [Koike, 77] e analogamente ao PFF mede a flutuação dos períodos de pitch (usando janela de três amostras).

Utilizando a frequência para substituir o período e variando a janela de dados significativos, pesquisadores propuseram diversas medidas de

perturbação. Dessa forma, tem-se na literatura o quociente de perturbação de frequência (FPQ), o quociente de perturbação de período de “pitch” (PPQ), razão de *jitter* (JR) que fornece a relação entre a média das perturbações pelo período médio, proposto para destacar a redução do *Jitter* para frequências fundamentais relativamente altas. Além da razão de *Jitter*, existe também o fator de *Jitter*, análogo ao JR, porém definido no domínio da frequência. A nomenclatura atual engloba todos esses fatores de perturbação de período (frequência) numa única denominação : *Jitter*.

Por outro lado, quando a perturbação é calculada sobre a amplitude do sinal, a maioria das medidas descritas acima têm seu correspondente e a denominação é *Shimmer*. Uma outra característica do sinal a ser considerada além das perturbações de período e amplitude é a forma de onda do sinal. A relação sinal ruído fornece informações para a discriminação de vozes patológicas e normais.

• Medidas acústicos no domínio da frequência

A análise no domínio da frequência possibilita um outro conjunto de medidas acústicos da voz. O espectrógrafo é um instrumento utilizado na análise de frequência da voz, o qual verifica a distribuição espectral de energia de um determinado trecho da voz, sustentado por um banco de filtros. Como resultado dessa análise tem-se um gráfico tempo x frequência, sendo a densidade de energia indicada pela intensidade. Além disso, a frequência fundamental e os formantes das vogais sustentadas são visualizados no espectrograma.

Diversos estudos comprovam que existem diferenças significativas entre espectros de vozes normais e patológicas. Isso pelo fato do fluxo de ar turbulento originário do fechamento incompleto da glote ou por vibração “não balanceada” das pregas vocais, as componentes espectrais do ruído podem ser originadas.

Essas componentes são distribuídas ao longo do espectro de forma variada e a extensão da distribuição depende da gravidade da patologia. Além disso, a presença do ruído contribui para a rouquidão, que é o primeiro sintoma para inúmeras patologias.

Proposto por Shoji em 1992 a relação de potências em alta frequência (HFPR) calcula a média do logaritmo da densidade espectral de potência em intervalos de frequência bem definidos. Algumas outras medidas definidas no domínio da frequência são : relação harmônicas de ruído (HNR) ; intensidade relativa de harmônicas (Hr) determinado somando-se todas as intensidades das harmônicas (da segunda harmônica em diante) normalizadas pela intensidade total do espectro de voz (fundamental, harmônicas e ruído), sendo que o limite estabelecido para pessoas normais é de 67,5%; harmônicas relativas (Ra) que é a relação entre as componentes harmônicas e não harmônicas obtidas em um segmento de três períodos; relação de soproidade (Br) que estabelece a razão entre a energia em altas frequências e a energia total do sinal; nível espectral de ruído (SNL) que analisa as intensidades espectrais em intervalos de frequência e por último a energia de ruído normalizada (NNE) que considera o sinal composto de componentes periódicas mais ruído aditivo.

1.2. Revisão da literatura

Diversas técnicas alternativas de auxílio ao diagnóstico têm sido propostas utilizando somente a voz do paciente. Foram propostos diversos índices quantitativos para diagnosticar e acompanhar a evolução de tratamentos de doenças da laringe (BAKEN et al, 1987). Vários métodos de diagnóstico automático utilizando a voz do paciente têm sido pesquisados, tais como filtragem inversa (CHILDERS et al, 1992), redes neurais (MARTINEZ et al, 2000), (RITCHINGS et al, 2001), (JO et al, 2001), medidas acústicas da voz (FROHLICH et al, 1998), entre outras.

Vários métodos não invasivos, como o acústico, tem sido aplicados para avaliação clínica de distúrbios de voz, na qual tem o potencial de fornecer medidas objetivas, ou seja, valores numéricos ou computadorizados, associados com a produção de voz. Essas medidas podem oferecer suporte quantitativo para decisões subjetivas da videoestroboscopia. Os efeitos de uma intervenção cirúrgica ou de um curso de terapia de voz podem ser avaliadas comparando as medidas do pré e do pós tratamentos.

Nos últimos anos muitas fontes de perturbações que afetam os sinais acústicos tem sido investigadas sistematicamente. Esses artefatos incluem efeitos de microfone (Titze, 1993), distorções na gravação e taxa de amostragem

insuficiente (Titze, 1987). É sabido que distúrbios de voz deveriam ser adquiridos em uma sala anti-ruído, utilizando um microfone com condensador acoplado posicionado lateralmente e a poucos centímetros da boca do paciente.

No trabalho *Objective Evaluation of Vocal Pathology using Voice Spectrography* [Rontal, 75], os autores apresentaram o uso do espectrograma para a avaliação da situação clínica da reabilitação vocal e do tratamento cirúrgico e médico de uma variedade de doenças das cordas vocais. Trabalhos anteriores discutiram o uso do espectrograma no estudo de doenças da laringe, concluindo que os padrões produzidos podem ser aplicados a diversas situações clínicas.

Este trabalho apresenta diferentes espectrogramas de vozes normais e com patologias, analisando-as e apontando diferenças. Uma das conclusões mais importantes deste artigo é a respeito do espectrograma da voz. Primeiramente, segundo os autores, não é possível se obter uma medida exata ou quantificar um espectrograma da voz. Mudanças nos padrões do 5 espectrograma da voz são percebidas qualitativamente. Os espectrogramas não apresentam diferenças entre tipos de lesões de cordas vocais, pois muitas patologias apresentam características acústicas similares. Os autores concluem que uma avaliação final das condições das cordas vocais deve ser feita por todos os meios possíveis, e não baseado apenas em um único método.

No trabalho *Automatic Classification of Laryngeal Dysfunction using the Roots of the Digital Inverse Filter* [Deller, 80], os autores apresentaram uma técnica automática de classificação de disfunção da laringe, utilizando os pólos e zeros de um filtro inverso (FI) digital. A informação para a classificação está contida na posição dos zeros do polinômio do filtro inverso. Um vetor padrão de características é processado por um procedimento de agrupamento (*clustering*) automático para classificar casos. Segundo os autores, dois procedimentos de análise são possíveis na população de sinais de voz. Primeiro a classificação entre casos, auxiliando no diagnóstico, e segundo, a avaliação dentro de um mesmo caso, onde se tem um auxílio na avaliação da evolução da terapia.

Nas simulações, foram utilizados dois tipos de sinais sintéticos. Os primeiros foram subdivididos em duas categorias de anomalia:

1. Casos de performance de filtro (como *Jitter*, *Shimmer*, entre outros);

2. Casos de inserção de ruído nos sinais.

Em geral, os resultados mostraram que a análise automática dos padrões pode ser um discriminante efetivo de diferentes aberrações acústicas. Os autores concluíram que este método demonstrou significativa habilidade para identificar simulações de anomalias na laringe. Porém, eles ressaltaram que, como foram utilizados sinais sintéticos e devido à enorme variabilidade inerente a sinais patológicos reais, pesquisas adicionais são necessárias para aplicar esta técnica clinicamente.

No trabalho *A method of spectra-analysis for abnormal voice signals* [Guoxin, 89], os autores utilizaram um sistema de análise espectral de sinais de voz de pacientes com doenças na laringe. Estudos mostraram que quando alguma desordem ocorre na laringe, a voz do paciente muda sua forma de onda e seu espectro de potência de acordo. Os resultados deste artigo mostraram que este método é capaz de reconhecer características acústicas de vozes anormais. Os autores concluíram que este método de análise-espectral é um método efetivo para auxiliar médicos na pesquisa da fisiologia e da patologia da voz, e também para monitorar o processo do tratamento.

No trabalho *Detection of Laryngeal Function Using Speech and Electrolottographic Data* [Childers, 92], os autores desenvolveram uma medida quantitativa para a avaliação da função da laringe, usando a fala e dados de eletrolotografias (EGG). Foram desenvolvidos dois procedimentos para a detecção de patologias da laringe:

1. Medida de distorção espectral usando métodos de *Pitch* com *Linear Predicting Code* (LPC) e quantização vetorial;
2. Análise do sinal de eletrolotografia (EGG), medindo a diferença do intervalo de tempo e amplitude.

Os autores utilizaram vinte e três pacientes com patologias e cinquenta e duas vozes normais. A vogal utilizada foi o /i/ sustentado. O melhor resultado adquirido com o primeiro método foi de 75,9% (22 corretos/29 sujeitos) de detecção correta de patologias. Já o segundo método apresentou 69% de detecção correta (20 corretos/29 sujeitos). A probabilidade de falsos positivos foi de 9,6% para sujeitos normais.

No trabalho *Evaluation of Neural Classifiers using Statistic Methods for identification of Laryngeal Pathologies* [Rosa, 98], os autores propuseram um

método quantitativo de classificação de patologias da voz, de forma a padronizar o diagnóstico entre especialistas da fala. O objetivo deste trabalho é utilizar redes neurais para classificar padrões de patologias através de medidas acústicas obtidas do resíduo do sinal (filtragem inversa). Foram utilizados 119 falantes entre normais e patológicos. Os vinte e dois tipos de patologias foram reagrupados para produzir apenas seis grupos baseados na acústica e nas similaridades patológicas da voz. Este procedimento foi justificado em experimentos prévios pelos autores, onde utilizaram todas as disfonias e obtiveram uma grande variabilidade nos resultados através da rede neural. Foram aplicados métodos estatísticos nas medidas acústicas, e estes foram os vetores de entrada da rede neural para identificar doenças da laringe. Os autores testaram diversas combinações entre as medidas acústicas e os métodos estatísticos. A média de acerto de classificação dos grupos de patologias foi de 62,33%.

No trabalho *An algorithm for the automatic differentiation between the speech of normals and patients with Friedreich's ataxia based on the shorttime fractal dimension* [Accardo, 98], os autores desenvolveram um algoritmo baseado nas técnicas de combinação de padrões acústicos, para obter um discriminante automático de vozes normais e com a patologia ataxia de Friedreich. Ataxia de Friedreich é uma doença degenerativa do sistema nervoso central, que é caracterizada pela desordem dos movimentos da articulação, hipotonia e tremor. A técnica proposta utiliza medidas fractais de dimensão de tempo-curto. Para comparação, os autores utilizaram as medidas de **cruzamentos por zero** (*zero-crossing*) e taxa de energia (*energy rate*), que são estimados na tarefa de classificação. Foram utilizados oito pacientes com ataxia de Friedreich e nove pessoas sem patologias na laringe ou disfunções neurológicas. Os sujeitos falaram três frases diferentes.

Os resultados mostraram que todas as medidas são capazes de diferenciar os dois grupos, porém o parâmetro de dimensão fractal fornece uma classificação com maior índice de acerto. O valor médio de classificação para os medidas fractais, taxa de energia e cruzamento por zero são, respectivamente, 96,1%, 92,1% e 94,1%. Com isso, os autores concluem que o método pode ser aplicado para classificação automática de vozes, normais *versus* patológicos.

No trabalho *Acoustic 'breathiness measures' in the description of pathologic voices* [Frohlich, 98], os autores propuseram uma medida acústica para soprosidade, com o objetivo de fazer a distinção entre diferentes condições de fonações fisiológicas em casos de vozes patológicas. O índice proposto é o *glotal to noise excitation ratio* (GNE), e este é comparado a outros sete índices descritos na literatura. A avaliação deste método é feita com base na capacidade de distinguir entre diferentes mecanismos bem definidos de fonação patológica. Para os testes, foi utilizado um banco de 454 vozes, entre patológicas e normais, como grupo de referência, e seis vogais diferentes isoladas.

Os resultados mostraram que o GNE, entre todas as medidas feitas, apresentou a mais significativa distinção entre os diferentes grupos de mecanismos de fonação patológica.

No trabalho *Signal processing and statistical procedures to identify laryngeal pathologies* [Rosa, 99], os autores utilizaram técnicas de processamento de sinais para extrair características acústicas da voz, e estas foram utilizadas em redes neurais para determinar as condições glotais relacionadas à patologia da laringe. Diversos índices acústicos foram calculados, como *Jitter* e *Shimmer*. Foram combinadas vinte e duas patologias duas a duas, e um segundo indicador de habilidade de identificação foi o nível de separação entre os grupos (*clusters*) formados entre os pares de patologias.

Após as transformações, as medidas acústicas foram fornecidas à entrada da rede neural. Cada neurônio de saída representa uma condição das cordas vocais (agrupadas em 6 grupos diferentes). A rede neural foi treinada minimizando a entropia relativa do conjunto de treinamento. Nos resultados experimentais, diversas combinações de índices produziram diferentes percentuais de acerto ao processo de classificação.

Algumas conclusões dos autores são: medidas de *Jitter* e *Shimmer* para calcular movimentos anormais das cordas vocais produzem identificações mais confiáveis nas redes neurais; O melhor resultado obtido foi de 82,22% de identificação das condições 8 da glote; e por fim, a discriminação entre vozes patológicas e normais foi alcançada, conforme as próprias palavras dos autores “perto de 100%”.

No trabalho *Laryngeal pathology detection by means of class-specific neural maps* [Hadjitodorov, 00], os autores propuseram um método de

classificação de vozes normais e de vozes patológicas com doenças na laringe. O método proposto para aumentar a precisão de detecção de patologia da laringe é baseado na modelagem da função densidade de probabilidade (fdp) dos falantes normais e patológicos por meio de dois protótipos de mapas de distribuição (PDM – *Prototype Distribution Map*), desenvolvido pelos autores. A fdp dos vetores de entrada de vozes desconhecidas como patológicas ou normais, também é modelada pelo PDM. Para os testes foram utilizadas 400 vozes, 100 normais e 300 pacientes com patologias na laringe. Os falantes pronunciaram a vogal sustentada /a/ para ser analisada. Para comparação, os autores utilizaram métodos clássicos de reconhecimento, como análise discriminante linear (LDA), o método *k-nearest neighbour* (k-NN) e o *self-organizing map* (SOM). A precisão na classificação do método proposto foi de 95,1%, enquanto que para LDA foi de 69,95%, para o k-NN foi de 90,95% e para o SOM 90,46%.

No trabalho *Acoustic Analysis of speech for detection of Laryngeal Pathologies* [Martinez, 00], os autores implementaram um sistema automático de detecção de patologias da laringe utilizando a análise acústica da fala no domínio frequência.

Diferentes técnicas de processamento foram aplicadas: cepstrum, mel-cepstrum, delta cepstrum, delta mel-cepstrum e FFT. Os sinais processados foram utilizados como parâmetro de entrada de uma rede neural. Eles afirmaram que a rede neural é um excelente sistema de classificação e especialista em dados com ruído, incompletos, sobrepostos, etc... Neste trabalho, o sistema é treinado para distinguir entre vozes normais e patológicas (duas classes), e para classificar vozes normais, diplofonia e rouca (três classes). Foi utilizada a vogal sustentada /a/ e as vozes patológicas foram diagnosticadas previamente (vinte e quatro amostras de patologias). Para a extração de características, uma janela *hamming* foi aplicada a cada segmento e então os padrões foram extraídos dos dezesseis primeiros coeficientes cepstrais. Os outros processamentos foram cepstrum em escala mel com dezesseis coeficientes, delta cepstrum e delta mel-cepstrum com trinta e dois coeficientes, e a FFT com cento e vinte oito coeficientes. Os resultados experimentais mostraram que a análise cepstral obteve a melhor performance (91%) para separar as duas classes e 81,91% para as três classes. A técnica que teve o pior desempenho foi a FFT, segundo os autores, devido à

dimensão 9 dos padrões que dificultou o treinamento da rede. Os autores concluem que é possível utilizar esta ferramenta para a análise quantitativa no processo de diagnóstico, assim como acompanhar a evolução da terapia de reabilitação.

No trabalho *Adaptive estimation of residue signal for voice pathology diagnosis* [Rosa, 00], os autores propuseram uma técnica não invasiva para auxiliar especialistas da voz a fornecer um diagnóstico preciso de doenças da laringe e das pregas vocais. A técnica utilizada é a filtragem inversa do sinal de voz, utilizando filtros Kalman e Wiener. Os autores utilizaram três vogais /a/, /e/ e /i/. A frequência de amostragem do sinal foi de 22050 Hz e o nível de quantização de 16 bits. A técnica de filtragem inversa descrita na literatura, como antes visto, é desenvolvida no segmento mais estacionário do sinal de fala. Porém, muitas patologias não possuem características de ‘quasi-estacionariedade’. Para superar este problema, os autores propuseram três algoritmos adaptativos para estimação do modelo autoregressivo (AR). Foram utilizadas sete medidas acústicas para diferenciar entre falantes normais e patológicos. Nos resultados, as vogais /a/ e /e/ foram as que obtiveram os melhores resultados. Das diversas conclusões relatadas pelos autores, uma delas afirma que a combinação destes filtros permite uma ótima discriminação de doenças com características similares e que a melhor característica (medida acústica) para a discriminação patológica foi *Jitter* com 54,79%.

No trabalho, *Pathological voice quality assessment using artificial neural networks* [Ritchings, 01], os autores descreveram um protótipo de um sistema para a avaliação objetiva da qualidade da voz em pacientes em recuperação de vários estágios de câncer da laringe. Foi utilizada a vogal sustentada /i/, e esta foi avaliada subjetivamente por especialistas da fala e da língua de acordo com sete níveis de qualidade (0 = menos anormal e 6 = mais anormal). O sistema extrai os principais parâmetros, como a impedância dos sinais (Eletroglotografia, EEG) e estes são usados para treinar e testar as redes neurais (ANN). O sistema utilizado para capturar os sinais da impedância elétrica consiste de dois eletrodos, um em cada lado do pescoço. Setenta e sete vozes anormais foram utilizadas para treinar e testar uma rede neural do tipo *Multilayer Perceptron* (MLP) com sete saídas. O algoritmo de treinamento utilizado foi o *backpropagation*.

A função de erro utilizada foi a entropia cruzada. Os melhores resultados foram obtidos mediante uma combinação de medidas acústicas do sinal, alcançando um índice de acerto de 92%. Os autores concluem que um sistema de rede neural pode ser treinado para fornecer uma classificação objetiva da qualidade da voz.

No trabalho, *Screening of Pathological Voice from ARS using Neural Networks* [Jo, 01], os autores descreveram um método de classificação de vozes patológicas ou normais baseados em um sistema de resposta automática (ARS – *Automatic Response System*). As vozes são classificadas utilizando uma rede neural, cujos parâmetros de entrada são características acústicas como *Jitter*, *Shimmer*, razão harmônicas/ruído (NHR), entre outros. O objetivo principal deste trabalho é poder classificar vozes, patológicas ou não, usando rede pública de telefone, de forma que mais doutores não especialistas no assunto e pacientes possam diagnosticar suas vozes sem ir ao hospital (a idéia não é emitir o diagnóstico exato nem o tratamento). A taxa de classificação é comparada a taxa de classificação obtida utilizando vozes que foram gravadas diretamente em *Digital Audio Tape* (DAT).

Foram utilizados setenta e seis casos de patologias e trinta e cinco casos de vozes normais para treinar a rede neural. As taxas de classificação do ARS ficaram entre 88% e 91%. Os autores concluem que não foi observado diferença significativa entre ARS e DAT em termos de taxa de classificação. Baseado nisso, é possível separar patológicos de normais com o sistema ARS e pode ser aplicado a um sistema automático de diagnóstico. Porém, o número total de vozes não é ainda suficiente para generalizar o desempenho, necessitando mais dados para isso.

No trabalho, *Acoustic analysis of voice in children with noduli vocales*, *International Journal of Pediatric Otorhinolaryngology* [Niedzielska, 01], os autores utilizaram testes acústicos em sinais de voz de crianças (4-14 anos) com nódulos vocais. Estes testes foram feitos antes e depois do tratamento. As medidas acústicas da voz foram comparados com um grupo controle de crianças sem patologia. Os resultados da investigação foram analisados acusticamente. Foi mostrado que os valores de *Jitter*, *Shimmer*, frequência fundamental e a relação sinal-ruído mudaram significativamente entre as crianças com nódulos vocais e sem patologias na laringe. Estas medidas tendem a se aproximar dos

valores normais durante o tratamento. A conclusão dos autores é que o método acima permite a detecção de desordens do funcionamento da laringe, podendo ser útil no processo de avaliação do tratamento.

1.3. Motivação e objetivos

Diante da importância dos procedimentos que modulam a voz e da necessidade de se entenderem as alterações na vibração das pregas vocais em ambos os sexos, este trabalho tem o intuito de prover auxílio aos especialistas na prevenção, diagnóstico e tratamento de alterações vocais e laríngeas. Indiretamente, através da análise da fala pode-se identificar as disfonias, uma vez que esta informação está codificada na voz. Sendo um método não-invasivo, pode ser realizado em qualquer pessoa, inclusive eliminando o uso de equipamentos e outros que incomodariam o paciente.

Dada a natureza numérica da análise acústica, os diagnósticos são quantitativos e portanto objetivos. Dessa forma, acrescentam subsídios ao especialista de voz que favorecem um diagnóstico adequado. Além disto, elimina dúvidas eventuais que exames por laringoscopia indireta possam apresentar.

Dessa forma a evolução do tratamento pode ser avaliada, pois as medidas acústicas (extraídas do sinal de voz) indicam precisamente o nível de melhora do paciente, permitindo que o especialista mantenha ou altere a medicação ou a forma de conduzir o tratamento.

Nos estudos que empregam medidas acústicas, verifica-se que a grande maioria busca relacionar estes valores obtidos do sinal de voz com características perceptuais da voz. Como a classificação das vozes analisadas é realizada em termos de graduação de rouquidão, aspereza e soproidade, existem naturais diferenças entre as avaliações dos especialistas de voz que classificam estas vozes para os estudos. Outra dificuldade com este método de utilizar as medidas acústicas para avaliação automática da condição da laringe é que existe uma natural correlação entre estes índices perceptuais, acarretando na dificuldade de diferenciação entre vozes “praticamente normais” daquelas que apresentam “leves” níveis de alteração.

Considerando as dificuldades apresentadas no diagnóstico diferencial das patologias da laringe e do trato vocal, este trabalho busca relacionar as medidas

acústicas extraídas do sinal de voz diretamente com as patologias, ou seja, possibilitar uma identificação automática destas apenas com o uso do sinal de voz digitalizado.

Para todas as medidas acústicas, utilizou-se a teoria bayesiana para avaliar a capacidade discriminatória destas, para selecionar as que são mais eficientes individualmente e combinadas. São confrontadas as patologias para verificar a existência de separabilidade estatística entre elas, considerando cada uma das medidas acústicas analisados.

O objetivo deste trabalho é o de desenvolver um sistema que possibilite a identificação automática de patologias do trato vocal, em particular da laringe, utilizando-se apenas o sinal de voz digitalizado. Os benefícios que podem advir desse sistema são:

1. Utilização de sinais digitais e modelos matemáticos, tem-se baixo custo;
2. Facilidade de uso, pois a partir do sinal digitalizado tem-se como resultado, a identificação probabilística das patologias da laringe;
3. Possibilidade de um pré-diagnóstico das patologias da laringe em estágios que não exigem intervenção cirúrgica;
4. Acompanhamento de tratamentos clínicos;
5. Gerar uma base de dados clínica.

Estes benefícios motivam o estudo da voz com objetivo de discriminar as patologias da laringe permitindo que o especialista de voz (otorrinolaringologista ou fonoaudiólogo) tenha uma ferramenta útil auxiliando o diagnóstico das patologias da laringe.

Esse trabalho tem a colaboração da Faculdade de Medicina da USP de Ribeirão Preto. Existe a preocupação de conciliar interesses de pesquisa com interesses clínicos, como descritos nos principais objetivos do estudo, ou seja:

- Organizar de um banco de dados base de patologias da laringe ;
- Aplicação da teoria de decisão, no caso a teoria Bayesiana ;
- Investigação do comportamento e importância de cada medida acústica na discriminação das patologias da laringe ;
- Desenvolvimento de um software de caráter social, que possa ser utilizado como triagem na área de otorrinolaringologia, visando prevenir, diagnosticar e tratar alterações vocais e laríngeas, facilitando e priorizando o atendimento.

1.4. Descrição técnica

O banco de sinais de voz digitais é de aproximadamente 500 sinais, este é composto por vozes normais e patológicas (diagnosticadas por médicos) e um outro banco de voz com aproximadamente 100 sinais de vozes (sem diagnóstico). A partir das medidas acústicas (38 medidas acústicas, explanadas em capítulos posteriores) extraídas desses sinais, tem-se informações numéricas dos sinais de voz e com isso utilizando-se da teoria de decisão, mais precisamente da teoria Bayesiana, realizamos um processo de discriminação de 4 patologias da laringe (Edema, Fenda, Nódulo e Pólipo) mais a voz normal.

O procedimento ocorreu da seguinte forma: as 38 medidas acústicas foram testadas univariadamente, ou seja, individualmente obtendo-se o percentual discriminatório de cada medida acústica com relação as 4 patologias mais a voz normal. Dessa forma, obtém-se uma pós-classificação desses sinais, que é então comparada com o pré-diagnóstico (estabelecido pelo médico) fornecendo assim as porcentagens de acerto de cada medida acústica na discriminação de cada uma das 4 patologias mais a voz normal em análise.

Com isso, pode-se avaliar o comportamento de cada medida acústica, ou seja, sua importância para a discriminação das 4 patologias mais a voz normal. E numa Segunda etapa as 38 medidas acústicas foram testadas multivariadamente, ou seja, as medidas serão combinadas 2 a 2, 3 a 3, 4 a 4, ..., 38 a 38, onde chegamos até 4a4 porque alcançamos 100% de acerto discriminatório para as 4 patologias mais a voz normal e dessa forma, entendemos que não seria necessário continuar as combinações porque os resultados são suficientes para a defesa, mais estaríamos continuando essas combinações posteriormente.

Assim, a partir da combinação estabelecida, um procedimento é inicializado gerando primeiro um processo de treinamento, ou seja, são obtidas médias e matrizes de covariância e depois um segundo processo inicia-se obtendo através da teoria Bayesiana (teoria de decisão), mais precisamente da máxima verossimilhança (explicitada no capítulo 5), a discriminação dentre as 4 patologias mais a voz normal em análise alcançada por cada combinação.

O estudo utilizando mais de uma variável permite o uso correlacionado das mesmas visando inferir, com maior precisão a discriminação das 4 patologias da laringe mais a voz normal em análise.

No capítulo 2 são relatadas a anatomia e fisiologia da laringe, além da apresentação de algumas patologias da laringe e da voz. No capítulo 3 encontram-se os fundamentos do processo de produção de fala culminando no modelo matemático do trato vocal, e algumas considerações linguísticas através da fonética e fonologia.

No capítulo 4 são detalhadas as medidas acústicas, que foram implementados por Rosa (1998).

Finalmente no capítulo 5 aborda-se a teoria de decisão, no caso a teoria Bayesiana utilizada no processo discriminatório das 4 patologias da laringe mais a voz normal e ainda os resultados alcançados levando-se em consideração as etapas univariada e multivariada, além das conclusões e propostas de trabalhos futuros.

Capítulo 2

Anatomia, Fisiologia e Patologias Laríngeas e Vocais

2.1. Anatomia básica da Laringe

A laringe é um órgão localizado na garganta que atua com um papel crucial no ato de falar e respirar. A laringe é o ponto no qual o trato aero-digestivo separa-se em dois caminhos : o curso de ar inspirado através da traquéia, nos pulmões, e no alimento engolido que entra no esôfago passando para o estômago.

Devido a sua localização, a laringe tem 3 importantes funções:

- controle do fluxo de ar durante a respiração ;
- proteção do corredor do ar e
- produção do som para a fala.

Figura 2.1. Estrutura da Laringe

A laringe consiste de uma estrutura de cartilagem com tecido mole envolvente. A parte mais proeminente da laringe é a cartilagem tireóide cuja

porção anterior da cartilagem tiróide pode ser facilmente apalpada na face anterior da garganta, área conhecida como “pomo de adão”.

Situado superior a laringe (considerado algumas vezes parte da própria laringe) localiza-se um osso em forma de U chamado hióide. O osso hióide é unido superiormente a mandíbula e inferiormente à laringe por meio de músculos e tendões. Essas uniões são importantes na elevação da laringe durante a deglutição e fala.

Na parte mais inferior da laringe esta localizada a cartilagem cricóide, a qual tem forma de um anel com sua parte maior posicionada na porção posterior da laringe. Abaixo da cricóide estão os anéis da traquéia. Vide figura 2.1.

No centro da laringe estão as pregas vocais. As pregas vocais são uma das partes mais importantes da laringe, porque atuam como papel chave em todas as funções vitais e de fonação.

As pregas vocais são feitas de músculos cobertos por uma camada fina chamada mucosa. Há uma prega direita e uma esquerda, dando forma a um "V" quando visto de cima (figuras 2.2 e 2.3). Na porção posterior de cada prega vocal existe uma pequena estrutura feita da cartilagem chamada aritenóide. Muitos músculos pequenos, descritos abaixo, são unidos pelos aritenóides. Estes músculos puxam as aritenóides distantes uma da outra durante a respiração, abrindo desse modo o caminho do ar. Durante a fala as aritenóides e conseqüentemente as pregas vocais são unidas (figuras 2.2 e 2.3).

Enquanto o ar passa pelas pregas vocais nesta posição, elas abrem e fecham muito rapidamente. A pulsação rápida do ar que passa através das pregas vocais produz um som que é então modificado pelo restante do trato vocal para produzir a fala.

Figura 2.2. À esquerda mostra as pregas na posição aberta. As pregas devem abrir como esta durante respirar. À direita, as pregas são mostradas na posição fechada como durante a fala.

Figura 2.3. Imagens laringoscópicas da laringe. A. Glote na posição de repouso; B. Glote durante a atividade. 1. Glote; 2. Pregas vocais; 3. Epiglote; 4. Comissura anterior; 5. Cartilagens aritenóides e 6. Comissura posterior.

2.1.1. Músculos da laringe

O movimento da laringe é controlado por dois grupos de músculos. Os músculos que movem as pregas vocais e os músculos localizados dentro da laringe, denominados músculos intrínsecos.

Os músculos intrínsecos são mostrados na figura 2.4 abaixo. As pregas vocais são abduzidas primeiramente por um par de músculos, cuja localização estende-se da porção posterior da cartilagem cricóide até a cartilagem do aritenóide. Este músculo é chamado posterior cricoaritenóideo (quase todos os músculos na garganta são nomeados indicando começo e fim).

Diversos músculos ajudam a aduzir e enrijecer (tensão) as pregas vocais. O próprio corpo das pregas vocais é composto de um músculo chamado tiroaritenóideo. Um músculo chamado interaritenóideo percorre de um aritenóide a outro e junta essas duas partes de cartilagem. O músculo cricoaritenóideo lateral, assim como o músculo posterior cricoaritenóideo, funciona também do aritenóide até a cartilagem cricóide.

O músculo cricotiróide funciona da cartilagem cricóide até a cartilagem da tiróide. Quando se contrai, a cartilagem da tiróide inclina-se para a frente,

aplicando uma tensão sobre as pregas vocais e desse modo aumentando o *pitch* da voz.

Os músculos extrínsecos são chamados também de músculos de cinta (pela semelhança com cintas). Não afetam o movimento das pregas vocais, mas levantam e abaixam a laringe inteira. Este movimento é especialmente importante para a deglutição. Os cantores treinados, por exemplo, desenvolvem

também o controle apurado destes músculos para ajudar melhorar a qualidade de suas vozes.

Figura 2.4. Localização de músculos na Laringe.

2.1.2. Mecanismos de produção de voz : fluxo de ar e vibração das pregas vocais.

A produção da voz é dependente do fluxo e pressão do ar através das pregas vocais. A figura 2.5 mostra uma seção cruzada vertical através da laringe. Abaixo das pregas vocais está a traquéia, que conduz aos pulmões. As pregas vocais são realmente pregas do tecido, descritas mais detalhadamente abaixo. Nesta figura as pregas vocais são separadas, mostrando como seria durante respiração.

Acima das pregas vocais está uma segunda prega de tecido denominada falsa prega vocal. As falsas pregas vocais são importantes ao impedir que substâncias entrem na traquéia durante deglutição. Elas não fazem o papel principal na fala, e, ao contrário das verdadeiras pregas vocais, elas não devem ter contato entre si durante a fala.

Acima da prega vocal está uma estrutura cartilaginosa flexível em forma de língua chamada epiglote. A epiglote é projetada sobre a abertura da laringe quando engolimos, contribuindo para impedir que algum material chegue aos pulmões.

Figura 2.5. Seção cruzada da Laringe

A posição das pregas vocais durante a fala depende do tipo de sons que estão sendo feitos. Um agrupamento de sons é chamado vozeado, porque sua produção depende diretamente da vibração das pregas vocais. Outro agrupamento é chamado de não vozeado, e para estes sons as pregas vocais estão usualmente abduzidas e o som é produzido por uma outra parte da boca e da garganta. Exemplos de sons não vozeados são "f" e "s". Estes sons não são produzidos pela laringe mas pelo fluxo de ar turbulento em determinadas partes da boca.

2.1.3. Vibração das pregas vocais

Para produzir sons visando a fala, as pregas vocais são primeiramente aduzidas pelos músculos da laringe. Quando estão aduzidas, a ação dos músculos respiratórios gera pressão de ar imediatamente abaixo das pregas vocais. Eventualmente a pressão abaixo das pregas vocais excede a tensão que as une, e um pulso de ar escapa através das pregas.

Enquanto o ar flui rapidamente através da laringe, esta diminui a pressão (fenômeno chamado efeito de Bernoulli) e as pregas vocais são aduzidas. A pressão abaixo das pregas aumenta novamente, e o processo repete-se. Os rápidos processos de abdução e adução produzem vibração das pregas vocais que podem ser verificados pelo exame estroboscópico. Cada vez que as pregas vocais abrem, elas produzem um fluxo de ar que altera a pressão do ar produzindo sons, que irão formar a fala.

É importante ressaltar que as pregas vocais não produzem sons vibrando como “cordas de guitarra”, por exemplo. Ao contrario disso, o som é produzido pelas mudanças de pressão geradas por pequenos fluxos de ar que passam através dos movimentos pregas vocais.

A estrutura microscópica das pregas é extremamente importante em termos do caminho em que o som é produzido e nas técnicas médicas, cirúrgicas e fonoterapêuticas utilizadas para tratar distúrbios da voz.

A figura 2.6 mostra uma vista sobre uma seção transversal com uma das pregas vocais. Pode-se ver que é realmente uma prega do tecido, e o termo “corda vocal” pode ser um tanto enganador.

Figura 2.6. Seção transversal das pregas vocais.

A prega vocal é coberta por uma fina camada de tecido chamada de epitélio. Subjacente a mucosa esta uma segunda camada chamada lâmina própria, e sob esta está o músculo que compõe o corpo da prega vocal (chamado músculo tiroaritenóideo). Na prega vocal saudável, a camada mais superficial da

lâmina própria está completamente frouxa, de modo que o epitélio possa vibrar em forma de onda sobre as camadas subjacentes.

2.1.4. Onda mucosa

A figura 2.7 mostra um diagrama esquemático de uma seção transversal de abdução e adução das pregas vocais. Observe que as pregas não abrem de uma vez, a parte mais inferior primeiramente começa a abduzir e para um ponto mais atrasado nesse instante a parcela superior das pregas separa-se. Esta maneira complexa de abertura e fechamento produz o que é chamado de onda mucosa.

Uma boa onda mucosa é necessária para que as pregas possam abduzir em uma forma uniforme e simétrica. Qualquer coisa que interfira com esta onda mucosa, tal como um inchamento ou um cisto nas pregas, perturba a onda mucosa e causa piora na qualidade da voz. Quando se examina a laringe, a onda mucosa pode ser estudada através da luz estroboscópica.

Figura 2.7. Seção transversal da abertura e fechamento das pregas vocais.

2.1.5. Articulação

O som produzido pelas pregas vocais consiste de uma série de harmônicos, mostrada na figura 2.8. O harmônico mais baixo, que é geralmente

o mais forte, é chamado a frequência fundamental. Para homens a frequência fundamental é aproximadamente de 100 Hz; é ligeiramente mais elevada nas mulheres e nas crianças. O gráfico da figura 2.8 mostra que diminui intensidade em cada harmônico quando do aumento da frequência.

Figura 2.8. Harmônicos produzidos pelas pregas vocais.

O comprimento do trato vocal, estendendo-se das pregas vocais aos lábios, é de aproximadamente 17 cm nos adultos. A forma do trato vocal mudará quando movermos nossa língua, mandíbula, palato e lábios. Dependendo da posição destas estruturas e da forma geral do trato vocal, determinadas frequências serão amplificadas pelo trato e outras poderão ser atenuadas ou permanecer inalteradas. Aquelas frequências que são amplificadas são chamadas de formantes.

Figura 2.9. Efeito das pregas e trato vocal.

O efeito das pregas vocais e do trato vocal é mostrado na figura 2.9. Os harmônicos são modificados na amplitude pela função de transferência do trato vocal. Alguns exemplos da posição da cavidade oral para determinadas vogais, e os formantes produzidos por essa posição são mostrados na figura 2.10. Note que sons diferentes envolvem posições diferentes da língua produzindo uma mudança significativa nos formantes.

Figura 2.10. Posição da cavidade oral para algumas vogais.

Em geral, o formante F_1 varia muito com a altura da língua (acima ou abaixo a boca) ou com a abertura da mandíbula e F_2 varia com a posição da língua (para frente ou para trás). Para extrair a frequência fundamental de um sinal de voz aplica-se um filtro passa-baixa para eliminar a influência dos formantes. O filtro deve ser capaz de excluir a menor frequência do subsistema articulatório (formante F_1) mantendo, porém, a frequência fundamental. A vogal que apresenta a maior frequência F_1 é a vogal /a/. Esta vogal é muito usada no exame visual da garganta, pois a abertura da mandíbula e a posição da língua

permitem uma boa visão a olho nu ou com o auxílio de algum dispositivo ótico (KENT et al, 1992).

Portanto, o processo de modificar a forma do trato vocal para produzir diferentes sons é chamado de articulação.

2.2. Fisiologia da Laringe

Funcionalmente, a laringe tem um papel importante em quatro mecanismos essenciais. A laringe funciona primeiramente como um esfíncter*, protegendo o fluxo de ar da possível penetração de líquidos, alimentos e os outros resíduos. Tem também um papel na respiração, através da dilatação muscular da abertura laringea. A terceira função da laringe é a fonação e parece ser uma aquisição filogenética atrasada. A quarta, e menos reconhecida, função da laringe é seu papel em levantar e em esticar, que é chamada função de esforço de fechamento. Assim, os principais papeis da laringe são:

1. proteção;
2. respiração;
3. fonação e
4. esforço de fechamento.

Embora os estímulos fora da laringe tal como a vista e o som pareçam afetar a função laringea, os receptores dentro da laringe tem um papel importante em modular o movimento e posição dos músculos laringeos. As funções laringeas básicas são o resultado de reflexos polisinápticos interrelacionados. Embora a função protetora seja inteiramente reflexiva e involuntária, o esforço de fechamento, a respiração e a fonação podem ser iniciado voluntariamente ainda que estejam reguladas através de um conjunto de reflexos. A figura 2.11 mostra um diagrama da fisiologia da laringe.

Figura 2.11. Fisiologia da Laringe

**Esfíncter é um músculo que interrompe uma passagem ou fecha um orifício natural. Quando esta relaxado permite que materiais passem através da abertura. Quando contraído, fecha a abertura.*

2.2.1. Proteção

A laringe age como um esfíncter necessário para fechar o fluxo de ar, impedindo a aspiração de alimentos, líquidos e outros materiais estranhos que possam entrar na cavidade faríngea. O exemplo principal desta função protetora da laringe é o fechamento da deglutição, durante a qual a musculatura laringea fecha o fluxo de ar durante o processo de engolir. Embora o reflexo glótico do fechamento possa ser classicamente eliciado pela estimulação do nervo laringeo superior direito, deve-se saber que outros estímulos sensoriais (ex: a estimulação do principal nervo aferente craniano, dos nervos sensoriais somáticos espinal e nervos sensoriais especiais) é também capaz de elicitar este reflexo básico como resposta.

2.2.2. Adução Laringea

As etapas básicas da adução laringea durante a função protetora podem ser vistas na figura 2.12 abaixo. O fechamento esfíncteriano, que é inteiramente reflexivo e involuntário, produzido por resultados da estimulação do nervo laringeo superior bilateral na adução de três camadas musculares dentro da estrutura laringea : as pregas aríepiglóticas, as pregas vocais verdadeiras e falsas. Quando posicionadas pela contração muscular, as falsas e verdadeiras pregas vocais selam e impedem que partículas de alimentos invadam o fluxo de ar.

As pregas ariepiglóticas agem como uma elevação junto a laringe, permitindo que o alimento passe sobre um ou outro lado da epiglote ao longo da fossa piriforme produzida entre cada prega e a parede lateral faríngea. Vide figura 2.13 abaixo.

Figura 2.12. Adução laringeana.

**Laringe durante a fala
(Adução das pregas vocais)**

**Laringe durante a respiração
(Abdução das pregas vocais)**

Figura 2.13. Exemplos de abdução das pregas vocais.

2.2.3. A elevação laringea

O movimento ascendente da laringe é produzido principalmente pela contração do músculo da tireóide e pela relaxação do músculo do esternohióide. Esta ascensão do complexo laringotraqueal e do movimento do terço posterior da língua juntas comprimem a epiglote posteroinferiormente. A configuração

protetora resultante da epiglote direciona o alimento engolido para as fossas piriforme.

2.2.4. Respiração

A laringe é a passagem do fluxo de ar, e assim que o ar entra no sistema respiratório esta deve estar amplamente aberta. Durante a inspiração, as pregas vocais abduzem muito da linha central resultando num aumento da abertura. Esta abdução é executada principalmente pelo músculo posterior cricoaritenóideo. Entretanto, durante a expiração há uma abdução relativa das pregas vocais e um estreitamento da abertura. A cricotireóide tem também um papel crítico durante a fase expiratória da respiração. Durante a respiração completa este movimento é extremamente rápido, porém a inspiração muito profunda resulta nas pregas vocais se apresentarem quase em nível de encontro às paredes laterais da cartilagem da tireóide. Além disso, de acordo com alguns pesquisadores, a laringe inteira é deslocada para baixo na inspiração e para cima na expiração. Estabelece-se agora que a glote abre uma fração de segundo antes que o ar seja inspirado pela descida do diafragma. O alargamento da glote com estouros rítmicos da atividade nos nervos laringeanos recorrentes e superiores é um efeito direto do centro respiratório medular. Esta ritmicidade é acentuada pela hipercapnia (nível acima do normal de dióxido de carbono na corrente sanguínea) e obstrução ventilatória e comprimida pela hiperventilação e resultante hipocapnia (nível abaixo do normal de dióxido de carbono na corrente sanguínea). A abdução inspiratória, é de fato, síncrona com a inspiração. O grau de atividade abdutor parece variar diretamente com resistência ventilatória e uma pressão sensitiva é acionada pela resposta influenciada pelo vago estímulo aferente. As etapas básicas da abdução laringea durante a respiração são mostradas na figura 2.14.

2.2.5. Fonação

A fonação humana, quando usada como uma modalidade de comunicação, pode ser ambas intelectual (fala) e emocional. A fala representa o efeito combinado da laringe, da língua, do palato, e das estruturas relacionadas ao vestíbulo oral. A produção do tom fundamental é devido à vibração das pregas vocais uma com a outra, gerada pela passagem do ar entre elas

(fenômeno passivo da vibração das pregas vocais). Entretanto a regulação da qualidade acústica não é um fenômeno passivo.

Figura 2.14. Etapas básicas da abdução laringeana durante a respiração.

2.2.6. Abdução e tensão da prega vocal

A adução das cartilagens aritenóides é realizada principalmente pelos músculos interaritenóides e cricoaritenóides laterais e esta ocorre no início da produção de fala.

Durante a fonação as pregas vocais verdadeiras são posicionadas também perto da linha central pela tensão da corda fornecida pelo músculo cricotiróide. Os músculos da tiroaritenóide fornecem uma modificação isométrica mais fina, refletida pela forma final das pregas vocais. A frequência da vibração depende da:

1. massa vibratória de ambas as pregas
2. sua tensão anteroposterior
3. função *damping* para *pitch* altos
4. pressão subglótica

Além disso, a atividade dos músculos laringeanos extrínsecos afeta o *pitch* alterando a relação espacial entre as cartilagens cricóide e tireóide.

2.2.7. A Teoria Mioelástica da Fonação

Uma das características mais importantes na vibração das pregas vocais é a ocorrência das ondas que viajam na mucosa das superfícies inferior à superior da prega vocal. Estas são referidas como ondas “mucosais”. A existência de uma mucosa flexível é essencial para a ocorrência da onda mucosal. Esta teoria da vibração das pregas vocais é chamada de teoria mioelástica da fonação. As etapas básicas desta ação aero-mecânica são:

1. adução das pregas vocais ;
2. glote fechada ;
3. abertura progressiva das pregas vocais como um zíper (da borda inferior para a borda superior) ;
4. a glote abre ;

A Figura 2.15 a seguir mostra as etapas que são repetidas ciclicamente até que a pessoa saudável se decida voluntariamente prender a fonação.

Figura 2.15. Etapas básicas da Fonação.

2.3. Examinando a laringe

A laringe é posicionada abaixo na garganta e requer algumas habilidades e equipamentos especiais para um exame.

2.3.1. Laringoscopia indireta

A forma mais simples de examinar a laringe é chamada laringoscopia indireta. O especialista pode colocar um espelho pequeno na parte traseira da garganta e dobrá-lo para baixo para a laringe. A luz pode ser refletida para baixo e a laringe pode ser vista no espelho. Laringoscopia indireta é rápida e fácil proporcionando uma vista tridimensional da laringe na cor verdadeira. Entretanto, alguns indivíduos toleram o exame por pouco tempo e ainda, o paciente não pode ser examinado durante a fala normal desde que o espelho deva permanecer na parte traseira da garganta.

2.3.2. Endoscopia flexível e rígida

Existem dois tipos de instrumentos óticos especiais que podem ajudar o médico no exame da laringe. O instrumento mostrado à esquerda na figura 2.16 é um nasofaringoscópio. A parte curvada do escopo é um cabo fiberoptico flexível que pode ser passado através do nariz e através da faringe até que se tenha uma visão superior das pregas vocais. Usando este endoscópio flexível, a laringe pode ser examinada durante a fala normal. Entretanto, a definição com este escopo é reduzida desde que a imagem deve viajar através de um pacote de fibras óticas pequenas. O instrumento à direita na figura 2.16 é chamado de telescópio de 90 graus. Este é colocado na parte traseira da boca, onde há um prisma que reflete a luz a 90 graus, sendo que é um pouco mais difícil para o paciente, mas dá uma imagem extremamente desobstruída e ampliada das pregas vocais.

Figura 2.16. Instrumentos utilizados para examinar a laringe : à esquerda o Nasofaringoscópio e à direita o telescópio 90 graus.

Uma câmera pode ser acoplada a cada um destes escopos e o exame pode ser gravado para a documentação. A figura 2.17 abaixo mostra uma imagem capturada através de um telescópio de 90 graus das pregas vocais normais. As pregas estão parcialmente abertas e a abertura na traquéia pode ser vista entre elas.

Figura 2.17. Pregas vocais parcialmente abertas e a abertura na traquéia entre elas.

2.3.3. O exame estroboscópico

Durante a fala, as pregas vocais vibram 100 vezes por segundo ou mais. Isto é demasiado rápido para ser visto sem um equipamento especial. A fim de examinar com mais cuidado as pregas vocais em ação, uma fonte de luz especial chamada estroboscópio de luz é utilizada. O estroboscópio emite um *flash* de luz curto e brilhante. Se o *flash* do estroboscópio for repetido a mesma taxa de vibração das pregas vocais, estas parecerão "congeladas" no tempo (este é o mesmo princípio da luz de ponto utilizada por mecânicos de automóveis). Se a taxa de disparo da luz for ajustada de modo que seja um pouco mais rápida ou mais lenta do que a taxa de vibração das pregas, as pregas parecerão mover-se

em movimentos lentos. O exame do estroboscópio é extremamente útil porque permite que vejamos como as pregas vocais estão funcionando. Alguns distúrbios de voz são causados por mudanças em alguma(s) das várias camadas de tecido da prega vocal.

Estas mudanças podem não ser aparentes em um exame do espelho ou fiberoptico regular da laringe, mas podem ser vistas com um exame cuidadoso do estroboscópio.

2.3.4. Laringoscopia direta

Outro tipo de exame é chamado de laringoscopia direta (figura 2.18), e este envolve diretamente olhar para laringe. Devido ao reflexo envolvido no processo, a laringoscopia direta é feita mais frequentemente no quarto operando-se sob anestesia geral, portanto, permite apenas uma avaliação orgânica/estrutural e não funcional da laringe. Durante a laringoscopia direta, biopsias podem também ser feitas de alguma lesão suspeita na laringe ou em outras partes da garganta.

A figura 2.18 mostra como a laringoscopia direta é feita. O especialista mantém um instrumento chamado laringoscópio em sua mão, e olha através deste instrumento para examinar a laringe. Um microscópio pode também ser usado para olhar as pregas vocais e outras partes da laringe sob ampliação elevada.

Figura 2.18. Exame de laringoscopia direta : como é feito e um exame.

2.4. Patologias laringeas e vocais

Na tabela 2.1, encontram-se os oito principais sintomas que os pacientes com distúrbios vocais apresentam. Estes sintomas em geral não ocorrem individualmente, aparecendo, conseqüentemente, de forma combinada.

Tabela 2.1 Principais Sintomas (extraído de [COL96]).

Rotidão	Este sintoma reflete uma vibração aperiódica das pregas vocais. Outros termos também utilizados pelos pacientes: voz "rotuca", "áspera" ou "raspada".
Fadiga Vocal	Pacientes queixam-se de cansaço após a fala prolongada e que a fala contínua requer muito esforço.
Soprosidade	Pacientes queixam-se de que são incapazes de pronunciar sentenças completas sem ficar sem ar e precisar reabastecer o suprimento de ar para continuar falando.
Extensão fonatória reduzida	Este sintoma está em geral associado a cantores que se queixam de experimentar dificuldades em produzir notas que, anteriormente, não apresentavam. Tipicamente, estas notas ocorrem na extremidade superior de sua extensão de canto.
Afonia	Afonia significa ausência de voz. O paciente fala em sussurro e pode, às vezes, queixar-se de uma variedade de sintomas, inclusive secura na garganta, dor e uma grande dificuldade para tentar falar.
Quebras de frequência	Um paciente pode queixar-se de saltos periódicos de voz e de quebras de voz. A voz parece fora de controle, e o paciente relata jamais saber que som sairá. Também pode ser descrito como o uso inadequado de falsete ou puberfonia. Este sintoma é relatado por jovens adolescentes que utilizam uma frequência inapropriadamente aguda como voz habitual ao invés da voz masculina típica de frequência mais grave.
Voz tensa/comprimida	Estes pacientes relatam que é difícil falar. Isto pode incluir inability de fazer com que a vocalização comece ou se mantenha. Eles relatam que é um esforço falar, experimentam muita tensão ao fazê-lo e tornam-se fatigados devido ao esforço envolvido.
Tremor	Os pacientes podem queixar-se de que a voz está cambaleante ou trêmula. Eles são incapazes de produzir voluntariamente um som estável sustentado.

A figura 2.19 é um exemplo de uma laringe normal. As pregas vocais verdadeiras são brancas, encontram-se na linha central da fonação, e as estruturas circunvizinhas são cor-de-rosa claro. Na sequência serão apresentados exemplos das doenças da laringe, servindo de parâmetro para comparação.

Figura 2.19. Laringe normal

Os nódulos (fig. 2.20) são as lesões benignas mais comuns das pregas vocais, tanto em crianças como em adultos. Eles são causados pelo abuso contínuo da laringe e pelo mau uso da voz. Estes abusos podem ser: cantar, berrar, tossir e limpar a garganta excessivamente, por exemplo. Os nódulos podem variar em número (até 4 nódulos), e sua resposta a terapia vocal é muito boa. À medida que os nódulos bilaterais aproximam-se um do outro na vocalização, há uma fenda glótica aberta de cada lado do nódulo. Esta fenda glótica aberta resulta numa falta de adução completa das pregas vocais. Esta aproximação incorreta leva a soprosidade na voz e ao desperdício de ar.

As características acústicas do nódulo apresentam *Jitter* e *Shimmer* elevados. Outro fato são as evidências de ruído no espectro, grau do qual depende a severidade da rouquidão e do tamanho da lesão (COLTON et al ,1996).

Figura 2.20. Nódulos nas pregas vocais.

Os pólipos vocais (fig. 2.21) comumente ocorrem na mesma região das pregas vocais como os nódulos, e são mais frequentes unilaterais do que bilaterais. Os pólipos e os nódulos são ambos relacionados ao hiperfuncionamento vocal e apresentam algumas semelhanças físicas.

Figura 2.21. Pólipos nas pregas vocais.

Estes têm a mesma etiologia e diferem apenas em grau. Um pólipo é maior e mais vascularizado, edematoso e inflamatório do que um nódulo. As vozes de pacientes com pólipos unilaterais caracterizam-se por disfonia severa. A prega vocal normal vibra em uma frequência, enquanto a lesão abafa seriamente a vibração da prega envolvida, resultando no que é percebido como rouquidão e sopro, muitas vezes, requerendo limpeza continuada da garganta. As características acústicas são semelhantes ao nódulo, *Jitter* e *Shimmer* aumentados são esperados, assim como o aumento do ruído espectral. O principal sintoma vocal é a rouquidão (BOONE et al, 1994).

A figura 2.22 é um exemplo de um paciente que tem o refluxo ácido na laringe (Doença laringofaríngea de refluxo). A laringe está vermelha e inchada. Este paciente tem também leucoplasia, que é um crescimento branco premalignante nas pregas vocais. À esquerda na figura 2.22 é o primeiro exame, e o à direita na figura 2.23 exame realizado depois de seis semanas de tratamento médico. Nota-se que as mudanças nas pregas vocais estão invertidas.

Figura 2.22. Paciente com refluxo ácido na laringe.

Os cistos vocais são benignos, lesões das pregas vocais acopladas sobre a superfície da mucosa da prega vocal (fig. 2.23). Estes são tratados geralmente com cirurgia ou terapia da fala.

Figura 2.23 À esquerda cisto nas pregas vocais e à direita após remoção.

Presbylaryngis é uma condição causada pelo afinamento do músculo e tecidos das pregas vocais com envelhecimento mostrada na figura 2.24. As pregas vocais têm menos volume do que uma laringe normal e consequentemente não se encontram na linha central. Em consequência, o paciente tem uma voz rouca, fraca, ou soprosa. Esta circunstância pode ser corrigida pela injeção de gordura ou de outro material em ambas as pregas vocais para se conseguir um fechamento melhor.

Figura 2.24. Afinamento do músculo das pregas vocais.

A Laringite é uma inflamação das pregas vocais e da laringe. Pode ser decorrente de exposição a agentes nocivos (tabaco, drogas, álcool), agentes ambientais (poeira), abuso vocal ou pela hidratação inadequada das pregas vocais. O muco grosso e pegajoso impede que as pregas vibrem em um líquido de maneira uniforme. A laringite também pode decorrer de infecções respiratórias superiores que exercem um efeito generalizado sobre a mucosa do

trato respiratório, incluindo a laringe. Frequência e perturbação de amplitude acima do normal são esperadas, assim como o ruído espectral aumentado. O principal sintoma vocal é a rouquidão (COLTON et al,1996), vide figura 2.25.

Figura 2.25. Má hidratação das pregas vocais.

Paralisia das pregas vocais são lesões que afetam o nervo vago em alguma parte ao longo do seu curso, desde a base até a laringe (fig. 2.26). Estas lesões resultam em uma paresia (fraqueza) ou paralisia nos músculos. O nervo vago é o principal nervo que supre a laringe através de fibras sensoriais que controlam todos os músculos da laringe. Este possui duas ramificações, os nervos laríngeos superior (NLS) ou recorrente (NLR). O nervo laríngео superior controla os músculos cricotireóideos e o recorrente os músculos restantes da laringe.

As lesões do NLS ou NLR podem afetar a posição das cordas vocais. Na paralisia NRL, a lesão pode ser unilateral ou bilateral e pode ser do tipo adutor ou abdutor, dependendo dos músculos afetados. A etiologia da paralisia da prega vocal é variada.

Os sintomas perceptuais mais comuns da paralisia unilateral são a sopro e a rouquidão. Ocasionalmente, a diplofonia pode estar presente. A paralisia bilateral do tipo adutor causará sopro severo ou afonia, porém, uma voz quase normal pode estar presente no tipo abdutor. Acusticamente falando, uma maior aperiódicidade (*Jitter* e *Shimmer*), uma extensão de frequência reduzida, níveis de ruídos mais elevados e uma extensão de intensidade vocal reduzida estão presentes (BOONE et al, 1994);(COLTON et al, 1996).

Figura 2.26. Paciente com paralisia das pregas vocais: à esquerda antes da tiroplastia e à direita após a tiroplastia.

O Edema refere-se a um acúmulo de fluido em algum lugar na prega vocal. Ele pode ocorrer profundamente na prega vocal ou em camadas mais superficiais. Quando ocorre na primeira camada da lâmina própria, é referido como Edema de Reinke. Algum inchaço aparente das pregas vocais, especialmente quando localizado, pode também ser um sinal de cisto.

Edema é uma reação natural do tecido a trauma e mau uso. Além do abuso vocal, o edema de Reinke crônico é mais freqüentemente associado ao fumo. Dos sinais acústicos, a freqüência fundamental da fonação é mais baixa do que esperada para o sexo e a idade do paciente. Pode haver freqüência aumentada e perturbação da amplitude, bem como a presença de ruído espectral devido à rouquidão (COLTON *et al*, 1996).

Edema de Reinke é o edema bilateral da camada subepitelial no espaço de Reinke, isto é, na região onde o epitélio cilíndrico das pregas vocais se transforma em plano estratificado. As pregas vocais ficam enormes com um edema, pobre em vasos o que modifica completamente o aspecto anatômico da região glótica. O edema de Reinke é limitado às pregas vocais e atinge até a região aritnóidea, onde alcança maior espessura.

O sintoma clássico do Edema de Reinke é a disfonia, produzida por pregas vocais mais flácidas (KUHL, 1982), vide figura 2.27.

Figura 2.27. Edema de Reinke nas pregas vocais.

Alterações estruturais mínimas são pequenas malformações das pregas vocais, ou seja, alterações pequenas desde o nascimento que podem ou não cursar com alterações da voz.

Podemos classificá-las em: sulco, ponte, vasculodisgenesia, cistos e diafragma laríngeo. Os sulcos são pequenas depressões sobre as cordas vocais. Pontes são algumas alças formadas pela mucosa que recobre as pregas vocais.

As vasculodisgenesias são malformações de algumas veias, muito pequenas, sobre as pregas vocais. Os cistos são pequenas bolsas de retenção de líquido e o diafragma laríngeo é uma mucosa que segura uma prega vocal na outra.

Geralmente essas alterações ocorrem juntas e podem se manifestar em qualquer idade, devido a um abuso vocal, por exemplo, vide figura 2.28.

Figura 2.28. Alterações estruturas mínimas : à esq. Cisto e à direita diafragma laríngeo.

Câncer ou carcinoma no sistema vocal é uma doença ameaçadora da vida que requer manejo médico-cirúrgico, vide figuras 2.29 e 2.30. Os cânceres de lábio e intra-orais, raramente, contribuem para mudanças vocais, porém, podem exercer efeitos negativos óbvios sobre articulação e podem ser fatais. Lesões orais extensas, envolvendo a língua, ou câncer palatino ou velar, podem afetar seriamente a articulação e a ressonância vocal. A rouquidão persistente é conhecida como um dos principais sintomas do câncer.

Em geral, há 3 classificações de câncer laríngeo, dependendo da região da lesão: supraglótica, envolvendo estruturas como as pregas ventricular e ariepiglótica, a epiglope, as cartilagens aritenóideas e as paredes da hipofaringe; glótica, da comissura anterior até as extremidades do processo vocal das aritenóideas; e subglótica, envolvendo a cartilagem cricóidea e a traquéia. Os sinais do câncer da laringe podem incluir: nódulo no pescoço, um alargamento na laringe, aumento de volume no pescoço e rouquidão.

Figura 2.29. Câncer nas pregas vocais.

Figura 2.30 Tumor nas pregas vocais.

Capítulo 3

Fala: Produção, Modelo, Fonética e Fonologia

3.1 O processo de produção de fala : teoria acústica

A teoria acústica de produção da fala procura modelar matematicamente o processo de geração do sinal de fala pelo aparelho fonador humano. As bases dessa teoria foram brilhantemente apresentadas por Fant (1960), através da publicação de *"Acoustic Theory of Speech Production"*.

Podemos dividir o aparelho fonador humano (fig.3.1) em três componentes principais:

Figura 3.1. Aparelho fonador humano.

•**Pulmões:** localizados no interior da caixa torácica, os pulmões controlam a intensidade do fluxo de ar que passa pela laringe.

•**Laringe:** localiza-se logo acima da traquéia, e é formada por cartilagens e tecido muscular. Destacam-se as pregas vocais, que são pregas de mucosa com musculatura e representam um papel fundamental no processo de produção da fala. As pregas vocais são formadas por dois pares de músculos. Durante a respiração normal as pregas estão relaxadas e abertas; no processo de produção de voz, no entanto, as pregas se tencionam e vibram com a passagem do ar. A taxa de

vibração das pregas vocais está diretamente relacionada com a frequência fundamental (grave/agudo) do sinal de voz: nos sons mais agudos, as pregas estão mais contraídas e portanto vibram mais depressa. O comprimento das pregas também influi na taxa de vibração; é por isso que as mulheres, cujas pregas vocais são mais curtas que as dos homens, possuem um tom de voz normalmente mais agudo.

•**Trato vocal:** porção do aparelho fonador humano que se estende desde a glote até os lábios (Figura 3-1). Os diversos elementos formadores do trato vocal são denominados de *articuladores*. O trato vocal funciona como uma *caixa de ressonância*, que atenua ou amplifica certas frequências do pulso produzido na glote (laringe). O movimento dos articuladores determina o formato do trato vocal e, por conseguinte, as suas características de ressonância.

A maneira mais simples de modelar o trato vocal consiste em considerá-lo como sendo um tubo cilíndrico com seção transversal de área uniforme, com uma extremidade aberta correspondente aos lábios e uma fonte de excitação sonora na outra extremidade. Um sistema como o descrito acima funciona como uma caixa de ressonância, onde certas frequências do sinal sonoro gerado na entrada do tubo são amplificadas, ao passo que outras são atenuadas.

As frequências em que ocorre ressonância são dependentes do comprimento do tubo: no caso de um tubo de comprimento L , as ressonâncias ocorrem para os comprimentos de onda:

$$l = 4L, 4L/3, 4L/5, 4L/7, \text{ etc.},$$

os quais correspondem às frequências:

$$f = c/4L, 3c/4L, 5c/4L, 7c/4L, \text{ etc.},$$

onde c é igual à velocidade de propagação do som no meio em questão.

Considerando-se um valor de L igual a 17cm, que é um valor típico para o comprimento do trato vocal, e fazendo-se c igual a 340 m/s (velocidade do som no vácuo), encontramos valores de ressonância em 500Hz, 1500Hz, 2500Hz, etc. .

Essas frequências de ressonância correspondem às frequências onde ocorre a máxima amplificação do sinal de entrada, e são normalmente denominadas de *formantes*. Note que os valores dos formantes são independentes da fonte de excitação: seus valores dependem única e exclusivamente da configuração do trato vocal.

O espectro típico do sinal produzido na laringe é mostrado na Figura 3.3(a).

Ele corresponde a uma sequência discreta de harmônicas, onde o espaçamento entre as harmônicas é igual à frequência fundamental. A energia dessas harmônicas tem uma queda da ordem de 12dB/oitava, por isso a maior parte da energia do sinal concentra-se nas baixas frequências.

De acordo com o modelo fonte-filtro o sinal de fala pode ser considerado como sendo o produto do espectro em frequência do trem de pulsos produzido na laringe pela função de transferência do trato vocal (KENT *et al*, 1992). Assume-se, nesse caso, que a laringe e o trato vocal funcionam como entidades independentes. Essa é, na verdade, uma simplificação do modelo, pois na verdade existe um certo acoplamento entre a laringe e o trato vocal, o que significa que a função de transferência do filtro não é totalmente independente da fonte.

Figura 3.2. (a) Espectro do trem de pulsos glotal (b) Espectro do trem de pulsos glotal filtrado pela função de transferência do trato vocal.

Ao passar pelo trato vocal, portanto, o pulso produzido na laringe sofre um processo de "filtragem", conforme ilustra a Figura 3.2(b). A curva sobre o espectro representa a função de transferência do trato vocal multiplicada com o espectro do sinal glotal, e os picos dessa curva correspondem às frequências de ressonância (formantes).

Além do efeito de filtragem do trato vocal, devemos ainda levar em conta o efeito da *radiação*. Este é um fenômeno que ocorre quando o som escapa dos lábios em direção ao ambiente. O efeito de radiação é equivalente ao de um filtro passa-altas, com amplificação da ordem de 6dB/oitava; para modelá-lo, basta acrescentar um zero à função de transferência do trato vocal.

Levados em conta os aspectos acima discutidos, podemos descrever O modelo fonte-filtro para a produção de vogais é ilustrado na Fig. 3.3 e a análise no domínio frequência é dada pela equação:

$$P(f) = U(f) \cdot T(f) \cdot R(f) \quad (1)$$

onde: f - frequência em Hz ; $P(f)$ é o espectro do sinal de fala, $U(f)$ o espectro do pulso glotal, $T(f)$ a função de transferência do trato vocal e $R(f)$ o efeito de radiação.

Figura 3.3. Representação do modelo fonte-filtro para vogais [Ken92].

O trato vocal humano não é rígido nem tampouco possui seção transversal de área uniforme, onde a movimentação dos articuladores (língua, lábios, mandíbula, etc.) durante o processo de produção da fala determina alterações na área da seção transversal ao longo do tubo. O efeito dessa alteração é a modificação do padrão de ressonância do trato: cada configuração do trato corresponde a um padrão de formantes diferente, cada um desses padrões correspondendo a uma vogal em particular.

Existem outros tipos de sons, além das vogais, que podem ser produzidos pelo aparelho fonador humano. A produção desses sons também pode ser explicada por meio do modelo fonte-filtro.

O processo de produção das vogais nasalizadas, por exemplo, é semelhante ao das vogais orais; nesse caso, no entanto, ocorre uma abertura do véu palatino, permitindo a passagem de ar também pela cavidade nasal. Podemos modelar o trato vocal, nesse caso, não mais como sendo um tubo uniforme, mas sim como dois tubos em paralelo.

As *consoantes fricativas* (/f/, /v/, /s/, /ʃ/, etc.), por sua vez, são produzidas quando ocorre uma turbulência no fluxo de ar devido à existência de uma constrição ao longo do trato vocal. As fricativas podem ser *vozeadas* ou *não-vozeadas*. No caso das não-vozeadas (ex: /f/, /s/, /ʃ/) existe apenas o ruído de turbulência, ao passo que nas vozeadas (ex: /v/, /z/, /ʒ/) ocorre ainda a vibração das pregas vocais. As *consoantes plosivas* (/p/, /b/, /t/, /k/, etc.), por sua vez, são produzidas por meio de uma obstrução total à passagem do ar ao

longo do trato vocal, seguida de uma liberação abrupta do ar retido. Elas também podem ser classificadas como *vozeadas* (ex: /b/, /d/, /g/) ou *não-vozeadas* (ex.: /p/, /t/, /k/)

Os fonemas aqui indicados seguem a notação do Alfabeto Fonético Internacional (IPA).

O processo de produção da voz é matematicamente modelado como mostra a figura 3.3 (RABINER *et al*, 1978) ; (DELLER JR., 1987), sendo que mais adiante nesse capítulo discutiremos mais a respeito do modelo matemático de produção de voz.

Há um chaveamento entre as fontes de excitação periódica e ruído branco, que produzem os sons vocálicos e fricativos, respectivamente.

A variação nos parâmetros do modelo para o trato vocal e para radiação do som podem produzir os sons plosivos.

Os ganhos para vozeados e fricativos correspondem à amplitude do sinal gerado.

Figura 3.3. Modelo Matemático simplificado para a produção de sinais vocais.

3.2 Modelo Matemático do Trato Vocal

Para facilitar a construção do modelo matemático que represente o sistema fonador humano, simplifica-se as estruturas acima descritas em somente dois blocos distintos, responsáveis pela produção da fala, através da junção do excitador “pulmonar” ao gerador de oscilações “glotal”, de modo que se tenha somente um único “subsistema” capaz de produzir o sinal oriundo das cordas vocais. Outra simplificação útil é a de considerar o “subsistema” ressoador como uma estrutura matemática linear, mesmo que variante no tempo (pelas características fonéticas associadas com as alterações nos órgãos supra-glotalis), pois as equações tornam-se mais fáceis de serem manipuladas de modo algébrico. Com tais pensamentos, tem-se abaixo a representação gráfica de tais considerações.

Figura 3.4 - Representação Simplificada do Sistema Produtor de Fala

Uma vez que o trato vocal foi estudado em porções distintas (fig. 3.4), através da avaliação do gerador de excitação, do trato supraglotal e dos efeitos de perda energética (concentrados no modelo de radiação dos lábios), determina-se agora, através da junção de todas as “peças”, o modelo completo do trato vocal.

Tal modelo pode ser visto na figura 3.5. abaixo.

Figura 3.5 - Modelo Representativo do Trato Vocal

O valores de ganho, A_v e A_n , levam em consideração os efeitos de maior ou menor pressão exercido pelo pulmão na fonação.

O modelo do pulso glotal é necessário para que se produza um sinal similar ao produzido pela vibração das cordas vocais de modo a excitar adequadamente a supraglotal artificial e produzir o sinal de voz desejado. Considerando que para sinais vocálicos tem-se uma onda quase-periódica, utilizando um gerador de trem de pulsos, separados por um período de tempo denominado período de pitch (ou frequência fundamental), o sistema que produz o sinal advindo das cordas vocais é dado por:

$$g(n) = \begin{cases} \frac{1}{2}[1 - \cos(\pi n / N_1)], & 0 \leq n \leq N_1 \\ \cos(\pi(n - N_1) / (2N_2)), & N_1 \leq n \leq N_2 \\ 0, & c.c. \end{cases} \quad (2)$$

Este sistema não necessariamente é único e definitivo. Foi construído com base em propriedades espectrais e similiaridade do sinal produzido com uma onda real glotal. Os parâmetros N_1 e N_2 são definidos com base na curva desejada para o sinal de saída deste modelo, com base no efeito de filtragem passa-baixa característico deste sistema.

O modelo para o sinal quando na necessidade de emissão de sinal não-vocálico é obtido na produção de sinal aleatório sem necessidade de

consideração restritiva de distribuição de probabilidade das amostras produzidas. Normalmente emprega-se geradores de sinais aleatórios com espectro de ruído branco.

Indicar as limitações deste modelo é importante para que se compreenda corretamente os resultados que serão produzidos pelas análises sobre as vozes (principalmente pela condição patológica que se deseja compreender e até certo ponto, reproduzir).

O modelo é adequado para sinais de voz que pouco variam durante o tempo, ou seja, vocálicos e não-vocálicos (considerando a estacionariedade por períodos de tempo entre 10 a 20 ms). Sinais plosivos não são adequados para a representação apresentada, dada sua natureza transiente.

Apesar de sucinto, os elementos básicos apresentados neste capítulo servem de base para compreender as manipulações matemáticas que serão realizadas sobre o sinal de voz, a fim de obter-se parâmetros acústicos que representem o sinal de voz, além da descrição da nomenclatura utilizada em análises acústicas e sistema de produção de fala.

3.3. Fonética e Fonologia

3.3.1 Considerações de natureza linguística

Pode-se definir a Linguística com sendo a ciência que estuda a linguagem. Este estudo é centrado normalmente nos aspectos funcionais dessa capacidade do homem mas envolve também os pontos de vista físicos como no caso da **Fonética**

Para melhor compreender os diversos aspectos da produção da fala, é comum utilizar alguma forma de classificação que permita identificar quais são as unidades básicas da fala.

Essa não é uma tarefa trivial, pois a fala é constituída por sinais acústicos de natureza inerentemente contínua, e não por segmentos discretos, conforme faz supor a idéia de classificação.

A Fonética e a Fonologia são dois ramos da Linguística que se preocupam com os aspectos acima mencionados. Elas diferem entre si no enfoque sob o qual analisam o processo de produção da fala, bem como quanto ao modo de classificação das unidades básicas da fala.

Levando-se em conta essas diferenças, pode-se dizer que a Fonética e a Fonologia são ciências complementares, que desempenham um papel fundamental no estudo da fala como mecanismo de comunicação.

3.3.2. Fonologia

A Fonologia é o ramo da linguística que estuda os sons constituintes da fala segundo seu aspecto funcional. Isso significa que a Fonologia trata do papel exercido por esses constituintes dentro do sistema de organização da fala, sem se preocupar com as suas propriedades acústicas ou articulatórias.

A Fonologia estuda uma unidade básica da língua: o fonema. Isto significa que uma sentença qualquer numa determinada língua pode ser descrita como sendo uma seqüência de fonemas. O fonema é uma unidade abstrata, que restringe sua área de atuação a um domínio psicológico e não físico, e que portanto não apresenta características acústicas. O que diferencia um fonema de outro é o seu papel distintivo dentro da língua. Podemos dizer, por exemplo, que no português, /p/ e /b/ representam fonemas diferentes, pois é a partir deles que podemos distinguir palavras como /pasta/ e /basta/.

Por serem unidades abstratas, os fonemas representam classes de sons. Isso significa que dois sons diferentes entre si, mas que não representem papel distintivo dentro da língua, pertencem a uma mesma classe. Considere por exemplo a palavra *porta*. Dependendo da região do Brasil em que nos encontremos, veremos o "r" de *porta* ser pronunciado de várias maneiras diferentes. Cada uma dessas variantes do fonema /r/ constitui um *alofone* de /r/.

Diferentemente dos fonemas, os alofones não possuem papel distintivo na língua (no exemplo acima, não importa qual dos alofones de /r/ seja utilizado, a palavra *porta* terá sempre o mesmo significado). Se um fonema representa uma classe de sons, um alogone representa, por sua vez, uma sub-classe da classe fonema. Uma outra forma de fazer essa diferenciação é dizer que um alogone é um fonema para o qual foram especificadas características além daquelas necessárias para diferenciá-lo de um outro fonema, ou seja, um alogone é um fonema *superespecificado*. A notação normalmente utilizada representa os fonemas "entre barras" (//), enquanto que os alofones são representados "entre colchetes" ([]).

A realização física de um fonema, por sua vez, é denominada de *fone*. O fone não é uma unidade abstrata como o fonema, mas sim uma unidade física real (trecho de sinal acústico). A cada fonema corresponde um número infinito de fones, todos eles com um grau de semelhança suficiente que permita classificá-los como sendo realizações acústicas pertencentes à mesma classe.

3.3.3 Fonética

A Fonética é o ramo da linguística que estuda os sons constituintes da fala de acordo com suas características acústicas e articulatórias. Nesse caso, as unidades da fala são estudadas como sinais sonoros e não como entidades abstratas, como ocorre na Fonologia.

3.3.4. Fonética Articulatória

A fonética articulatória procura descrever os diversos sons da língua de acordo com a dinâmica (posição e movimentação) dos articuladores que constituem o trato vocal humano (lábios, língua, mandíbula, etc.). Podemos levar em conta tanto o *modo de articulação* como o *ponto de articulação* na classificação dos sons constituintes da fala.

3.3.4.1. Modo de articulação: Ao classificar os sons da língua segundo o modo de articulação, estamos levando em conta o caminho percorrido pelo fluxo de ar ao longo do trato vocal, bem como o grau de obstrução por ele encontrado durante esse trajeto. De acordo com esse critério, podemos dividir os sons da fala em:

- **Vogais e Ditongos:** na produção das vogais, a passagem do ar pelo trato vocal é livre, e a configuração dos articuladores é estacionária. Ex.: /a/, /e/, /i/, /o/, /u/.

Já os ditongos são encontros vocálicos. Eles são semelhantes às vogais, no entanto diferem destas no que diz respeito à dinâmica dos articuladores do trato vocal: se no caso das vogais os articuladores se encontram em uma configuração estacionária, os ditongos apresentam

uma característica dinâmica, pois a posição dos articuladores varia à medida em que ocorre a transição de uma vogal à outra.

Ex: /a/, /e/.

- **Nasais:** na produção das nasais a passagem de ar pelo trato também é livre; nesse caso o fluxo de ar passa também pela cavidade nasal, o que ocasiona um acoplamento desta com a cavidade oral. Ex.: /m/, /n/.
- **Líquidas:** são semelhantes às vogais, com uma obstrução no eixo central do trato vocal.. Ex.: /l/, /r/. Algumas, como o /l/, deixam escapar o ar lateralmente: são as laterais. Outras, como o /r/, obstruem essa passagem rapidamente ou intermitentemente: são as vibrantes.
- **Fricativas:** nas consoantes fricativas ocorre uma obstrução parcial à passagem do fluxo de ar em algum ponto do trato vocal, o que ocasiona a geração de um ruído de turbulência. As fricativas podem ser sonoras (ex.: /v/, /z/, /Z/) ou não sonoras (ex. /f/, /s/, /S/), dependendo da ocorrência ou não de vibração das pregas vocais.
- **Plosivas (oclusivas):** no processo de produção das consoantes plosivas ocorre uma obstrução total à passagem do fluxo de ar, seguida de uma liberação abrupta (burst) desse fluxo retido com ruído. As plosivas também podem ser classificadas como sonoras (ex.: /b/, /d/, /g/) ou não sonoras (ex.: /p/, /t/, /k/).

A figura 3.6 ilustra as três classes de sons que podem ser produzidos. Ela corresponde a digitalização da locução “especial” falada por um locutor masculino, gravada a uma taxa de 11025Hz, com resolução de 16 bits por amostra. Ao total, foram gravados 1,486 segundos de voz ou 16383 amostras diferentes. É possível identificar sons vozeados nas seções A, D e F. Os sons fricativos são representados pelas seções B e E. A obstrução do trato vocal para a produção de um som plosivo é mostrada na seção C, e o som plosivo (acompanhado de um som vocálico) é mostrado no início da seção D.

Figura 3.6 - Sinal digital de voz para a palavra "especial".

3.3.4.2. Ponto de articulação: Os sons da fala também podem ser classificados de acordo com o ponto de articulação, ou seja, o ponto do trato vocal onde ocorre a obstrução máxima à passagem do fluxo de ar. Esse tipo de classificação é válido para as consoantes, uma vez que, no caso das vogais, não se pode falar exatamente em "obstrução" à passagem do fluxo. De acordo com o ponto de articulação as consoantes podem se classificadas como:

- **Bilabiais:** as consoantes bilabiais são formadas a partir da constrição dos lábios superior e inferior. Ex.: /p/, /b/.
- **Labiodentais:** são produzidas com a constrição entre o lábio inferior e os dentes superiores. Ex.: /f/, /v/.
- **Dentais:** nesse caso, a ponta da língua toca os incisivos superiores. Ex.: /t/, /d/.
- **Alveolares:** a constrição ocorre entre a ponta da língua e os alvéolos (porção do trato situada entre o palato e os incisivos superiores). Ex.: /l/, /R/.
- **Palatais:** são produzidas com o dorso da língua próximo ao palato duro. Ex.: /'/, /ø/.
- **Velares:** nesse caso, a constrição ocorre entre o dorso da língua e o palato mole. Ex.: /k/, /g/.

No caso das vogais, muito embora não se possa falar exatamente em obstrução à passagem do fluxo de ar, podemos nos referir à localização do ponto de constrição máxima.

Nesse caso, as vogais podem ser divididas em *anteriores*, *centrais* e *posteriores*. Elas podem ainda ser classificadas de acordo com o formato dos lábios (arredondados ou não arredondados), bem como pelo grau de abertura dos maxilares (aberta ou fechada).

3.3.5 Fonética Acústica

Procura analisar os sons da fala como sinais acústicos, por isso leva em conta suas características espectrais e da forma de onda. De acordo com a Fonética Acústica, podemos classificar os sons da fala em:

- vogais: as vogais apresentam forma de onda com características periódicas. Cada vogal é caracterizada espectralmente pelo seu padrão formântico. São sinais de alta energia, concentrada principalmente nas regiões de baixa frequência.
- ditongos: são semelhantes às vogais, mas apresentam um padrão formântico dinâmico.
- fricativas: são sinais de forma de onda aperiódica e com baixa intensidade.

Espectralmente, a energia se concentra nas altas frequências. As fricativas sonoras apresentam alguma periodicidade: em seus espectrogramas podemos identificar uma quantidade considerável de energia nas regiões de baixas frequências (barra de vozeamento), devida à vibração das pregas vocais.

- plosivas: são caracterizadas por um silêncio seguido de uma explosão (período curto de alta energia), devida à liberação do ar retido durante a obstrução do trato vocal. No caso das plosivas sonoras também ocorre uma barra de vozeamento devida à vibração das pregas vocais.

- líquidas: são consoantes sonoras semelhantes às vogais, pois possuem padrão formântico definido. No entanto, são sinais mais curtos e de menor energia. No caso das laterais, ocorre a presença de zeros, devido à existência de uma cavidade sobre a língua.
- nasais: são semelhantes às vogais, mas apresentam menos energia, devido à atenuação introduzida pela presença dos zeros nasais, e também pelo fato do ar escapar somente pela cavidade nasal.

Capítulo 4

Medidas Acústicas

4.1. Análise acústica da voz

O aparato vocal no que diz respeito a seu estudo segue duas vertentes :

- ✓ a primeira, considera os modelos dos diversos subsistemas envolvidos visando a síntese de sons e
- ✓ a segunda, segue o sentido inverso, a partir do som irradiado analisa os diversos subsistemas envolvidos.

Na síntese, modelos fisiológicos, acústicos ou essencialmente matemática tem sido elaborados objetivando a irradiação sonora através de computadores.

Na análise, em contrapartida, obtém-se informações no sinal real para posterior inferência no aparato vocal. Levando em consideração a complexidade do aparato vocal, diversos parâmetros acústicos são definidos.

Focaliza-se a glote para possíveis inferências sobre patologias das pregas vocais, o sinal produzido pela adução e abdução (pulso) destas é de fundamental importância.

Dentre os métodos para obtenção do pulso glotal tem-se :

- ✓ aquisição do sinal e processamento para eliminação das componentes relativas ao trato supra-glotal, onde destacam-se o cepstrum (deconvolução homomórfica do sinal) e a filtragem inversa residual ;
- ✓ Medidas de contato direto através da pele do pescoço do paciente (a chamada eletroglotografia) e
- ✓ Utilização de longos tubos acústicos sem reflexão para a eliminação do trato supra-glotal.

O sinal (pulso) obtido através dos métodos descritos acima contém informação sobre o “estado” das pregas vocais e, portanto, a análise ou construção de parâmetros acústicos baseados neste sinal permite inferir sobre o “estado” destas.

Basicamente, três características de um sinal são analisadas :

- ✓ período (ou frequência) ;

- ✓ amplitude e
- ✓ forma de onda.

4.1.1. Análise do sinal sonoro

Essa análise corresponde a avaliação inicial que o especialista utiliza no diagnóstico clínico. A comprovação da soproidade, aspereza ou rouquidão constitui-se na análise acústica do sinal de voz.

A utilização da instrumentação eletrônica nessa análise visa a quantificação e fornecimento de medidas mais precisas das características do sinal de voz. Essa análise pode ser realizada sobre sinais diretos ou indiretos (cepstrum ou resíduo). Para os sinais diretos, os sinais são adquiridos mais facilmente, entretanto, apresentam limitações na extração dos parâmetros pelo fato da estrutura supra-glotal poder mascarar pequenas variações indicando início patológico. Mesmo utilizando microfone de contato na garganta, o tecido do pescoço comporta-se como um filtro passa-baixas e o sinal fica dependente da localização dos eletrodos.

Já para os sinais indiretos, o som é modelado por excitação periódica de um tubo acústico. Essa periodicidade é obtida através da vibração das pregas vocais. Dessa forma, retirando-se o efeito do tubo tem-se a excitação, ou seja, o pulso glotal.

Uma das maneiras de retirada do efeito do trato vocal é a utilização da filtragem inversa do sinal, ou seja, adotando-se o enfoque acústico e estimando-se as contribuições espectrais labiais e do trato vocal, pode-se calcular o inverso destas contribuições, que aplicado ao sinal, tem-se o pulso glotal.

No caso do filtro inverso ser obtido através de um modelo linear de produção de fala, o resultado será o sinal residual, sendo este um enfoque puramente matemático.

Outro enfoque para obtenção do pulso glotal é a técnica cepstrum, que não utiliza filtragem, porém distorce o sinal objetivando separar as componentes espectrais relativas à glote das componentes relativas ao trato vocal e lábios.

De maneira geral, pode-se subdividir a análise acústica do sinal em três grupos :

- ✓ medidas de perturbação ;
- ✓ medidas de ruído e
- ✓ medidas realizadas sobre o sinal residual.

4.2. Medidas acústicas : visão geral

4.2.1. Medidas de Perturbação

As perturbações, ciclo-a-ciclo, da frequência fundamental e de sua amplitude são importantes na avaliação de patologias do trato vocal. Isto inclusive é comprovado pela análise de frames de filmes de alta velocidade.

Alterações extremas e abruptas na frequência fundamental durante a fonação fazem com que estas vozes sejam percebidas como ásperas. O acréscimo de ruído de alta frequência, proveniente da incorreta adução das pregas vocais caracteriza a voz rouca. Desta forma, pode-se perceber a importância da avaliação da frequência fundamental (ou pitch) na análise das disfonias.

A avaliação das perturbações ciclo-a-ciclo de frequência fundamental (*jitter*) e de amplitude (*shimmer*) necessita que seja estimado o período de abertura e fechamento das pregas vocais, e o nível com que a glote abduz-se e aduz-se. Nas primeiras pesquisas, o emprego de papéis milimetrados e medidores de tensão do sinal de voz gravado em mídia magnética permitiam o registro da fonação e a avaliação das perturbações. Com o incremento da tecnologia digital, computadores possibilitaram o desenvolvimento de técnicas automáticas para determinação dos períodos sucessivos do sinal glotal.

4.2.2. Aquisição do sinal de voz

O sinal ideal para análise desta perturbação é o pulso glotal, que possui um conjunto de harmônicas, múltiplos da frequência fundamental, com um *roll-off* de 6 dB por oitava, que corresponde diretamente ao fluxo de ar produzido pela movimentação cíclica das pregas vocais. Entretanto, o nível de energia deste sinal é muito baixo, além de ser extremamente difícil de se obtê-lo diretamente. Cabe então desenvolver métodos para estimá-lo, ou determinar suas propriedades temporais (*jitter* e *shimmer*).

As técnicas de estimação do pitch dependem do modo com que a voz é obtida do paciente. O emprego de microfones tipo cardióide é muito difundido pela facilidade com que é operado e adquirido, sendo especificado com ganho dinâmico, máxima resposta em frequência e unidirecional, de modo que o sinal obtido seja a representação mais fiel da voz produzida pela pessoa. Este sinal contém informações sobre a laringe e pregas vocais, mas também traz as características das estruturas de ressonância.

Microfones tipo acelerômetro, ou de contato, por sua vez, produzem um sinal de voz livre da influência da cavidade supraglotal. A colocação deste na pele do pescoço do paciente, na região da tireóide, gera um sinal de voz com características próximas ao sinal glotal. Entretanto, a atenuação produzidos pelas paredes do pescoço pode mascarar disfonias ainda em um estágio inicial.

Um sinal mais próximo ao glotal é obtido pela eletroglotografia (ROTHENBERG *et al*, 1988; CHILDERS *et al*, 1990), onde um transdutor é colocado ao redor do pescoço do paciente e alterações na sua resistência, compatíveis com movimentação das pregas vocais. Entretanto, ainda questiona-se a relação entre o sinal glotal e o obtido por esta técnica: variação inter-pessoas da anatomia do pescoço e problema com patologias onde não ocorre o contato entre as pregas vocais podem levar à erros de estimativa (ASKENFELT *et al*, 1980).

4.2.3. Período e amplitude da Frequência Fundamental (F_0)

Uma vez determinado o sinal à ser utilizado na avaliação das perturbações, determina-se o valor da amplitude e período de pitch ao longo do sinal estimado.

Para os sinais advindos de acelerômetros e eletroglotógrafo, o emprego de técnicas de busca de pico (*peak-picking*) ou pesquisa de cruzamentos por zero (*zero-crossing*) são comumente empregadas devido a clara definição dos períodos e amplitudes nestes sinais.

O sinal advindo de microfones convencionais (tipo cardióide, por exemplo) trazem consigo as características espectrais do trato supraglotal e desta forma, a determinação dos pontos de busca (valor máximo ou mínimo do sinal de voz, ponto de cruzamento por zero) é dificultada.

Classicamente, existem dois grupos de métodos para determinação do período de pitch: baseados em eventos e os baseados em médias de tempo curto (*short-term*). Os métodos baseados em eventos buscam pontos chaves no sinal de voz e a diferença entre estes caracteriza o *i*-ésimo período de frequência fundamental. Neste grupo encontram-se as medidas por busca de pico, cruzamento por zero e mais recentemente, formas de onda similares (*waveform matching*), onde esta última é baseada em uma minimização do erro médio quadrático entre sucessivos períodos de pitch (MILENKOVIC, 1987). Estas medidas tornam-se versáteis quando é aplicado um filtro passa-baixa no sinal de voz. (TITZE et al, 1993) recomendam que o filtro seja do tipo FIR (fase linear) com frequência de corte entre o primeiro e o segundo harmônico (supõe-se que seja com base em uma estimativa prévia de F_0).

Dentro do grupo de estimadores baseados em médias de tempo curto, encontram-se métodos baseados em autocorrelação (DUBOWSKI, 1976), por diferença média de magnitude, ou AMDF (ROSS, 1974; Moorer, 1974), por filtragem inversa (MAKSYM, 1973; DAVIS, 1979), por análise cepstral (NOLL, 1964; 1966), por compressão harmônica (SCHROEDER, 1968), além de outros. Apesar da precisão destas técnicas, diretamente não devem ser usadas na avaliação das perturbações de frequência fundamental por realizarem uma ponderação média do valor de F_0 , e assim desconsiderar algumas oscilações do valor médio que correspondem à alterações na laringe e pregas vocais. O emprego de métodos híbridos pode ser considerado útil na estimação tanto da amplitude quanto do período da frequência fundamental.

Com relação às medidas de amplitude ciclo-a-ciclo, não há uma padronização sobre quais pontos do sinal de voz devem ser medidos e sua diferença considerada como *shimmer*. Os valores comumente medidos como amplitude de um ciclo são o ponto máximo e mínimo do sinal, apesar dos estudos ainda não indicarem a relação direta entre estes picos e o sinal glotal propriamente dito. Verifica-se que o pico mínimo em um período do sinal de voz é posterior ao fechamento real da prega vocal (energia sonora mínima), mesmo sem a aplicação de filtros passa-baixa (TITZE et al, 1987). A influência da cavidade de ressonância no sinal de voz é responsável por este distanciamento entre os sinais (glotal e de voz).

Uma vez que o processamento do sinal de voz é realizado em computadores, as peculiaridades do processo de discretização devem ser consideradas na análise das perturbações de F_0 .

As medidas de *jitter* são limitadas pela taxa de amostragem enquanto que o *shimmer* tem precisão relacionada à discretização das amostras. Estudos consideraram este problema. Verificou-se que a quantização das amostras deve ser, sem o emprego de interpolação, no mínimo de 9 bits para garantir uma margem de erro de 0.1% em vozes normais na avaliação do (TITZE *et al*, 1987).

Horii (1979) constatou que a partir de 20 kHz de taxa de amostragem, obtém-se valores reduzidos de *jitter*, definindo assim esta frequência como mínima para estudos de perturbação de pitch, uma vez que a resolução desta medida é de 0.05 ms. Posteriormente, Titze *et al* (1987), consideraram que 500 amostras por período de frequência fundamental são necessárias para ter-se um erro de estimativa do *jitter* em 0.1%. Este requisito é elevado demais pois para uma voz masculina, por exemplo, de 100 Hz exigiria uma taxa de amostragem de 50 kHz.

O uso de interpolação nas medidas de amplitude e frequência fundamental permite a redução destes requisitos de taxa de amostragem e discretização. Realizando interpolação parabólica pode-se reduzir o requisito de taxa de amostragem para 50 amostras por ciclo, mantendo o nível do erro no cálculo do *jitter*.

4.2.4. *Jitter e Shimmer*

Com a determinação precisa das amplitudes e períodos de frequência fundamental, é possível efetuar as medições acústicas para identificação de patologias ou característica perceptual. Os índices acústicos encontrados na literatura apresentam diferenças entre o tipo de microfone empregado, armazenamento da voz gravada e, principalmente, tipo de fala produzida pelo paciente (contínua ou sustentada).

Estes parâmetros baseiam-se em uma estabilidade ideal das cordas vocais, obtendo-se então uma variância nula. Alterações involuntárias ou proveniente de patologias produzirão padrões errôneos de vibração nas pregas vocais que serão detectados por medidas baseadas na alteração ciclo-a-ciclo da amplitude

ou frequência fundamental. Pode-se definir *jitter* e *shimmer* como uma medida de dispersão do período e da amplitude ciclo-a-ciclo, respectivamente.

Considerando a magnitude da dispersão das perturbações ciclo-a-ciclo, Horii (1979) obteve o *jitter* médio, sendo calculado pela seguinte expressão:

$$JM = \frac{\sum_{i=1}^N |P_i - P_{i-1}|}{N - 1} \quad (3)$$

onde P_i corresponde ao período de pitch.

O emprego do módulo desta medida pode ser substituído pelo desvio padrão das diferenças entre períodos sucessivos, mantendo-se o sinal, uma vez que estas medidas (*jitter* sinalizado) possuem distribuição normal (teste χ^2).

Lieberman (1961, 1963) verificou através da distribuição das diferenças entre períodos sucessivos que os casos patológicos apresentavam-se na região desta curva, acima de 0.5 ms. Com isto, definiu a medida PFF (*Perturbation Frequency Factor*), ou seja, a integral da distribuição da frequência com variação entre períodos sucessivos superiores à 0.5 ms. Inclusive comentou que esta medida é sensível ao tamanho e posição de crescimentos (pólipos, cistos, nódulos) na laringe do paciente. Mas Hall (1995), empregando medidas de *jitter* e *shimmer* em sinais de vozes e de eletroglotografia verificou que não foi possível discriminar significativamente nódulos vocais em mulheres.

Considerando apenas as alterações de sinal na frequência fundamental, Hecker e Kruehl (1971) desenvolveram a medida DPF (*Directional Perturbation Factor*), sendo definido pelo percentual de alterações de sinal da frequência fundamental. Entretanto, estes pesquisadores empregaram fala contínua (*Rainbow Passage*) e assim, seus resultados devem ser ponderados com cuidado pois a prosódia e características psicológicas influenciam a F_0 . Posteriormente, Sorensen e Horii (1984) empregaram fala sustentada para avaliação desta medida.

Juntamente com Murry e Doherty (1980), verificaram que esta medida (para perturbações tanto em amplitude como em frequência fundamental, mesmo que os primeiros autores não a tenham usado para amplitude), é estatisticamente significante na discriminação de carcinomas na laringe. Também verificou-se que não há diferença significativa entre homens e mulheres quando o DPF é empregado. Sorensen e Horii tentaram inclusive

explicar os valores elevados para esta medida - característica patológica - em termos de alterações nos mecano-receptores do sistema reflexogênico que mantém o equilíbrio de forças na musculatura das pregas vocais, apesar desta explicação ser apenas uma suposição teórica.

Reconhecendo a dependência da perturbação de períodos de pitch em relação à frequência fundamental, Horii (1979) empregou a medida JR (*Jitter Ratio*) para evitar que com o aumento de F_0 , ocorresse uma redução do *jitter* médio. Isto advém da estabilidade “artificial” obtida nesta circunstância pois o aumento de tensão na prega vocal aumenta a adução da mesma, restringindo o espaço glotal e permitindo que os níveis volumétricos de ar sejam menores. Esta medida é definida pela média da diferença absoluta das perturbações dos período de F_0 pelo período médio no trecho analisado. Multiplica-se este valor por 1000 apenas para evitar valores muito pequenos. Horii (1982) verificou estatisticamente que não existem diferenças entre medidas usando este índice obtidas para oito diferentes fonemas vocálicos ingleses. Sorensen e Horii (1983) encontraram diferenças significativas entre homens e mulheres para esta medida.

Seguindo uma abordagem semelhante, Murry e Doherty (1980) usaram a media JF (*Jitter Factor*) que é definida por:

$$JF = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |F_i - F_{i+1}|}{\frac{1}{N} \sum_{i=1}^N F_i} \quad (4)$$

onde F_i denota o recíproco da frequência fundamental.

Verificou-se que esta medida pode ser empregada na diferenciação entre pessoas normais e pacientes disfônicos, discriminando inclusive pacientes com tumor de pacientes com paralisia unilateral. Entretanto, verificou-se que DPF para perturbações em F_0 é mais sensível à alterações de massa na laringe.

Ainda considerando a relação entre frequência fundamental e medida de perturbação, Honjo e Isshiki (1980) consideraram o uso de medidas de frequência em semi-tons. Com isto, determina-se a média da variação em semi-tons, da frequência fundamental à cada período. Constatou-se que valores acima de um semi-tom caracterizam-se por elevada disfonia.

Wendhal (1963, 1966) estudou os desvios ciclo-a-ciclo de período e amplitude através de síntese de voz. Produzindo alterações de *jitter* e *shimmer* no sinal sintetizado proporcionais, este pesquisador determinou que estas medidas relacionam-se diretamente com patologias de característica áspera, pois o aumento de *jitter* e *shimmer* é proveniente do aumento da percepção da aspereza na voz do paciente.

Considerando a avaliação da aspereza, Deal e Emanuel (1978) desenvolveram um índice acústico que se relaciona-se com o aumento de ruído espectral característico deste padrão de voz. Para evitar a polarização por diferenças naturais na frequência modal de diferentes pessoas, empregou-se o coeficiente de variância sobre os períodos e amplitudes de frequência fundamental. Obteve-se então as medidas PVI (Period variability index) e AVI (Amplitude variability index), que correspondem ao coeficiente de variância multiplicado por uma constante (1000). Aplica-se o logaritmo decimal sobre a medida AVI por se relacionar linearmente com a percepção de aspereza. Apesar de bons resultados, estes pesquisadores, que empregaram equipamentos analógicos, não as consideraram eficientes na identificação de aspereza na voz de pacientes (medida de ruído SNL, que posteriormente será apresentado).

Entretanto, Klingholz e Martin (1985) acharam que medidas de *jitter* (mais especificamente o *Jitter Factor*) são capazes da diferenciação entre distúrbio de voz hipo- e hiperfuncionais (a hiperfuncionalidade é uma característica das vozes ásperas) e questionaram os impulsos dente-de-serra usados por Wendhal como sinal sintetizado na avaliação da aspereza. Estes pesquisadores, através de voz sintetizada, mostraram que o nível de ruído espectral proporcionado por perturbações de frequência fundamental é muito superior que os produzidos por alterações de amplitude.

Cabe ressaltar que Gelfer (1995) cita que com o aumento da frequência fundamental, ocorre um mascaramento da condição áspera da voz, pois a hiperfuncionalidade da prega vocal cria uma condição de estabilidade “artificial” desta. Mulheres geralmente apresentam maior aspereza do que homens, apesar de diagnosticadas (através de equipamentos) como normais.

Apesar da possibilidade de uso na avaliação de tremor vocal (Winholtz e Ramig, 1992), as alterações lentas nos períodos de pitch em fala sustentada não estão relacionadas com patologias e podem elevar as medidas de *jitter*, o mesmo

sendo válido para o *shimmer*. Koike (1973) desenvolveu a medida RAP (*Relative average perturbation*), também denominada PPQ (*Period perturbation quotient*) por Takahashi e Koike (1975).

Esta medida também foi aplicada sobre medidas de amplitude, ciclo-a-ciclo, sob o nome de APQ (*Amplitude perturbation quotient*). Davis (1979) e Koike *et al* (1977) buscaram estabelecer a fundamentação matemática destas medidas - RAP, PPQ e APQ - para a discriminação entre patologias e grupo de controle normal.

No cálculo destes parâmetros, determina-se a diferença entre o período/amplitude atual da frequência fundamental e uma estimativa empregando média móvel deste valor. Assim, tem-se a seguinte expressão:

$$RAP = \frac{\frac{1}{N-P+1} \sum_{i=1}^{N-P+1} \left| \frac{1}{P} \sum_{j=1}^P (X_{i+j-1}) - X_{i+0.5(P-1)} \right|}{\frac{1}{N} \sum_{i=1}^N X_i} \quad (5)$$

onde N é a quantidade de períodos/amplitudes usados no cálculo

P é o tamanho da janela para média móvel

X_i é a i -ésima medida de período/amplitude, ciclo-a-ciclo

O tamanho da janela variou conforme o pesquisador: Koike (1973) empregou uma janela de 3 amostras para medidas de oscilação de F_0 e 11 amostras para medidas de amplitude, enquanto Davis (1979), através de teste sistemático do tamanho das janelas sobre um sinal proveniente de filtragem inversa, sugere o uso de $k=5$ para medidas de perturbação em amplitude.

Patologias como laringite crônica, paralisia unilateral e tumor foram discriminadas por estes parâmetros, enquanto seu uso em vozes de pacientes com nódulos não foram significativamente diferentes as vozes normais.

Seguindo um caminho distinto, Schoentgen e Guchteneere (1995) analisaram o *jitter* através de séries temporais. Sua fundamentação está na existência de correlação entre os sucessivos períodos de pitch. Considerando que as medidas de dispersão supõem independência estatística entre amostras, é necessário aplicar um método de descorrelação das medidas de variação de ciclo. Isto é obtido pelo uso de séries temporais por dispensarem a hipótese de

independência estatística. A estimação de um modelo AR possibilita a determinação de resíduos (erros de estimação) e determina-se o PPQ modificado, ou seja:

$$PPQ_{\text{mod}} = \frac{\frac{1}{N-q} \sum_{i=q+1}^N |Z_i|}{\frac{1}{N} \sum_{i=1}^N P_i} \quad (6)$$

onde Z_i é o i -ésimo resíduo da estimação do modelo AR

q é a ordem do modelo AR

N é a quantidade de períodos de frequência fundamental estimados

P_i é o i -ésimo período de pitch.

A ordem do modelo é determinada a posteriori baseado em teste estatístico χ^2 para verificação da brancura do sinal residual, uma vez que o algoritmo de Levinson permite o incremento da ordem do modelo a cada iteração. A aceitação da hipótese no teste motiva a parada da recursão nesse algoritmo.

Esta abordagem permite a redução da diferença entre pessoas, pois elimina a correlação do *jitter*. Verificou-se que a curva de distribuição dos resíduos da estimação é mais estreita (leptocúrtica) que a distribuição do *jitter*, uma vez que o intervalo total é reduzido. Esta distribuição sofre uma redução do seu centro (deslocamento da média).

Verificou-se que a distância entre as distribuições dos resíduos entre pessoas normais e pacientes (utilizada na verificação de separação entre populações) foi reduzida, provocada pela redução no intervalo interquartil.

Através do coeficiente de correlação serial (r_k) sobre as medidas de amplitude da frequência fundamental, Koike (1969) e Iwata (1972) foram capazes de determinar padrões no correlograma para o padrão normal de voz e patologias como paralisia unilateral e carcinoma. Empregando acelerômetro e “lag” variando de 1 até 15, demonstra que as medidas de amplitude são úteis na discriminação de patologias.

De modo a padronizar a maioria das medidas de perturbação (*jitter* e *shimmer*) em torno dos conceitos que as fundamentam, Pinto e Titze (1990)

tentaram unificar os parâmetros acústicos em torno de derivadas de ordem superior das medidas de perturbação.

Foi sugerido o uso de medidas estatísticas como média, mediana e desvio padrão para avaliar a extensão da perturbação. Isto é confirmado por Askenfelt e Hammarbert (1986) que consideraram o desvio padrão das derivadas discretas de primeira ordem como melhor índice para aplicação no diagnóstico de patologias (eles empregaram fala contínua em seu trabalho).

Foi avaliado a robustez desta medidas em rejeitarem modulações (senoidal, ruído gaussiano e tendência linear). Esta unificação teve o intuito de criar uma terminologia comum entre os pesquisadores e facilitar a comparação entre resultados.

A maioria das medidas acústicas para determinação de *jitter* e *shimmer* foram descritas nesta seção. Variações deste conjunto são possíveis e comumente encontrados, como por exemplo, emprego de função logarítmica e eliminação da tendência linear.

Entretanto, alguns esclarecimentos ainda se fazem necessários. Na determinação dos parâmetros de perturbação, é necessário considerar a quantidade de ciclos que serão analisados para obtenção de uma medida confiável. Titze *et al* (1987) recomendam o emprego de no mínimo 20 ciclos para determinação dos índices acústicos.

Verificou-se que as medidas de perturbação decaem assintoticamente para um valor estável a partir desta quantidade de ciclos. Deve-se notar que a maioria das medidas acústicas empregam a média estatística; o emprego de estimativas robustas como a mediana possivelmente permite a redução desta quantidade.

A faixa de idade, tanto das amostras normais quanto dos pacientes tem influência sobre as medidas acústicas. Murry e Doherty (1980) verificaram que DPF e JF são mais elevados para pessoas de maior idade; isto provavelmente está associado ao envelhecimento das pregas vocais e órgãos anexos envolvidos na produção da fala. É importante ser cuidadoso na comparação com resultados da literatura e valores dos índices produzidos para pessoas idosas.

Um ponto de discordância entre pesquisadores é o tipo de voz a ser empregado na análise das perturbações (sendo válido para demais tipos de medidas). Alguns trabalhos empregam a fala contínua considerando-a como a

mais natural que a pessoa pode produzir, reduzindo os efeitos psicológicos da metodologia para gravação.

Entretanto, Horii (1979) afirma que as medidas obtidas para este tipo de fala devem ser cuidadosamente consideradas pois alterações articulatórias do trato supraglotal influenciam os parâmetros acústicos, mascarando patologias e evidenciando disfonias inexistentes. O emprego de voz sustentada permite a fixação do trato supraglotal em uma configuração quase fixa, e “somente” alterações na laringe e pregas vocais influenciaram o sinal de voz. Através do emprego de eletroglotografia ou acelerômetros, estas considerações não se fazem necessárias.

As vozes sustentadas são caracterizadas pela produção de fonemas vocálicos. A escolha destes também varia conforme o pesquisador. Os fonemas mais empregados na literatura são o /a/, /e/ e /i/, sendo esta escolha motivada pelas características da estrutura formante, menor amplificação das componentes não harmônicas (ruído) pela cavidade supraglotal, facilidade na determinação dos períodos e amplitudes de frequência fundamental.

Cabe citar que o fonema /i/ tem características mais ásperas que os outros dois fonemas. Isto advém da característica de maior tensão das pregas vocais, causada pela elevação das aritenóides. A estabilidade do sinal produzido é superior, mas devido a característica ressoadora da configuração do trato supraglotal, uma quantidade de ruído proveniente da glote é amplificada. Com características opostas, o fonema /a/ é menos estável, mas o sinal de voz possui menos quantidade de ruído.

4.2.5. Medidas de ruído espectral

O ruído espectral corresponde, geralmente, ao aumento da produção de ar turbulento devido à incompleta adução das pregas vocais. A percepção acústica destas alterações no sinal de voz é a soproidade.

Patologias que produzem percepção de rouquidão também permitem que grande quantidade de ar atravesse a glote durante a fonação. Assim é importante determinar a quantidade de ruído existente no sinal de voz. Além disto, alterações na estrutura espectral, como deterioração energética das componentes harmônicas e aumento da largura de banda nas frequência

formantes devem ser contabilizados para uma melhor discriminação das patologias.

A presente seção apresenta as medidas acústicas que consideram estas modificações no espectro, buscando representar o comportamento incorreto das pregas vocais, seja por disfonias funcionais ou orgânicas.

Empregando um analisador de frequência de alta resolução (equivalente à densidade espectral de potência), Remacle e Trigaux (1991) verificaram que no intervalo entre 0 e 1 kHz é possível perceber visualmente diferenças entre uma voz normal e voz patológica provenientes de cistos e nódulos na corda vocal, com o emprego de uma escala linear ao invés de logarítmica. Descreveram o nódulo como possuindo o primeiro harmônico inalterado, mas os demais sofrem decréscimo de intensidade. Identificaram que casos de nódulos de pequena extensão eram muito similares à voz normal. Apesar do caráter qualitativo, pode-se perceber que o espectro permite identificar características particulares das disfonias.

Utilizando-se da densidade espectral de potência médio em composições fonéticas consoante-vogal, Naranjo *et al* (1994) à uma taxa de amostragem de 48 kHz, verificaram que a energia acima de 6 kHz a energia é significativamente maior para pessoas normais do que para pacientes disfônicos. Apesar de empregar sílabas, somente a porção vocálica é analisada, pois a coarticulação representa uma não estacionariedade no trato supraglotal.

Shoji *et al* (1992-1), considerando pacientes com voz soprosa e usando frequência de amostragem em 44,1 kHz, determinou também a espectro de potência média para 20 segmentos de 243,8 ms, com sobreposição de 50%. Determinando-se a energia a partir de uma frequência de corte, percebeu-se que a partir de 6 kHz os pacientes com sopro são discriminados adequadamente.

Outra forma de visualizar os ruídos é através do espectro médio de longo tempo (LTAS ou *Long-term average spectra*). Frokjaer-Jensen e Prytz (1976) que corresponde a determinação da densidade espectral de potência média para um grande número de amostras.

Estas medidas da densidade espectral correspondem à retirada da ordenada tempo do espectrograma. Uma simplificação que pode ser empregada em voz sustentada pois admite-se novamente a estacionariedade do sinal de voz.

Assim, as discussões sobre resolução temporal e espectral permanecem válidas para estas medidas, qualitativas ou quantitativas.

Uma das medidas mais encontradas nos estudos de disfonias é o HNR (*Harmonic-to-noise ratio*), desenvolvido por Yumoto, Gould e Baer (1982). Originalmente, são consideradas as vozes roucas. O método baseia-se na suposição de que o sinal de voz é composto de uma componente periódica e outra, ruidosa (com distribuição de amplitudes com média zero).

Determina-se uma estimativa de um período de frequência fundamental sem ruído, $s_a(t)$, através da média de um conjunto N de trechos de comprimento igual ao i -ésimo período de pitch. Assim, esta medida pode ser obtida pela seguinte relação

$$HNR = \frac{N \cdot \sum_{t=1}^{T_i} s_a^2(t)}{\sum_{i=1}^N \sum_{t=1}^{T_i} [s_i(t) - s_a(t)]^2} \quad (7)$$

onde T_a é o comprimento da estimativa livre de ruído

T_i é o comprimento do i -ésimo trecho do sinal de voz

s_i é o i -ésimo trecho

O denominador corresponde à energia do ruído estimado. Foi considerado $N=50$ nesta estimação. Pode-se verificar a necessidade de uma estimativa precisa do período de frequência fundamental. Estes pesquisadores citam que esta técnica não pode ser empregada em casos extremos de rouquidão por não ser possível identificar os trechos periódicos, e assim, determinar HNR; porém verificaram que esta medida é estatisticamente significativa na avaliação de rouquidão.

Para contornar o problema da determinação dos períodos de F_0 na estimação de HNR, Awan e Frenkel (1994) desenvolveram um algoritmo empregando medidas de perturbação (*jitter* e *shimmer*), além de busca de picos e cruzamentos de zero, possibilitando uma redução na quantidade de erros na determinação dos trechos utilizados nessa medida acústica. Isto permite uma redução do desvio padrão de HNR para os grupos de patologia, aglutinando-os espacialmente, o que possibilita uma discriminação patológica mais robusta.

Outra medida para identificação de níveis de rouquidão é o Hr (*Relative harmonic intensity*), baseando-se na existência de proeminentes componentes de frequência fundamental no espectro de potência, usando escala linear, enquanto que as regiões laterais contém ruído proveniente da turbulência do ar na glote.

Assim, determinando esta curva sobre 4096 pontos no trecho mais estável do sinal de voz. Hiraoka *et al* (1984) definiram então esta medida como a percentagem de energia das componentes harmônicas a partir da segunda frequência harmônica, pela energia total do sinal de voz. Não foi utilizado a primeira harmônica pois verificaram que algumas vozes roucas apresentam elevada amplitude nesta componente, fazendo com que a medida acústica possua um grande valor e a discriminação patológica seja prejudicada; sem seu emprego, Hr apresenta valor limite para pessoas normais em 67,5%, enquanto que se esta componente fosse utilizada, este limiar sobe para 76,5%, para o fonema /a/ (que apresenta alta energia na formante mais baixa).

Empregando expansão de séries de Fourier, Kojima *et al* (1980) construiu um parâmetro acústico relacionando a energia entre componentes acústicas e ruidosas de um sinal construído a partir de três trechos com período de frequência fundamental.

Determina-se então a energia das componentes harmônicas (S_i) e não-harmônicas (N_i) da densidade espectral obtida do sinal gerado para cada segmento de três períodos de pitch até a frequência de 5500 Hz, percorrendo 325 ms do sinal de voz. Finalmente determina-se a medida R_a (*Relative harmonic*) através da expressão

$$R_a = 10 \cdot \log \frac{\sum_{i=1}^N S_i}{\sum_{i=1}^N N_i} \quad (8)$$

Esta técnica tenta ampliar a resolução da transformada de Fourier por utilizar um sinal altamente periódico, sendo capaz de diferenciar vozes soprosas de ásperas.

A medida da soproisidade desenvolvida por Fukazawa *et al* (1988) permite uma avaliação simples do ruído de turbulência produzido pela glote nesta condição disfônica. A medida Br (*Breathiness ratio*) corresponde ao percentual

de energia de alta frequência por energia total do sinal. Esta energia de alta frequência é obtida pela derivada discreta de segunda ordem (filtro passa alta com +12 dB/oitava) do sinal de voz, ou seja:

$$Br = \frac{\sum_{i=1}^{N-2} [f(t_i) - 2f(t_{i-1}) + f(t_{i-2})]^2}{\sum_{i=1}^N [f(t_i)]^2} \quad (9)$$

A vantagem desta medida é não depender da determinação de período de frequência fundamental, uma vez que algumas patologias dificultam esta estimação. Algumas medidas que foram apresentadas (HNR, por exemplo) dependem destas avaliações para permitirem a identificação de patologias. Além da extrema simplicidade de obtenção, possibilita discriminar entre disfônicos soprosos e pessoas normais. Os autores sugerem o uso adicional de outras medidas para melhorar a separabilidade entre câncer, pólipos e paralisia unilateral.

Investigando vozes ásperas, alguns pesquisadores (Sansone e Emanuel, 1970; Lively e Emanuel, 1970; Emanuel *et al*, 1973; Emanuel e Whitehead, 1979; Trullinger *et al*, 1988) desenvolveram uma medida denominada SNL (*Spectral noise level*) baseada na determinação das menores intensidades espectrais em intervalos de 100 Hz na curva média de 37 espectros de potência, com sobreposição de 256 pontos. Os intervalos estudados são 100 a 2600 Hz, 2600 a 5100 Hz, 100 a 5100 Hz, 100 a 8000 Hz e 100 a 2000 Hz.

Determinaram que existem vogais que apresentam mais características de aspereza que outras (/u/ e /i/), possivelmente pela característica das formantes estarem deslocadas na direção das altas frequências, pelo posicionamento da língua, e maior tensão das pregas vocais. Além disto, verificaram que o ruído espectral está relacionado com perturbações na frequência fundamental. Isto também é explicado pelos estudos de simulação de aspereza de Klingholz e Martin (1985). Verificaram que existe uma alta correlação entre esta medida e a aspereza quando empregado os intervalos 100 a 2600 Hz e 100 a 2000 Hz. Isto porque as componentes harmônicas são confundidas pelas amplitudes de ruído nas altas frequências.

Empregando um método adaptativo de filtragem, mais robusto que HNR e R_a , Kasuya *et al* (1986) construíram uma medida no domínio da frequência para

estimar o ruído entre as componentes harmônicas, que suporta alterações suaves na amplitude e período de F_0 . Considerando que o sinal de voz é composto por componentes periódicas e ruído aditivo, a medida NNE (*Normalized noise energy*) é quantificada pela expressão:

$$NNE = 10 \cdot \log \frac{\frac{1}{L} \sum_{i=N_L}^{N_H} \sum_{m=1}^L |\hat{W}_m(i)|^2}{\frac{1}{L} \sum_{i=N_L}^{N_H} \sum_{m=1}^L |S_m(i)|^2} \quad (10)$$

onde: $\hat{W}_m(i)$ é a estimativa espectral do ruído;

$S_m(i)$ é a densidade espectral de potência do sinal de voz;

L é a quantidade de janelas de dados que são utilizadas para análise;

N_L e N_H são os limites inferiores e superiores, em amostras, sobre o qual a medida é desenvolvida, sendo $N_L = N f_L T$ e $N_H = N f_H T$. As frequências f_L e f_H formam o intervalo de frequências na curva de densidade espectral de potência onde a estimativa de ruído é feita. T é o recíproco da frequência de amostragem. N é o número de pontos da transformada de Fourier utilizada no cálculo da curva de energia espectral.

A estimativa do ruído é feita com a seguinte suposição: na região das componentes não harmônicas existe apenas ruído; assim, nestes locais, a energia do ruído é igual a energia das próprias componentes. Já as componentes periódicas são compostas por sinal e ruído, utiliza-se o valor médio da energia das regiões não harmônicas laterais como estimativa do ruído. Assim, fica determinado o numerador da expressão, ou seja, ruído espectral.

As larguras de banda dos componentes harmônicos, utilizadas na estimação do ruído, são obtidas por processo adaptativo, baseado na frequência fundamental. São utilizados 7 períodos de pitch para a janela (Hamming) de dados de 40 ms de extensão, onde se determina o valor de NNE. Este processo é repetido à cada 20 ms e o valor médio de NNE representa a medida de quantidade de ruído para a voz analisada.

Patologias como câncer (do tipo T2, T3 e T4), paralisia unilateral e nódulos foram passível de identificação utilizando este parâmetro acústico para um intervalo de frequências entre 1 e 5 kHz, parecendo informar adequadamente a

quantidade de ruído turbulento que atravessa a glote. Entretanto, as disfonias pólipos, Edema de Reinke e laringite não foram identificadas adequadamente.

O emprego de análise cepstral foi considerada por De Krom (1993) na determinação da relação harmônico-ruído. O princípio do método é considerar que os picos na *quefreny* correspondente a frequência fundamental correspondem às componentes harmônicas do sinal. Retirando-os por um processo de liftering, obtem-se então o ruído cepstral (componentes não harmônicas) modulado pela estrutura formante, pois estas características ainda permanecem - baixo *quefreny* - no cepstrum modificado. Retornando ao domínio da frequência, a diferença de energia, em uma determinada banda de frequência, entre o sinal original e a componente ruidosa (ruído cepstral) fornece o valor de HNR.

Um processo adaptativo é empregado para determinar os picos harmônicos baseado na alteração de sinal da primeira derivada do cepstrum à direita e à esquerda destes picos. Também é realizado uma compensação no valor da componente ruidosa (domínio da frequência) pois as componentes harmônicas devem estar sobre o eixo das abscissas. Esta correção de linha de base é realizada, subtraindo-se do ruído espectral estimado, o módulo do valor mínimo da componente espectral de menor energia no intervalo entre picos harmônicos do espectro livre de ruído (transformada inversa do cepstrum modificado).

Verificou-se posteriormente que esta medida baseada em análise cepstral (DE KROM, 1995) consegue discriminar adequadamente vozes soprosas de ásperas, nos intervalos de frequência de 400 a 2000 Hz e 2000 a 5000 Hz.

Sabendo que a amplitude do primeiro harmônico é muito maior para vozes soprosas do que para vozes normais, De Krom (1990) determinou a diferença entre a energia dos dois primeiros harmônicos. Considerando que vozes soprosas apresentam queda suave de energia no intervalo entre a segunda formante e a banda de ruído turbulento de alta frequência (2000 a 5000 Hz), foi considerado a diferença de energia entre os intervalos acima apresentados. Também foi considerado as bandas entre 60 a 400 Hz e 5000 a 8000 Hz, além de uma medida de frequência fundamental baseada em técnica cepstral (DE NOLL, 1966).

Estas medidas de relação sinal-ruído (HNR), energia e declividade espectral, energia de harmônicos e algumas relações entre elas, tem objetivo de descrever o espectro (grande quantidade de informação) através de poucos valores e relacioná-los com as patologias.

Verificou-se que a percepção de aspereza varia conforme a vogal, confirmando estudos anteriores (SANSONE *et al*, 1970; LIVELY *et al*, 1970; EMANUEL *et al*, 1973; EMANUEL *et al*, 1979; TRULLINGER *et al*, 1988; KLINGHOLZ *et al*, 1985), tendo apresentado resultados de discriminação inferiores aos obtidos com patologias soprosas.

Isto indica que é necessário empregar medidas de perturbação de amplitude e período de *pitch* para identificar tais patologias, apesar de se conhecer a influência do *jitter* e *shimmer* no domínio espectral (KLINGHOLZ *et al*, 1985).

Empregando um método de *pitch* síncrono para avaliação da rouquidão, Muta *et al* (1988) desenvolveram a medida NS (*Noise to signal*), para determinar a quantidade de ruído entre os picos harmônicos no espectro de potência. Primeiro, é extraído um esboço do período de frequência fundamental para um trecho de voz, através de um método de estimação de *pitch* qualquer; em seguida, através da autocorrelação de um novo trecho de voz de comprimento quatro vezes o período de F_0 determinado se obtém um novo período de frequência fundamental que é comparado ao seu esboço. Caso sejam diferentes, um novo tamanho de amostras é utilizado até ambos serem idênticos.

Com a aplicação de uma janela de Hanning sobre as amostras de tamanho quatro vezes superior ao período de *pitch* encontrado, determina-se o espectro de potência destas amostras janeladas, do qual mantém-se somente 4 amplitudes, $P(k)$, por período de *pitch* (igualmente espaçadas).

Uma vez que se têm exatamente quatro períodos da fundamental de amostras da voz analisada, os picos harmônicos são separados exatamente à cada quatro unidades. Assim, determina-se a relação entre picos e vales (sinal/ruído) por:

$$NS = 10 \cdot \log \frac{\sum_{k=3}^{4L+2} \min_{i=-1,0,1,2} P(4k+i)}{\sum_{k=3}^{4L+2} P(k)} \quad (11)$$

onde L é o número de harmônicos empregados no cálculo (no caso, $L=16$).

Primeiramente, verificou-se a capacidade da medida na avaliação terapêutica de pacientes (pré- e pós-operatórios), mostrando ser útil nesta função. Resultados mostraram que o método é sensível à *jitter*, *shimmer* e ruído aditivo, e insensível à pequenas modulações em amplitude e frequência.

Considerando patologias que produzem percepção soprosa, Hillenbrand et al (1994) empregaram um conjunto de medidas de modo a relacionar com este tipo de voz. Suas medidas acústicas baseam-se na suavidade do sinal glotal, o que caracteriza a grande amplitude do primeiro harmônico; a perda de periodicidade nos harmônicos, principalmente nas médias e altas frequências; e na maior energia de alta frequência devido ao ruído turbulento.

Definiu-se a medida CPP (Cepstral peak prominence) como a diferença em amplitude entre o pico cepstral e o seu valor correspondente em uma regressão linear em um intervalo no cepstrum de quefrequency de 1 msec até o máximo quefrequency. Opera como uma medida de periodicidade pois existe relação direta entre a amplitude do pico e quão periódico é o sinal.

Outra medida acústica é o RPK (Pearson r at autocorrelation peak), que é obtida pelo cálculo de coeficiente de Pearson entre o sinal de voz e a sua versão deslocada k amostras, onde k corresponde ao ponto de ocorrência do segundo máximo da função de autocorrelação do trecho analisado. Esta medida é realizada a cada 10 ms, usando uma janela de 30 ms, sendo que o valor médio destas estimativas é apresentado como resultado final. Novamente, esta medida serve para avaliar a periodicidade do sinal de voz., pois espera-se obter valores baixos quando o parâmetro é calculado sobre vozes soprosas.

Também foi empregado a medida P/A (Peak to average ratio) dada pela razão entre pico máximo do trecho analisado pela amplitude média deste mesmo trecho, retificada em onda completa. O resultado final é dado pela média dos valores determinados em sucessivos segmentos de 10 ms.

Além destas medidas sugeridas, Hillenbrand et al (1994) sugerem a utilização dos parâmetros Br, da razão entre a energia acima de 4 kHz (alta frequência) pela energia abaixo desta frequência (baixa frequência) e da amplitude do primeiro harmônico.

Verificou-se que as medidas RPK e CPP correlacionam-se com as patologias com características soprosas, além de simples de serem determinadas. Confirmando os estudos de Fukazawa et al (1988), Br também serviu para discriminação de sopro.

Em um estudo para determinação de rouquidão, Shoji et al (1992) sugere o emprego de sentença fonética /hai/ para avaliação da eficiência glotal. O parâmetro H-Index é calculado pela diferença entre os dois primeiros máximos do envelope de uma curva de densidade espectral de potência. Esta curva é determinada sobre os primeiros 275 ms da sentença fonética, sendo que cada ponto desta curva é determinado pela média de 20 amplitudes no espectro de potências do trecho analisado e ao final é aplicado um filtro suavizador por janela triangular. O princípio fisiológico desta medida é que durante o fonema /h/, a capacidade harmônica da glote na produção de um sinal é baixa (ruído turbulento). Após a coarticulação, o fonema /a/ deve apresentar elevada energia harmônica.

A capacidade humana de aduzir a prega é então determinada, permitindo avaliar a existência ou não de patologias na glote.

Pode-se perceber pelas medidas apresentadas que a capacidade de produção de ruído turbulento na glote é caracterizado pela proeminência do primeiro harmônico em relação aos demais. A partir do segundo harmônico, a amplitude destas componentes é reduzida, fazendo com que a curva de densidade espectral de potência (para médias frequências) torne-se cada vez mais horizontal. Além disto, o ruído de alta frequência eleva-se. Estas três características descrevem as patologias perceptualmente definidas por soprosas. A percepção da rouquidão se deve pela produção simultânea de aspereza e sopro, sendo que a aspereza é um fenômeno caracterizado principalmente por *jitter* e *shimmer*.

Notou-se que a maioria dos trabalhos não tentam discriminar patologias diretamente; apenas relacionar seus índices com identificação perceptual.

4.2.6. Avaliação por Resíduos

Os métodos anteriormente apresentados empregam medidas tanto no domínio do tempo como no domínio da frequência, buscando representar a forma de onda (normal ou patológica) através de uma pequena quantidade de medidas. A avaliação das estruturas da laringe e cordas vocais exigiria que o sinal considerado não sofresse influência do efeito de ressonância proporcionado pela cavidade supraglotal.

O emprego de fala sustentada (usando fonemas vocálicos) visa minimizar esta influência; entretanto, conforme o fonema empregado, características mais ou menos ásperas podem desviar a discriminação para uma patologia, mesmo que seja de pouca extensão.

Sugere-se então o emprego de uma estimativa do sinal de velocidade volumétrica de ar através da estimativa inversa do modelo de produção de voz. Este sinal glotal permitirá, assim, a avaliação do funcionamento e condição das pregas vocais. Rothenberg (1973) empregou uma máscara pneumotacógrafa para avaliar a velocidade volumétrica de ar na boca, juntamente com um filtro inverso para avaliação de apenas 2 formantes (a resposta em frequência da máscara opera como um passa-baixa com $f_c=1$ kHz) e obteve sinais glotais bastante próximos do real.

Entretanto, surgem alguns problemas nesta avaliação:

1. é difícil determinar os parâmetros do modelo inverso do sinal de voz
2. o sinal glotal contém somente informação de baixa frequência
3. são considerados modelos lineares para o trato vocal, que são invalidados para grandes aberturas glotais.

Como alternativa para esta técnica, pesquisadores (KOIKE et al, 1975; DAVIS, 1979; PROSEK et al, 1987; ESKENAZI et al, 1990) sugerem o uso do resíduo proveniente de uma filtragem inversa como fonte de informação sobre a estrutura glotal, uma vez que elimina-se a correlação proporcionada pela supraglotal. Além disto, o sinal residual mantém informação de alta frequência e de periodicidade, permitindo sozinho a discriminação de grande número de patologias.

O sinal residual corresponde ao erro de estimação de um modelo AR para o trato vocal, baseado nas amostras de fala sustentada. O emprego da fala sustentada tem por objetivo manter a supraglotal em um estado estacionário (sem variação das formantes).

As equações de Yule-Walker do modelo paramétrico são solucionadas pela técnica de autocorrelação (usando o algoritmo de Levinson), onde o número de coeficientes estimados é 14. Davis (1979) investigou a influência da quantidade de coeficientes e verificou que com o aumento deste, ocorre um incremento na quantidade de picos no espectro de potência (alta resolução) do filtro AR sem uma relação com o número de formantes.

Também não emprega-se pré-ênfase de alta frequência de forma a compensar a influência dos lábios no sinal de voz pois não há melhoras significantes na discriminação das disfonias.

O segmento de voz analisado compreende, em média, 138 ms de duração, sendo que o ponto inicial de estimação dos coeficientes é feito a partir de 400 ms do começo da fonação. Este procedimento visa obter a porção mais estável da voz através de uma padronização que permita a construção de métodos automáticos de avaliação do trato vocal, eliminando a entonação inicial.

Definiu-se um conjunto de parâmetros acústicos que caracterizem o sinal residual obtido através de poucas medições. A periodicidade é determinada diretamente através dos índices acústicos PPQ (*Pitch perturbation quotient*) e APQ (*Amplitude perturbation quotient*), com procedimento de cálculo semelhante ao de Koike (1973). São empregados 5 pontos para o cálculo das médias móveis destas duas medidas. O intuito é obter-se avaliação de *jitter* e *shimmer*, minimizando o efeito de variações lentas que ocorrem nos sinais de voz, que tendem a induzir discriminações patológicas errôneas.

Também avaliando a periodicidade do sinal residual (e assim, do sinal de voz), a média PA (*pitch amplitude*) corresponde ao segundo pico de maior amplitude na autocorrelação normalizada do sinal residual. Assim, quanto mais próximo proeminente o pico for, mais periódico é o sinal de voz.

Considerando a distribuição das amostras de resíduo, verificou-se que a avaliação da curtose pode ser útil na identificação de patologias. O índice acústico EX (*Coefficient of excess*) avalia o grau de achatamento da distribuição de amplitudes do sinal residual.

Assim, curvas leptocúrticas caracterizam vozes normais e platicúrticas as vozes disfônicas. O cálculo da curtose corresponde à razão entre o momento de quarta ordem da distribuição pelo quadrado do momento de segunda ordem.

Emprega-se a avaliação no grau de nivelamento (*flatness*) da magnitude espectral, tanto do filtro inverso, quanto do sinal residual pois a proeminência dos picos espectrais determina vozes sadias. Assim, os parâmetros SFR (*Spectral flatness of residue*) e SFF (*Spectral flatness filter*) correspondem à medidas de mascaramento de formantes e mascaramento de harmônicos da frequência fundamental, respectivamente, por ruído.

Verificou-se que as alterações orgânicas ou funcionais nas pregas vocais e laringe refletem-se diretamente no sinal residual e por consequência, estes parâmetros podem ser empregados na discriminação patológica. A medida acústica PA é de grande importância, pois correlaciona-se fortemente com a percepção acústica das patologias.

4.2.7. Medidas Realizadas sobre o pulso glotal

O pulso glotal estimado pela filtragem inversa (anulação dos efeitos do trato supra glotal) pode ser analisado por vários parâmetros para se determinar sua forma.

• Quociente de Abertura e de Fechamento

Basicamente o quociente de abertura mede a relação entre o tempo de abertura da glote pelo tempo total (período). Similarmente o quociente de fechamento é a razão entre o tempo no qual a glote permanece fechada e o período. A relação entre ambos é:

$$OQ = 1 - CQ \quad (12)$$

• Quociente de Velocidade

Este quociente determina a razão entre o tempo no qual a glote permanece aberta e aquele no qual permanece fechada.

• Índice de velocidade

$$SI = \frac{Ta - Tf}{Ta + Tf} = \frac{SQ - 1}{SQ + 1} \quad (13)$$

• Relação Excitação Glotal/Ruído – GNE

Proposto por Michaelis (1997), este parâmetro é baseado na correlação entre os coeficientes de envelopes de Hilbert para diferentes bandas de frequência. Esta medida indica quanto do sinal vocálico provém da vibração das pregas vocais e quanto provém da turbulência provocada pela rouquidão.

O processo descrito por Michaelis consiste no seguinte algoritmo :

- ✓ Alterar a frequência de amostragem para 10Khz.
- ✓ Realizar a filtragem inversa do sinal de voz através de um filtro LPC de 13ª ordem com a utilização de uma janela de Hanning de 30ms de comprimento de deslocamento de 10ms.
- ✓ Calcular os envelopes de Hilbert (correspondente ao valor absoluto do sinal analítico derivado da transformada de Hilbert) para 3 frequências de 1KHz de largura de banda (500, 1500 e 2000 Hz)
- ✓ Calcular a função de correlação cruzada entre os diversos envelopes
- ✓ Escolher o maior valor entre as funções de correlação cruzada
- ✓ Escolher o máximo entre os valores obtidos em (5)

Para o cálculo das correlações cruzadas a normalização utilizada foi: inicialmente subtrair a média de cada um dos envelopes calculados, e em seguida dividir cada elemento pela norma do vetor.

A seguir a figura 3.17 mostra um exemplo de voz normal da correlação cruzada entre os envelopes de Hilbert de 500 e 1000Hz para uma janela de 60ms .

Figura 3.17. Exemplo de Voz normal com correlação cruzada.

• Índice do Ruído Turbulento (TNI)

O Índice do Ruído Turbulento foi proposto por Mitev *et al* (2000) de forma a mensurar o fator de ruído em um sinal de voz. Este trabalho é baseado em um medidor de frequência fundamental proposto por Milenkovic (1987). Diferentemente dos parâmetros HNR e NNE que são os mais utilizados para a medição da soproidade da voz, o TNI mede o ruído através da diferença entre sinais subsequentes.

A equação que calcula o TNI é

$$TNI = 100 \left(1 - \frac{1}{N-1} \sum_{n=1}^{N-1} R(t_n, T_n) \right) \quad (14)$$

onde

$R(t_n, T_n)$ é a função de auto-correlação normalizada pela energia

$$R(t_o, T) = \frac{\sum_{i=0}^{T_w-1} v(t_o + i) v(t_o + i + T)}{\sqrt{\sum_{i=0}^{T_w-1} v^2(t_o + i) \sum_{i=0}^t v^2(t_o + i + T)}}$$

onde:

n = ciclo a ser calculado

t_n = início da ciclo glotal

T = Período instantâneo do ciclo

T_w = mínimo período de duração do *pitch* na janela a ser considerada

$$R(t_n, T_n) = \max[R(T, T)] \quad T_{\min} \leq T \leq T_{\max} \quad t_n - \delta \leq t \leq t_n + \delta$$

No caso de semelhança o valor de $R(t_n, T_n)$ será próximo a 1 e o valor de TNI será baixo. No caso de sinal ruidoso o somatório terá um valor menor que 1 e o valor do TNI será elevado.

4.2.8. Outras técnicas acústicas

Individualmente, os parâmetros acústicos determinados não conseguem identificar perceptualmente as patologias, e assim, efetuar um diagnóstico diferencial correto.

Diversos pesquisadores realizaram tentativas de empregar mais de uma medida acústica buscando melhores resultados.

Sanderson *et al* (1992), considerando medidas de frequência fundamental, intervalo de pitch, percentual vocálico e *jitter* médio, para fonação contínua, concluíram que é possível avaliar tratamento terapêutico de pacientes disfônicos através da análise de suas vozes, sendo confirmado por Fex *et al* (1994), que empregaram as medidas F_0 , APQ, PPQ, NNE e variação entre níveis espectrais entre períodos da fundamental. Wolfe e Steinfatt (1987) empregaram medidas de desvio padrão e coeficiente de variação das perturbações de primeira ordem do período da fundamental, *jitter* médio, *jitter factor* (JF), *jitter ratio* (JR) e DPF para discriminação entre características perceptuais da voz.

Foi verificado que avaliação acústica da fala permite diagnosticar fadiga laríngea (*myasthenia laryngis*), que corresponde à um primeiro estágio das patologias orgânicas (EUSTACE *et al*, 1996). Hartman (Hartman *et al*, 1984). Acoustic measurement of voice quality in central dysphonia. *Journal of Communication Disorders*, 17, 425-440 cita que é possível detectar acusticamente disfonias centrais, geralmente associadas à problemas neurológicos que acabam afetando o funcionamento das pregas vocais.

Entretanto, outros trabalhos indicam a dificuldade de uso de medidas acústicas no pré-diagnóstico de patologias. Martin *et al* (1995), empregando medidas de *jitter*, *shimmer* e HNR, citam sobre a dificuldade de diagnosticar soprosidade. Nittrouer *et al* (1990), usando outras medidas (SNR, relação entre primeiro e segundo harmônicos, perturbação de frequência e amplitude da F_0) citam que os resultados não são suficientemente significativos para um uso exclusivo de avaliação acústica.

Utilizando modelamento biomecânico da fala, Smith *et al*, 1992 mostra ser possível alterar estes modelos de forma a representar paralisias unilaterais e bilaterais. Herzel *et al*, 1994, suscita o uso de teoria de dinâmica não-linear (teoria do caos) para diagnóstico de patologias, visto que o sistema de produção de fala é complexo e não linearidades são características do mesmo.

Gubrynowicz *et al* (1980) utilizando um classificador *fuzzy* com medidas de dispersão e assimetria das diferenças entre períodos sucessivos de *pitch* foram capazes de distinguir entre pessoas normais e pacientes disfônicos. Através de classificação por vetores de quantização para medidas acústicas, Childers *et al* (1992) foram capazes de avaliar a condição das pregas vocais, também separando pacientes disfônicos de pessoas normais. Cita que a classificação empregando sinal proveniente de eletroglotografia provê resultados inferiores que os obtidos usando sinais de voz.

4.3. Medidas acústicas : descrição

4.3.1. Aquisição do Sinal de Voz

A aquisição do sinal de voz consiste na conversão das ondas sonoras em sinais elétricos através de um transdutor (microfone), filtragem desse sinal elétrico e conversão analógico-digital (A/D) desse sinal. Esse processo é ilustrado na figura 3.16.

A conversão das ondas acústicas em sinais elétricos é realizada através de um transdutor. As especificações técnicas dos transdutores encontrados no mercado variam de acordo com o modelo e a aplicação específica. Os microfones mais utilizados para aquisição de voz possuem basicamente as características mostradas na tabela 3.1.

Figura 3.16. Processo de aquisição do sinal de voz

Tabela 3.1 - Especificações técnicas para microfones

ESPECIFICAÇÃO	DESCRIÇÃO
Tipo	Unidirecional
Faixas de Frequências	10 a 20 KHz
Nível de Sensibilidade	-76 db a 1 KHz
Impedância	600 Ohms

O sinal de voz não é limitado em frequência. Ele possui componentes de frequência em uma faixa bastante larga, mesmo que com amplitudes muito baixas.

A filtragem do sinal analógico deve ser realizada através de um filtro passa-baixas, antes da conversão A/D (RABINER *et al*, 1978). Tal filtro, chamado filtro *anti-aliasing*, objetiva suprimir as componentes de frequência do sinal analógico superiores a metade da frequência de amostragem. Esse critério é conhecido como critério de Nyquist (PROAKIS *et al*, 1996). A conversão analógico-digital do sinal elétrico filtrado é realizada através de um amostrador.

O sinal amostrado é representado de forma digital, e pode-se então proceder a aplicação de técnicas de processamento digital de sinais (DSP). A frequência de amostragem utilizada deve ser tal que preserve a maior parte da informação contida no sinal de voz. O sinal de voz geralmente possui componentes de frequência superiores a 5KHz com amplitudes muito baixas. Assim frequências de amostragem superiores a 11KHz preservam grande parte da informação contida no sinal.

Outro fator a ser avaliado na conversão analógico digital é o número de níveis que o sinal amostrado poderá assumir.

Quanto maior for esse número, mais parecido o sinal digital será do sinal analógico original. Esses níveis são normalmente representados por uma cadeia de bits. Para sinais de voz uma resolução de 16 bits (equivalente a 65536 níveis) por amostra é apropriado.

A gravação da voz para análise foi realizada nos ambulatórios de Otorrinolaringologia e de Cirurgia de Cabeça e Pescoço, no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, da Universidade de São Paulo.

O grupo denominado “Normal” corresponde às pessoas sem qualquer disfonia, orgânica ou funcional, não-fumantes, que servem de controle nos estudos.

Para os diagnósticos (tanto pacientes quanto das pessoas normais) foi empregado a videolaringoscopia. Em alguns casos foi empregado videolaringoestroboscopia para dirimir dúvidas que ocorriam (com baixa frequência). Normalmente, mais de um especialista efetuava os exames e avaliavam a condição da laringe e das cordas vocais.

Na literatura, verifica-se o uso predominante de equipamentos analógicos, com gravação em fita magnética para posterior estudo, seja em computadores ou outro dispositivo como osciloscópios. Holmes (1975), analisando a gravação analógica em meio magnético, mesmo que o gravador seja de boa qualidade, provoca distorções de fase nas baixas frequências do sinal. Assim, as informações patológicas que a frequência fundamental podem ser alteradas, ou o sinal residual ou glotal ser descaracterizado. Sugere-se o uso de gravação em FM ou utilização direta de equipamento digital.

O processo de aquisição das vozes apoia-se no uso direto do computador, recebendo o sinal a ser digitalizado de um microfone. Com apenas uma conversão A/D, elimina-se grande quantidade de ruídos espúrios e assim, evita-se que o sinal de voz digitalizado possa apresentar características de vozes patológicas.

O microfone empregado na aquisição de voz foi um modelo LeSon SM58, unidirecional e de ganho dinâmico, permitindo que informação expúria do ambiente sobre o sinal de voz seja desprezível, juntamente com a boa resposta em frequência deste tipo de microfone.

A conversão analógico-digital foi realizada através de uma placa de aquisição de dados convencional marca SoundBlaster SB16, sendo que a taxa de amostragem utilizada foi de 22050 Hz e discretização das amostras em 16 bits.

O uso destes valores é confirmado pelos estudos de Horii (1979) e Titze, Horii *et al* (1987) para manter a margem de erro para jitter e shimmer abaixo de 0.1%. O controle automático de ganho foi definido no programa de gravação para reduzir perturbações lentas de grande amplitude na voz do paciente. Outra necessidade foi a de especificar um valor máximo para o ganho da placa de aquisição para evitar-se a saturação de amplitudes e produzir sinal ruidoso aditivo na conversão.

Trabalhos anteriores mostram divergência entre o tipo de fala empregada na gravação (sustentada ou contínua). Apesar da fala contínua ser mais próxima da realidade do paciente, as análises sobre este tipo de fonação devem ser cuidadosas, pois alterações nas articulações da supraglote modificam a densidade de potência espectral. Considerando que também que a escolha dos fonemas sustentados varia conforme a medida acústica realizada, efetuou-se a gravação dos fonemas /a/, /e/ e /i/. Com o microfone posicionado à 5

centímetros em direção da boca, solicitou-se que os pacientes pronuncia-se os fonemas (separadamente) por aproximadamente 5 segundos em média, tempo este controlado pelo operador, em níveis de intensidade e altura confortáveis para evitar tensão excessiva sobre as pregas vocais. Caso o paciente não suportasse este período de fonação, solicitava-se que o mesmo cessasse imediatamente de emitir o fonema pois verifica-se um aumento de tensão das pregas vocais para manutenção da vibração destas mediante uma menor quantidade de ar proveniente dos pulmões; isto produz uma estabilidade artificial na produção dos fonemas sustentados (GELFER, 1995). Espera-se que o fonema /i/ produza menor nível de jitter e shimmer que os demais fonemas pois as cordas vocais estão em uma condição de maior tensão que nos demais sons vocálicos.

A quantidade de tempo considerada (5 segundos) deve-se aos requisitos de convergência assintótica das perturbações - de amplitude e frequência fundamental - necessários para uma melhor avaliação patológica do paciente (TITZE *et al*, 1987).

Toda gravação é precedida de uma explicação deste processo de registro da voz ao paciente e realizado efetuado-se um teste para que este sofresse pouca influência psicológica, alterando significativamente sua voz. Algumas gravações foram descartadas e refeitas pois constatou-se diferença entre a emissão da vogal sustentada e o modo com que o paciente falava durante conversas com o mesmo antes da gravação.

Estes cuidados fazem-se necessários para evitar a produção de ruídos expúrios e assim, novas variáveis de que não se tem controle ou conhecimento sobre sua influência no sinal de voz, e assim, dificultar a análise acústica das patologias.

A cada gravação, automaticamente tomava-se nota de algumas informações do paciente. Cadastrou-se o nome completo dos pacientes, sua idade, sexo, atividade que exerce, fonema gravado e a patologia que possui, sendo esta última fornecida pelo especialista.

4.3.2. Cálculos preliminares sobre o sinal de voz

Uma vez que todo o sinal de voz apresenta diferenças em amplitude entre pacientes, algumas medidas acústicas podem ser influenciadas. Assim, é realizado um pré-processamento para normalização das amplitude entre -1 e +1. Isto é realizado por uma ponderação de todas as amostras digitalizadas em relação ao maior valor absoluto destas, que padroniza todas as medidas acústicas, desprezando-se efeitos de intensidade na voz.

Outra necessidade é definir os trechos vocálicos do sinal digitalizado. Apesar da possibilidade de desenvolver-se algoritmo específico para isto, foram eliminados os trechos não-vocálicos manualmente. Sabe-se que os sinais de voz apresentam diferenças de energia sonora entre as porções inicial e final dos trechos vocálicos, devido ao “ataque” do fluxo de ar pela glote e perda de capacidade volumétrica de ar, respectivamente. De Krom (1995) mostrou que estes intervalos no sinal de voz produzem maior perturbação em amplitude e frequência, ciclo-a-ciclo, o que pode caracterizar apenas uma condição de energia acústica do fluxo de ar e não uma evidência patológica completa. Apesar de De Krom (1995) mostrar que o intervalo inicial do trecho vocálico apresentar maior correlação com a percepção de aspereza na voz de pacientes, estes dois momentos no sinal de voz são retirados manualmente. São apagados aproximadamente 10 períodos de frequência fundamental em cada um destes intervalos. Se houvesse mais de 5 segundos de vocalização ainda restantes, o excedente ainda foi retirado para padronizar o tamanho das amostras de voz.

Verificou-se nos sinais digitalizados que a placa de aquisição utilizada (SoundBlaster SB16) produz um nível DC nos sinais digitalizados indevido, sendo necessário sua retirada. Apesar da influência sobre as medidas acústicas realizadas ser muito pequena, utilizou-se um algoritmo para retirada de tendência linear que para grande quantidade de amostras, apenas centra o valor médio das amplitudes em zero.

Assim, o pré-processamento é realizado na seguinte ordem:

1. Retirada de trecho não-vocálico, e de ataque inicial e final, do sinal de voz digitalizado
2. Retirada de tendência linear
3. Normalização em amplitude do sinal

O intuito deste pré-processamento é padronizar todos os sinais de voz para que a análise seja uniforme, evitando que os parâmetros acústicos sejam influenciados por características de gravação (mesmo com o controle realizado).

4.3.3. Medidas de Perturbação

As medidas de perturbação, ciclo-a-ciclo, da frequência fundamental e amplitude são importantes para avaliação das pregas vocais. Pesquisadores, começando por Lieberman (1963), descreveram um conjunto de medidas cuja magnitude de variação permitem diagnóstico diferencial das disfonias. Os objetivos deste tipo de parâmetro é primariamente auxiliar na discriminação de patologias que afetem a periodicidade de adução e abdução das pregas vocais. Disfonias como hipercinesia produzem um movimento irregular da mucosa que reveste as pregas vocais, apesar de organicamente as estruturas de vocalização não apresentarem moléstia. A aspereza (percepção detectada neste tipo de patologia) advém da tensão exagerada sobre a prega vocal.

Outro grupo de patologias que influenciam no movimento glotal são as que produzem voz rouca. Carcinoma, paralisia unilateral, corpectomia são casos em que as pregas vocais também apresentam periodicidade irregular, além de propiciarem o fluxo de ar turbulento pela glote.

Considerando estas características, determinou-se um conjunto de medidas de perturbação, baseado na literatura, empregando em alguns casos elementos estatísticos robustos para obter-se resultados mais precisos, que possam ser empregados na discriminação patológica.

Uma primeira necessidade é estimar o sinal glotal, ou os períodos e amplitudes ciclo-a-ciclo do sinal de voz, para que se possa determinar as perturbações propriamente ditas. Este é um fator crítico em pesquisas da área pois no caso de utilizar-se sinais de voz, a influência da estrutura supraglotal dificulta sua determinação. Objetivamente, o sinal glotal é sempre estimado. Mesmo com o emprego de eletroglotografia ou acelerômetros, os sinais produzidos não são exatamente idênticos ao produzido na laringe.

Considerando apenas o uso da voz, técnicas síncronas e assíncronas, uso de filtragem passa-baixa e medição manual dos períodos de pitch em papel milimetrado são os métodos mais empregadas na literatura.

Esta seção apresenta um algoritmo para estimação das amplitudes e períodos da fundamental utilizando o método cepstral e minimização de erro médio quadrático. Posteriormente, as medidas acústicas de perturbação, e o seu procedimento de cálculo, são apresentadas.

• Estimador síncrono para período e amplitude de F_0

O algoritmo para determinar os sucessivos períodos e amplitude da frequência fundamental realiza duas etapas distintas, sendo a primeira assíncrona e a seguinte, síncrona. A etapa assíncrona realiza uma estimação do período médio de frequência fundamental para um determinado segmento do sinal de voz, possibilitando que na segunda etapa seja realizado um refinamento desta estimativa. Isto se faz necessário pois algoritmos síncronos, ou orientados à eventos, em avaliações de laboratório, foram imprecisos para vozes com grande quantidade de oscilação. Resolveu-se então adotar uma técnica híbrida que pudesse ser robusta, independente dos requisitos computacionais.

A fase assíncrona realiza uma estimativa do pitch para um segmento de 4096 amostras (185.8 ms) utilizando o método de cepstrum (NOLL, 1967). Considerando o uso do cepstrum real, este é matematicamente definido por:

$$rc_s(n) = \text{real} \left\{ \mathcal{F}^{-1} \left\{ \log \left| \mathcal{F} \{ s(n) \} \right| \right\} \right\} \quad (14)$$

ou seja, o cepstrum real $rc_s(n)$ corresponde a porção real da transformada discreta inversa de Fourier do logaritmo do módulo da transformada discreta de Fourier do sinal de voz. O princípio é que a aplicação do logaritmo sobre um produto equivale à soma dos dois logaritmos. Com este intuito, procura-se realizar a separação do sinal de excitação (variação rápida) e a função de transferência do trato supraglotal (variação lenta), que são convoluídos para produzir o sinal de voz.

Assim, primeiramente teríamos o espectro do sinal de voz representado por:

$$S(\omega) = P(\omega) \cdot V_{VR}(\omega) \quad (15)$$

onde: $S(w)$ é a transformada discreta de Fourier do sinal de voz ;

$P(w)$ é a transformada discreta de Fourier do trem de pulsos periódico e

$V_{VR}(w)$ é a transformada discreta de Fourier do trato supraglotal e radiação dos lábios

Aplicando-se o logaritmo nesta expressão, transforma-se este produto em soma. Como a transformada de Fourier é linear, obtém-se então que o cepstrum real de tempo curto de um segmento do sinal de voz, ou seja:

$$rc_s(n; m) = \text{real}\left\{\mathfrak{T}^{-1}\left\{\log|P(\omega)|\right\}\right\} + \text{real}\left\{\mathfrak{T}^{-1}\left\{\log|V_{VR}(\omega)|\right\}\right\} \quad (16)$$

Apenas considera-se a porção real pois o logaritmo do espectro de potência é par. A análise do sinal resultante mostra que as componentes de variação lenta localizam-se nas baixas *quefrequency*, enquanto que as variações rápidas, provenientes dos harmônicos no espectro de potência do sinal de voz, ficam dispostos nas altas *quefrequency*. A linearidade (adição) obtida no cepstrum real permite que através de *liftering* (filtragem) seja possível separar as componentes harmônicas da estrutura formante.

Empregando um algoritmo de busca de pico, procura-se então o máximo pico cepstral na região de existência do período de frequência fundamental da voz, onde a abscissa deste pico corresponde ao *pitch* médio para o segmento de voz considerado. A região de busca utilizada é definida no intervalo de frequências entre 40 e 500 Hz, pois considera-se que a variação de frequência fundamental pode ser extrema para vozes ásperas e roucas.

Noll (1967) recomenda um processamento adicional para refinamento do período da fundamental encontrado. Isto porque perturbações temporais entre períodos pode propiciar que o segundo pico cepstral (múltiplo do primeiro) tenha uma magnitude superior, **produzindo erros de estimativa**. Como isto é algo comum para pessoas com disfonias, foi considerado a seguinte regra heurística:

1. Período de F_0 corresponde *quefrequency* atual (abscissa do pico máximo atual);
2. Se a *quefrequency* do pico máximo atual $\geq$ *quefrequency* do pico máximo anterior;
 - 2.1. Busca pico máximo no intervalo de ± 0.5 ms da metade da *quefrequency* do pico máximo atual;
 - 2.2. Se este novo pico máximo localizar-se entre ± 1 ms do *quefrequency* do pico máximo anterior;

2.2.1. Então Se a magnitude deste novo pico máximo for superior à $\frac{1}{4}$ da amplitude do pico máximo atual;

2.2.1.1. Período de F_0 corresponde à nova *quefrequency*;

2.2.2. Senão Se a magnitude deste novo pico máximo for superior à $\frac{1}{2}$ da amplitude do pico máximo atual;

2.2.2.1. Período de F_0 corresponde à nova *quefrequency*.

A partir desta estimativa assíncrona para o segmento considerado, é realizada a etapa síncrona, para determinar precisamente todos os períodos da fundamental neste trecho de voz analisado. Para auxiliar o processo de busca, é mantido paralelamente ao sinal de voz, uma versão filtrada do mesmo, na qual empregou-se um filtro passa-baixa FIR de ordem 100 com frequência de corte em 800 Hz.

Realiza-se uma pesquisa de todas as amplitudes mínimas (em segmentos periódicos baseado na estimativa da F_0) pois este evento é o que mais se aproxima de um efeito físico do sistema produtor de fala (adução completa das pregas vocais). Estes períodos de amplitude mínima ainda necessitam ser melhor definidos pois a discretização acarreta em perda de resolução. Para contornar este problema, empregou-se a técnica de minimização de erro médio quadrático entre amplitudes mínimas (TITZE *et al*, 1993), semelhante à metodologia de Milenkovic (1987).

Considerando que um sinal de voz sustentada pode ser definido pela soma entre uma componente periódica e outra aperiódica, ou seja:

$$s(t) = p(t) + e(t) \quad (17)$$

Considerando esta aproximação, a porção periódica pode ser definida por

$$p(t) = p(t - t_p) \quad (18)$$

onde t_p é o período de frequência fundamental para o segmento considerado.

A estimativa deste período se dá através do critério de minimização do erro médio quadrático. Para tanto, supõem-se que $e(t)$ é estacionário, com média zero e apresenta descorrelação entre períodos sucessivos. Sabendo que:

$$s(t) - s(t - t_p) = e(t) - e(t - t_p) \quad (19)$$

Assim, a função a minimizar-se é dada por:

$$mse(t_p) = \frac{1}{2} E \left\{ [e(t) - e(t - t_p)]^2 \right\} \quad (20)$$

ou

$$mse(t_p) = \frac{1}{2} E \left\{ [s(t) - s(t - t_p)]^2 \right\} \quad (21)$$

Para determinar o período de pitch que minimiza esta função, determina-se uma região de busca para que possa realizar-se a variação de t_p , reduzindo a quantidade de processamento. Titze et al (1993) sugerem que este intervalo tenha magnitude entre 5 a 15% da estimativa anterior do período de frequência fundamental, com centro definido nesta estimativa. No algoritmo implementado foi empregado 5%. Aplica-se uma interpolação parabólica usando o erro mínimo, $mse(t_p)$, calculado para determinar a melhor estimativa do período da fundamental.

Usando a interpolação parabólica diretamente sobre o sinal de voz no instante de amplitude mínima, determina-se o valor da amplitude para o presente ciclo (que será utilizado no cálculo do shimmer do sinal de voz).

Este critério empregado garante a robustez da estimativa de frequência fundamental para um ruído aditivo. Isto é útil pois muitas patologias (que fornecem percepção soprosa) se caracterizam pelo aumento de oscilações aleatórias e métodos para estimação de F_0 puramente orientados à eventos como cruzamentos por zero e coleta de picos (*peak-picking*) tem sua capacidade deteriorada nesta situação, muito comum em disfonias.

Este processo (etapa assíncrona/síncrona) é realizada sobre todo o sinal de voz padronizado, com movimentação da janela de análise. Pode-se dizer que a fase assíncrona opera como um indicador na busca dos períodos de frequência fundamental, pois o algoritmo é fortemente baseado na segunda etapa.

• PFF (*Perturbation Frequency Factor*)

Esta medida segue a definição de Lieberman (1963, 1967). Determinando as diferenças entre os períodos de frequência fundamental sucessivos, este pesquisador construiu sua curva de distribuição. Verificou que as patologias concentravam-se a partir de diferenças de período, ciclo-a-ciclo, acima de 0.5

ms. Assim, esta medida acústica é definida como a integral desta distribuição para perturbações superiores à 0.5 ms.

Uma definição equivalente é que PFF corresponde à percentagem de todas as diferenças de período da fundamental superiores à 1/2 ms.

• DPF (*Direction Perturbation Factor*)

Ignorando as variações de amplitude das diferenças entre os períodos glotais adjacentes, define-se DPF, segundo Hecker *et al* (1971), como sendo a percentagem de alterações do sinal algébrico destas diferenças em relação ao número de intervalos de pulsos glotais.

Para determiná-lo, determina-se primeiramente as diferenças entre períodos (ou amplitudes) de frequência fundamental, das quais se mantém apenas o sinal algébrico. Calculando a relação entre o número de mudanças de sinal pelo número de períodos (ou amplitudes) da fundamental, obtém-se o percentual de perturbação diferencial, ou DPF.

• JR (*Jitter Ratio*)

Por verificar-se que o aumento da frequência fundamental propicia uma redução na perturbação média (redução de jitter médio), este parâmetro foi construído para minimizar tal efeito.

A expressão que define esta medida é dada abaixo :

$$JR = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |P_i - P_{i+1}|}{\frac{1}{N} \sum_{i=1}^N P_i} \cdot 1000 \quad (22)$$

ou seja, a razão entre a perturbação média dos períodos de frequência fundamental pelo período de pitch médio. O fator aplicado (1000) na equação acima apenas evita a produção de valores de baixa magnitude.

Apesar desta medida apenas ser empregada na literatura para períodos sucessivos, esta expressão também foi utilizada considerando as amplitudes ciclo-a-ciclo, para verificar seu potencial para discriminação de patologias.

• JF (*Jitter Factor*)

De modo análogo à medida acústica JR, considerando o recíproco do período (ou da amplitude) da fundamental determina-se o índice JF, também com o intuito de produzir uma medida que não sofra influência das alterações da frequência fundamental média. Assim, tem-se a seguinte expressão:

$$JR = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |F_i - F_{i+1}|}{\frac{1}{N} \sum_{i=1}^N F_i} \cdot 100 \quad (23)$$

Novamente um fator multiplicativo é empregado para que não se produzam valores muito pequenos para o índice.

• Séries Temporais

Baseado na sugestão de Shoentgen *et al* (1995), deseja-se produzir uma medida que elimine a correlação existente entre perturbações adjacentes (tanto de período quanto amplitude).

Para determinar esta medida, determina-se os resíduos de uma estimação AR sobre as diferenças entre o i -ésimo período da fundamental pelo período de F_0 médio, ou seja, estima-se a seguinte série descorrelacionada:

$$Z_i = (P_i - \bar{P}) - a_1(P_{i-1} - \bar{P}) - \dots - a_q(P_{i-q} - \bar{P}) \quad (24)$$

onde q é a ordem do modelo, P_i é o i -ésimo período (ou amplitude) da frequência fundamental e $\bar{P}$ é a média do período (ou amplitude) da frequência fundamental.

Para determinar os coeficientes do modelo AR acima, e por conseguinte, os resíduos da estimação, utiliza-se o algoritmo de Levinson para solucionar as equações de Yule-Walker, pois este método permite que a ordem do modelo aumente a cada iteração.

Para determinar esta ordem, q , emprega-se um teste estatístico (teste χ^2) para avaliar a brancura do resíduo Z_i obtido. Neste teste, a hipótese nula é que os resíduos obtidos correspondem a um ruído branco. Em caso de rejeição, executa-se uma nova iteração do algoritmo de Levinson. Limitou-se em 20 o número máximo de iterações, pois nas análises de Schoentgen e Guchteneere

(1995) constatou-se que poucos modelos AR para estes resíduos verificam a hipótese nula para um número de coeficientes superior à 8.

Determinados os resíduos da estimativa paramétrica, aplica-se então uma média móvel (também denominado PPQ modificado) de comprimento 5 para eliminação das perturbações lentas que possam ocorrer no sinal residual, ou seja:

$$PPQ_{\text{mod}} = \frac{\frac{1}{N-5} \sum_{i=6}^N |Z_i|}{\frac{1}{N} \sum_{i=1}^N P_i} \quad (25)$$

Apesar do trabalho desses pesquisadores apenas citar o uso para perturbações em frequência, considerou-se coerente aplicá-lo sobre as amplitudes da fundamental pois certamente estas são correlacionadas com os valores vizinhos.

As medidas para este algoritmo são realizadas à cada 256 amostras sobre um conjunto de 512 estimativas de período (ou amplitude) de F_0 , sendo que determina-se ao final a mediana destes valores como índice representativo do sinal de voz.

• AVI/PVI (*Amplitude/Perturbation Variation Index*)

Baseado em estatística descritiva, aplica-se o coeficiente de variação dos períodos, ou amplitudes, da frequência fundamental, sobre todas as amostras padronizadas do sinal de voz. O coeficiente de variação é definido por:

$$CV = \frac{\frac{1}{N} \sum_{i=1}^N (P_i - \bar{P})^2}{\bar{P}^2} \quad (26)$$

onde P_i é o i -ésimo período (ou amplitude) da frequência fundamental,

$\bar{P}$ é a média do período (ou amplitude) da frequência fundamental.

O parâmetro acústico PVI corresponde ao coeficiente de variação aplicado sobre os períodos de frequência fundamental multiplicado por 1000 para que não se tenha valores muito pequenos, enquanto que a medida AVI, além de resultar do produto entre o coeficiente de variação pela mesma constante sobre

as amplitudes da fundamental, aplica-se o logaritmo decimal, pois existe a indicação de alto grau de linearidade positiva entre o nível de percepção acústica de aspereza e a aplicação da escala logarítmica na medida AVI.

• Médias Móveis

Na literatura estas medidas são encontradas sobre nomes como RAP (*Relative Average Perturbation*), PPQ (*Period Perturbation Quotient*) ou APQ (*Amplitude Perturbation Quotient*). O princípio que rege a aplicação destas medidas sobre as estimativas tanto de período quanto de amplitude da frequência fundamental é a existência de variações lentas, sem relação com as disfonias, que ocorrem no sinal de voz. Estas oscilações afetam medidas de jitter, elevando seu valor, induzindo a identificação errônea de patologias da laringe.

A expressão genérica para médias móveis é:

$$MM = \frac{\frac{1}{N-P+1} \sum_{i=1}^{N-P+1} \left| \frac{1}{P} \sum_{j=1}^P (X_{i+j-1}) - X_{i+0.5(P-1)} \right|}{\frac{1}{N} \sum_{i=1}^N X_i} \quad (27)$$

onde N é a quantidade de períodos/amplitudes usados no cálculo

P é o tamanho da janela para média móvel

X_i é a i -ésima medida de período/amplitude, ciclo-a-ciclo

P é a ordem, ou comprimento da janela de dados.

No cálculo deste parâmetro foi considerado para mensuração das perturbações de F_0 , janelas de comprimento 5 e 11, condizente com o trabalho de Davis (1979). Para perturbações em amplitude, utilizou-se um comprimento de 11 amostras.

• Perturbações de ordem superior

Classificou-se um conjunto de medidas que se aplicam sobre perturbações de ordem superior (tanto em período quanto amplitude da frequência fundamental), ou seja, sobre derivadas discretas com ordem variando entre 1 à

4, sendo que a perturbação de ordem zero é o desvio entre as amostras e a média destas amostras. Sobre este conjunto de 10 subsinais (5 para períodos e 5 para amplitudes de F_0) aplica-se um conjunto de medidas estatísticas, para avaliação da capacidade destas medidas em discriminar patologias.

Matematicamente tem-se que a perturbação de ordem zero é dada pela seguinte expressão:

$$p_i^0 = P_i - \bar{P} \quad (28)$$

onde P_i é o i -ésimo período (ou amplitude) da frequência fundamental,

$\bar{P}$ é a média do período (ou amplitude) da frequência fundamental.

A k -ésima perturbação é definida por:

$$p_i^k = \begin{cases} p_i^{k-1} - p_{i-1}^{k-1}, & k \text{ par} \\ p_{i+1}^{k-1} - p_i^{k-1}, & k \text{ ímpar} \end{cases} \quad (29)$$

Um conjunto de medidas estatísticas (clássicas e robustas) foi utilizado para caracterizar este “sinal” de perturbações de período e amplitude da fundamental. Como medidas de tendência central, utilizou-se a média e mediana das perturbações de ordem k .

Para avaliar a dispersão destas perturbações, utilizou-se desvio absoluto, desvio padrão, assimetria e curtose, para se avaliar o comportamento da distribuição. Finalmente, utilizou-se regressão linear para determinar o comportamento das amostras de perturbação através de estimação clássica (por mínimos quadrados) e robusta (por minimização do desvio absoluto do valor mediano). Nesta última abordagem são considerados os coeficientes da regressão linear e o desvio quadrático das regressões.

A medida de assimetria, baseada no momento de terceira ordem, permite avaliar a tendência lateral da curva de distribuição das perturbações, sendo descrita pela seguinte equação:

$$assimetria = \frac{\sum_{i=1}^N (p_i^k)^3}{N \cdot \sigma_{p^k}^3} \quad (30)$$

Determinando o grau de achatamento da distribuição das perturbações, utiliza-se a medida de curtose, definida pela expressão:

$$curtose = \frac{\sum_{i=1}^N (p_i^k)^4}{N \cdot \sigma_{p^k}^4} - 3 \quad (31)$$

Esta medida é semelhante ao parâmetro EX definido por Davis (1979), onde constatou-se que as disfonias são caracterizadas por curvas de distribuição platicúrticas (curtose maior que 0), enquanto que as curvas leptocúrticas são encontradas em pessoas normais.

A regressão linear clássica foi definida através do critério de mínimos quadrados, ou seja, minimiza-se o somatório do quadrado das diferenças, ou seja

$$S(a,b) = \sum_{i=1}^N (p_i^k - a - b \cdot i)^2 \quad (32)$$

Os coeficientes a e b , a partir são determinados por

$$b = \frac{N \cdot \left(\sum_{i=1}^N i \cdot p_i^k \right) - \left(\sum_{i=1}^N i \right) \cdot \left(\sum_{i=1}^N p_i^k \right)}{N \cdot \left(\sum_{i=1}^N i^2 \right) - \left(\sum_{i=1}^N i \right)^2} \quad (33)$$

e

$$b = \frac{\left(\sum_{i=1}^N p_i^k \right) \cdot \left(\sum_{i=1}^N i^2 \right) - \left(\sum_{i=1}^N i \cdot p_i^k \right) \cdot \left(\sum_{i=1}^N i \right)}{N \cdot \left(\sum_{i=1}^N i^2 \right) - \left(\sum_{i=1}^N i \right)^2} \quad (34)$$

O desvio da regressão linear é finalmente calculado como o desvio padrão dos resíduos desta estimação linear. O desenvolvimento destas medidas de regressão é verificar se as perturbações apresentam elevação ou redução linear significativa durante a fonação. Além disto, através do desvio de regressão, deseja-se determinar a dispersão das perturbações, eliminando a influência da inclinação linear que estes “subsinais” possam apresentar.

Para suportar oscilações rápidas aleatórias que possam ocorrer nas perturbações de ordem superior, utilizou-se de uma regressão linear robusta, menos susceptível à ruídos expúrios. O princípio da técnica é minimizar a expressão:

$$S(a,b) = \sum_{i=1}^N |p_i^k - a - b \cdot i| \quad (35)$$

Utilizando a simplificação em que se minimiza uma soma absoluta qualquer empregando a mediana (x_M) das amostras ou seja:

$$\sum_i |x_i - x_M| \quad (36)$$

Fixando inicialmente o coeficiente b (estimativa), o coeficiente a pode ser obtido através da expressão:

$$a = \text{mediana}\{p_i^k - b \cdot i\} \quad (37)$$

Uma vez determinado o coeficiente a , pode-se finalmente determinar o novo coeficiente b através da equação:

$$\sum_{i=1}^N i \cdot \text{sinal}\{p_i^k - a - b \cdot i\} = 0 \quad (38)$$

onde a função *sinal* é -1 para valores negativos, +1 para valores positivos e *sinal*(0)=0.

O objetivo de se minimizar a discrepância entre cada perturbação e a sua mediana é obter uma equação de reta linear mais representativa. Para esta determinação, utiliza-se de algoritmos iterativos e numéricos para obter-se os coeficientes a e b . Empregou-se o método da bisseção para obter estes valores, sendo utilizada uma estimativa inicial empregando o método dos mínimos quadrados. De forma análoga à regressão linear por mínimos quadrados, o desvio da regressão robusta é obtida pelo desvio padrão dos resíduos desta estimação.

A partir deste contingente de medidas estatísticas, pretende-se determinar a capacidades destas na discriminação patológica. A grande quantidade de medidas sobre os sinais de perturbação de ordem superior visa caracterizar adequadamente suas amostras e verificar o comportamento destas para cada disфонia.

4.3.4. Medidas de Ruído

A presença de ruído no sinal de voz é um grande indício de anormalidades nas pregas vocais. Patologias como pólipos, nódulos e cistos impedem a correta adução das pregas vocais, resultando em escape de ar que torna-se turbulento e produz a sensação acústica de soprosidade. Quando ocorrem perturbações na frequência fundamental (seja em magnitude ou nos períodos sucessivos), como em carcinomas, a voz torna-se rouca.

Baseado nestas considerações, é importante examinar a curva de densidade espectral de potência à procura de oscilações aleatórias entre as componentes harmônicas do sinal de voz, verificar o comportamento da envoltória espectral, relacionando a perda energética das formantes com o aumento da magnitude nas regiões espectrais não-formantes. Além disto, torna-se imprescindível qualificar estas informações, além da própria quantidade de ruído de alta frequência. Para isto, algumas medidas acústicas foram utilizadas para fornecer uma quantificação da turbulência glotal, sendo extraídas do trecho normalizado do sinal de voz.

• HPFR

Para avaliar o ruído de alta frequência, empregou-se a estimativa HPFR sugerida por Shoji *et al* (1992). Determina-se a partir do sinal de voz normalizado a média das curvas de densidade de potência espectral em escala logarítmica (contendo 1024 pontos cada uma), com deslocamento de 25,6 ms entre cada estimativa. Por fim, aplica-se a seguinte expressão, que corresponde a média das magnitudes espectrais em determinadas faixas de frequência.

$$HPFR_i = \sum_{\omega=fl_i}^{fh_i} \overline{S(\omega)} \quad (39)$$

onde fl_i e fh_i indicam o intervalo de; estimação da energia média espectral.

Utiliza-se uma janela de Blackman para reduzir o espalhamento espectral sobre cada trecho avaliado. Os intervalos considerados na análise são 60 a 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz e a energia acima de 6000 Hz até o limite de Nyquist (como empregou-se uma frequência de amostragem de 22050 Hz, o patamar máximo considerado é de 11025 Hz).

Uma vez determinadas estas medidas (HFPR) empregou-se medidas estatísticas (clássicas e robustas) para avaliar o comportamento da distribuição dos valores ao longo do sinal de voz. Utilizou-se a média, mediana, desvio absoluto, desvio padrão, índices de assimetria e curtose, coeficientes da regressão linear obtida por mínimos quadrados, coeficientes de regressão linear robusta, e os desvios médio quadrático de ambas as regressões. A descrição destes índices estatísticos pode ser encontrada na seção 3.10.9.

• Medidas de Energia Espectral

Com o objetivo de avaliar o comportamento da energia espectral das disfonias, determinou-se para cada trecho do sinal de voz normalizado contendo 1024 amostras, com deslocamento temporal, de 20 ms, a curva de densidade de potência espectral em escala logarítmica (utilizou-se uma janela de Hamming para redução do lóbulos laterais). Sobre este segmento espectral, determinou-se a magnitude média em determinadas faixas de frequência, a saber, 60 a 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz.

Uma vez calculado todas estas quatro medidas espectrais para a totalidade das amostras do sinal de voz, aplicou-se técnicas estatísticas para avaliar o comportamento da distribuição destas medidas, ou seja, média, mediana, desvio absoluto, desvio padrão, índices de assimetria e curtose, coeficientes da regressão linear obtida por mínimos quadrados, coeficientes de regressão linear robusta, e os desvios médio quadrático de ambas as regressões. Este grupo de índices estatísticos será, empregado em outros parâmetros acústicos, sendo totalmente descritos na seção 3.10.9.

• Br

Para quantificar o ruído turbulento em patologias com características perceptuais de sopro, foi sugerido por Fukazawa *et al* (1988) a medida Br (*Breathiness Ratio*), fornecendo avaliação deste ruído em relação às demais componentes.

No procedimento de cálculo, primeiramente aplica-se um filtro de pré-ênfase, utilizando ganho de +6dB/oitava a partir da frequência nula até 10 kHz.

Considerando que o ruído turbulento caracteriza-se por componentes aleatórias de alta frequência, acompanhado de redução harmônica, determina-se o índice acústico sobre o sinal filtrado pela seguinte equação

$$Br = \frac{\sum_{i=1}^N [s(t) - 2s(t_{i-1}) + s(t_{i-2})]^2}{\sum_{i=1}^N [s(t_i)]^2} \quad (40)$$

onde N é a quantidade de amostras do sinal de voz normalizado.

Apesar dos autores citados desenvolverem o cálculo apenas sobre um trecho de 25,6 ms, este índice acústico foi determinado sobre toda a extensão do sinal normalizado. Comparando-se os resultados com patologias similares às utilizadas por estes pesquisadores, não houve diferença estatística significativa.

• SNL

Para diferenciar vozes normais das que apresentam características ásperas, Sansone *et al* (1970), Lively *et al* (1970) e Emanuel *et al* (1973) desenvolveram a medida SNL (*Signal to noise level*) para quantificar esta característica perceptual. O procedimento de cálculo era realizado sobre curvas de densidade de potência espectral impressas em papel milimetrado.

Baseado nas descrições deste método, determinou-se a potência espectral para cada trecho de 4096 pontos, com deslocamento de 25,6 ms entre segmentos, sendo previamente aplicado uma janela de Hamming. Em cada uma destas curvas de energia, determinou-se o valor mínimo de magnitude à cada 100 Hz, dentro de um intervalo de frequências (100 à 2600 Hz, 2600 à 5100 Hz, 5100 à 8000 Hz, 100 à 5100 Hz e 100 à 8000 Hz).

Ao final, obtém-se a média destes valores mínimos, ou seja, valor parcial de SNL. Com todos os valores parciais extraídos das janelas do sinal de voz, aplica-se as medidas estatísticas descritas na seção 3.10.9.

• HNR

Vozes roucas são caracterizadas por componentes aleatórias na região da frequência formante, aparecimento de ruído de alta frequência e perda de

magnitude nas frequências harmônicas. Assim, para quantificar estes elementos do sinal de voz patológico, foi desenvolvido o índice acústico HNR (*Harmonic to noise ratio*) por Yumoto *et al* (1982).

O procedimento para obtê-lo exige a construção de um algoritmo para determinar os trechos periódicos de frequência fundamental sincronamente, pois é necessário extrair um conjunto de segmentos para se estimar esta medida acústica.

O algoritmo utilizado emprega duas fases distintas na estimação dos períodos de *pitch*: uma assíncrona e outra síncrona.

A primeira fase corresponde à determinação da F_0 média para um conjunto de 4096 amostras. Isto é realizado através da técnica de estimativa cepstral (NOLL, 1967).

A fase posterior utiliza o valor médio da frequência fundamental para buscar os cruzamentos por zero que definem o início e fim de um período da fundamental.

Primeiramente determina-se a amplitude mínima dentro de um intervalo de existência e a partir deste, encontra-se o próximo cruzamento por zero, à direita.

Com todas as estimativas dos cruzamentos de zero separados por um período da fundamental, pode-se determinar as formas de onda da fundamental, ciclo-a-ciclo.

No procedimento de cálculo da medida, é necessário determinar a média para um conjunto de 50 formas de onda periódicas, ou seja:

$$S(\tau) = \sum_{i=1}^{50} \frac{S_i(\tau)}{50} \quad (41)$$

onde τ corresponde ao máximo período da frequência fundamental, sendo τ_i O período do i -ésimo segmento periódico analisado (para os trechos com período $\tau_i < \tau$, considera-se $S_i(t) = 0$ para $\tau_i < t < \tau$).

Com a forma de onda média (S_m), que no domínio da frequência corresponde à um sinal imune de ruído, obtém-se a energia das componentes harmônicas através da expressão:

$$H = n \sum_{i=1}^r s_m^2(i) \quad (42)$$

A determinação da energia do ruído do sinal analisado baseia-se na diferença entre cada segmento $f_i(\tau)$ e forma de onda média, ou seja,

$$N = n \sum_{i=1}^r \sum_{t=1}^{r_i} [s_i(t) - s_m(t)]^2 \quad (43)$$

Assim, estima-se para um segmento do sinal de voz com tamanho igual à 50 períodos de frequência fundamental uma estimativa da relação entre a parte harmônica e a ruidosa. Para evitar suposições sobre qual o trecho mais estável do sinal de voz (onde se extrairá este índice acústico), utilizou-se de sobreposições de 10 períodos da fundamental, até a totalidade do sinal normalizado. A partir destas medidas ao longo do sinal de voz, aplicou-se as medidas estatísticas da seção 3.10.9 para avaliar o seu comportamento.

• Hr

Outra avaliação da rouquidão na fala sustentada é o Hr (*Harmonic ratio*), apresentado por Hiraoka *et al* (1984). Esta medida acústica baseia-se na densidade de potência espectral para avaliar a relação entre as componentes harmônicas e as não harmônicas.

Primeiramente aplica-se um filtro FIR passa-baixa de ordem 100 com frequência de corte em 8 kHz sobre o sinal de voz. Calcula-se então, à cada 25,6 ms, um periodograma de 4096 pontos (com aplicação de uma janela de Hamming) sobre o qual será determinada a relação harmônica.

Uma modificação é realizada no procedimento apresentado pelos autores. Utilizou-se do método de estimação de frequência fundamental por cepstrum (NOLL, 1967) ao invés de determiná-la a partir do periodograma (conforme citado pelos autores). A partir da F_0 pode-se determinar as componentes harmônicas por algoritmo de busca na região de existência destas, na curva de densidade de potência espectral. Esta região é definida em um intervalo de: $J:1/3$ do período da fundamental, centrada em intervalos múltiplos de F_0 .

A medida de relação harmônica é então definida por :

$$H_r = \left(\frac{\sum_{i=1}^P h_i}{P} \right) \quad (44)$$

onde P corresponde ao somatório dos máximos, harmônicos, na curva da densidade de potência espectral;

h_i é a magnitude das componentes harmônicas.

Os somatórios da equação 57 são desenvolvidos até a frequência de 10 kHz. A partir das estimativas da relação harmônica para cada segmento de 4096 pontos, aplica-se as medidas estatísticas apresentadas na seção 3.10.9, de modo a analisar o comportamento da distribuição destes índices.

• Ra

Para determinar a quantidade de ruído no sinal de voz, Kojima *et al* (1980) consideraram a utilização um trecho de tamanho igual à três períodos da fundamental do sinal de voz, e avaliaram a densidade espectral de potência do sinal resultante da repetição deste trecho.

A estimativa dos períodos da frequência fundamental foi feita através do algoritmo híbrido de duas fases (assíncrona e síncrona) descrito na seção 3.11.5.

Novamente se emprega o método cepstrum de Noll (1967) para determinar o período de F_0 médio. A partir destas estimativas (marcações dos cruzamentos de zero), construi-se um sinal com 4096 pontos empregando repetições de três períodos sucessivos da fundamental. O cálculo do índice acústico é realizado sobre o periodograma deste novo sinal, onde se estima o somatório quadrático das componentes harmônicas (H_i) e o somatório quadrático dos pontos de máximo na região entre as harmônicas (N_i). A determinação das componentes harmônicas sobre a curva de densidade espectral é feita através de busca na sua região de existência ($\pm 1/3$ do período da fundamental, centrada em intervalos múltiplos de F_0). Assim, à cada 325 ms do sinal de voz, determina-se uma avaliação parcial da medida, ou seja,

$$Ra = 10 \log \left(\frac{\sum_{i=1}^N H_i}{\sum_{i=1}^N N_i} \right) \quad (45)$$

onde N é o número de trechos de comprimento igual à três períodos da fundamental. Aplica-se este procedimento de estimativa sobre toda extensão do sinal de voz normalizado.

Também foi testado o efeito de elevação de três para seis períodos na composição do sinal analisado espectralmente. Emprega-se então as medidas estatísticas definidas na seção 3.10.9 para determinar o comportamento da distribuição deste índice no sinal de voz.

• NNE

Kasuya *et al* (1986) propõem estimar a quantidade de ruído, baseada nos intervalos entre as magnitudes harmônicas. Considerando que trechos do sinal de voz (utiliza-se janela de Hamming para reduzir os lóbulos espectrais laterais) são compostos por uma componente periódica e outra de ruído aditivo. Assim, no domínio da frequência, tem-se

$$S_m(\omega) = H_m(\omega) + N_m(\omega), \quad \omega = 0, 1, \dots, N-1 \quad (46)$$

onde m é o índice da janela de dados analisada $S_m(\omega)$ é a componente harmônica; $N_m(\omega)$ é a componente ruidosa e N é a quantidade de pontos da transformada discreta de Fourier.

Para avaliar a relação de energia entre as componentes do sinal de voz, utiliza-se a seguinte equação, correspondente ao índice acústico propriamente dito

$$NNE = 10 \log \frac{\frac{1}{L} \sum_{\omega=1}^{N_H} \sum_{m=1}^L |\hat{N}_m(\omega)|^2}{\frac{1}{L} \sum_{\omega=N}^{N_H} \sum_{m=1}^L |S_m(\omega)|^2} \quad (47)$$

onde $\hat{N}_m(\omega)$ é a estimativa espectral do ruído; $S_m(\omega)$ é a densidade de potência espectral do sinal de voz; L é a quantidade de janelas de dados que são utilizadas para análise; N_L e N_H são os limites inferiores e superiores, em amostras, sobre o qual a medida é desenvolvida.

Para estimar o ruído, considera-se que entre os picos harmônicos, a única componente existente é ruidosa. Na região do pico, a existência das duas componentes, não é possível separar-se $N_m(\omega)$ de $S_m(\omega)$. Assim, utiliza-se do

valor médio dos dois vales laterais de cada pico harmônico como sendo a componente ruidosa naquelas regiões, ou seja,

$$|\hat{N}_m(\omega)|^2 = \begin{cases} |S_m(\omega)|^2 & \omega \in D_j \\ \frac{\sum_{k \in D_j} |S_m(k)|^2}{2 N_j} + \frac{\sum_{k \in D_{j+1}} |S_m(k)|^2}{2 N_{j+1}} & \omega \in P_i \end{cases} \quad (48)$$

onde j indica uma região contendo um vale e um pico harmônico.

As regiões D_j e P_j são definidas com base na largura de banda da janela espectral (no caso, Hamming). Assim, estas são definidas por :

$$\begin{aligned} P_j &= \left\{ j \mid j_i - \frac{2N}{M} \leq j \leq j_i + \frac{2N}{M} \right. \\ D_j &= \left. \left\{ j \mid j_{i-1} + \frac{2N}{M} \leq j \leq j_i - \frac{2N}{M} \right. \right. \end{aligned} \quad (49)$$

onde M é a quantidade de amostras da m -ésima janela analisada.

O algoritmo desenvolvido primeiramente determina o período médio de frequência fundamental através do método de autocorrelação. Extrai-se então um trecho de tamanho $7 T_p$ (período médio de F_0) e aplica-se a transformada discreta de Fourier com 1024 pontos.

Sobre este resultado então, estima-se a posição de cada pico harmônico através de busca centrada nas frequências múltiplas da fundamental, permitindo então o cálculo de P_j e D_j .

Se houver sobreposição entre larguras de bandas sucessivas, estas são desconsideradas na estimação da medida acústica.

Finalmente, aplica-se a eq. (47) para estimativa da medida NNE em determinados intervalos N_l e N_h amostrais na transformada discreta de Fourier.

Os intervalos considerados (em termos de frequência) são 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz e 1000 à 5000 Hz.

• N/S

Outra medida empregada na avaliação de rouquidão foi sugerida por Muta *et al* (1988). Seu cálculo baseia-se em estimativa síncrona da frequência fundamental e posterior avaliação da relação sinal ruído. O procedimento é realizado em três etapas distintas. A primeira etapa corresponde à extração da frequência fundamental do sinal.

Utilizando uma estimativa do período da fundamental (F_0) baseada no método cepstral (NOLL, 1967), extrai-se um segmento com tamanho igual à 4 vezes àquele período estimado. Aplica-se uma janela de Hamming sobre o mesmo e determina-se a autocorrelação deste. Compara-se então a estimativa F_0 com a obtida da autocorrelação (F) e em caso de discrepância, considera-se $F_0=F$, executando novamente esta etapa (desconsiderando apenas a fase cepstral) até que se obtenha um trecho contendo exatamente quatro períodos da fundamental. Como havia sinais muito ruidosos, definiu-se um número máximo de tentativas igual à 20.

A próxima etapa corresponde à análise síncrona *do pitch*. Considerando o sinal janelado da fase anterior, $s(n)$, determina-se a curva de densidade espectral de potência através de:

$$H(\omega) = \left| \sum_{n=1}^{M_R} s(n) e^{-j \frac{2\pi \omega n}{M_R}} \right|^2 \quad (50)$$

onde M_R é igual à $4F$.

Finalmente, avalia-se a relação sinal-ruído (NS) através da expressão :

$$NS = 10 \log \frac{\sum_{\omega=3}^{4L+2} \min_i H(4\omega + i)}{\sum_{\omega=3}^{4L+2} H(\omega)} \quad i = -1, 0, 1, 2 \quad (51)$$

onde L é o número de harmônicos empregados no cálculo (no caso, $L = 16$).

Antecedendo o método apresentado, aplica-se um filtro FIR passa-baixa de ordem 100, com frequência de corte em 4800 Hz. A cada estimativa da medida acústica, desloca-se a janela de análise em 25,6 ms. Ao conjunto de

índices determinados sobre o sinal de voz, aplica-se os cálculos estatísticos apresentados na seção 3.10.9.

• HNR Cepstral

Uma interessante técnica de avaliação do ruído no sinal de voz é através de *liftering* do cepstrum real. Krom (1993) especifica um método que foi empregado neste trabalho. Sabe-se que as componentes de baixa oscilação do sinal de voz (formantes) apresentam correspondência nos baixos *quefreny* do cepstrum, enquanto os harmônicos são evidenciados nos elevados *quefreny*. Como o ruído proveniente de ar turbulento nas cordas vocais tem banda larga, pode-se empregar um processo adaptativo de *liftering* para remover os picos em *quefreny* múltiplos da frequência fundamental, mantendo-se apenas a informação da estrutura formante e o ruído de turbulência. A medida acústica é então definida por :

$$\text{HNRCeps}_{[f_i, f_h]} = 10 \log_{10} \left[\frac{\sum_{\omega=f_i}^{f_h} S^2(\omega)}{f_h - f_i + 1} \right] - 10 \log_{10} \left[\frac{\sum_{\omega=f_i}^{f_h} N^2(\omega)}{f_h - f_i + 1} \right] \quad (52)$$

onde f_i e f_h correspondem à frequência digital superior e inferior da banda espectral de análise ; $S(\omega)$ é a curva de densidade de potência espectral do trecho analisado do sinal de voz ; $N(\omega)$ é a curva de densidade de potência espectral do ruído existente do trecho analisado.

A estimativa do ruído no domínio cepstral é desenvolvida da seguinte forma: determina-se o cepstrum real ($C(\tau)$) em uma janela com comprimento de 4096 amostras, de onde estima-se o período da frequência fundamental pelo método de Noll (1967).

Filtra-se adaptativamente este cepstrum real, anulando todas as magnitudes na região de existência dos picos cepstrais múltiplos da frequência fundamental.

Mais especificamente obtém-se :

$$C_{\text{liftering}}(\tau) = \begin{cases} 0, n.T - 0,5.b \leq \tau \leq n.T + 0,5.b \\ C(\tau) \text{ c.c.} \end{cases} \quad (53)$$

onde n corresponde ao n -ésimo pico cepstral analisado ($n > 1$); b é a largura de banda dos picos cepstrais; T é o período da frequência fundamental.

A largura de banda é definida em cada pico periódico do cepstrum real nos pontos de alteração do sinal da derivada primeira desta curva, tanto à esquerda quanto à direita deste pico. Determina-se primeiramente uma aproximação da: densidade de potência espectral do ruído através de :

$$N_{ap}(\omega) = \Im\{C_{\text{lifering}}(\tau)\} \quad (54)$$

É necessário realizar uma compensação da linha de base do espectro do ruído estimado pois a estrutura harmônica do sinal de voz deve repousar sobre a abscissa (0 dB) na curva de densidade de potência espectral. Assim, a compensação pode ser definida por :

$$Bd(\omega) = \left| \min[S(\omega) - N_{ap}(\omega)] \right| \quad (55)$$

Assim, a estimativa final da densidade espectral do ruído é dado por :

$$N(\omega) = N_{ap}(\omega) - Bd(\omega) \quad (56)$$

A medida apresentada na eq. (56) é realizada sobre segmentos de 4096 amostras, à cada 25,6 ms, produzindo-se uma grande quantidade de amostras que são avaliadas através dos métodos estatísticos da seção 3.10.9. Os intervalos de frequência onde os cálculos foram realizados são 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz e 60 à 2000 Hz.

• Primeiro e segundo harmônicos

As medidas de componentes harmônicas foram também utilizadas pois vozes soprosas apresentam modificações bem claras de magnitude nestas frequências. Para sua determinação empregou-se um algoritmo de estimação de frequência fundamental por cepstrum (NOLL, 1967).

A partir desta medida, realizou-se uma busca centrada em ambos os picos harmônicos (frequências múltiplas de F_0) na curva de densidade de potência espectral de 4096 pontos, sendo seus valores fornecidos em dB. A

estimativa é realizada à cada 25.6 ms e sobre os valores obtidos aplica-se as avaliações estatísticas apresentadas na seção 3.10.9.

• CPP

A amplitude do pico cepstral pode ser útil na avaliação de vozes soprosas. A medida CPP (*Cepstral peak proeminence*) apresentada por Hillenbrand *et al* (1994) corresponde à diferença entre a amplitude do pico cepstral e o valor correspondente na regressão linear do trecho analisado (este trecho inicia-se a partir de 1 ms da curva de cepstrum real).

Estas estimativas são realizadas sobre janelas do sinal de voz de comprimento igual à 4096 amostras, com deslocamento de 25,6 ms.

• P/A

Esta medida acústica corresponde à razão entre a máxima amplitude no segmento de voz analisado pela amplitude média da retificação em onda completa deste mesmo trecho. Esta medida (*Peak to amplitude*) foi sugerida por Hillenbrand *et al* (1994) para ser aplicada entre segmentos sucessivos de 10 msec.

Neste trabalho, considerou-se uma janela de 4096 pontos, com deslocamento de 25,6 ms. Ao final das análises sobre todos os trechos do sinal de voz, são aplicadas as análises estatísticas da seção 3.10.9.

• RPK

Baseado no conceito da autocorrelação, Hillenbrand *et al* (1994) sugerem a medida RPK (*Pearson r at autocorrelation peak*). O método de cálculo inicia pela estimação do período de frequência fundamental, obtido no trabalho através do método de cepstrum (NOLL, 1967).

Utiliza-se então o trecho do sinal de voz analisado e sua versão deslocada um período da fundamental no cálculo do coeficiente de Pearson, ou seja,

$$r_p = \frac{\sum_{i=1}^{N-T} \{(s_i - \bar{s})(s_{i+T} - \bar{s})\}}{\sqrt{\sum_{i=1}^{N-T} (s_i - \bar{s})^2} \sqrt{\sum_{i=1}^{N-T} (s_{i+T} - \bar{s})^2}} \quad (57)$$

onde T corresponde ao período da fundamental estimado ; $\tilde{S}$ e $\tilde{S}_T$ correspondem respectivamente à média dos trechos analisado e deslocado um período de F_0 .

Os trechos analisados contém 4096 amostras, com deslocamento temporal de 25,6 ms. Aplica-se ao final do cálculo de todos os trechos disponíveis no sinal de voz, as análises estatísticas da seção 3.10.9, para avaliação do comportamento da distribuição dos valores obtidos.

4.3.5 Medidas no Sinal Residual

A terceira técnica para avaliação do trato vocal é através da análise do sinal residual (Davis, 1979). O fundamento desta metodologia está estimação do sinal glotal através do modelo inverso do trato vocal: assim, aplica-se o sinal de voz e obtém-se o sinal proveniente da glote.

O sinal residual, apesar de não ser o sinal glotal, corresponde ao resultado de uma filtragem inversa sobre o sinal de voz, onde o filtro construído é determinado considerando o modelo paramétrico do trato vocal. Este sinal corresponde ao sinal de excitação para produção de voz sintetizada.

Pesquisadores (KOIKE *et al*, 1975; DAVIS, 1979; PROSEK *et al*, 1987) verificaram que características patológicas podem ser determinadas através do sinal residual, pois conforme a disfonia e sua extensão, são produzidas alterações temporais e espectrais neste sinal. Assim, desenvolveram um conjunto de medidas acústicas que permitam relacionar a voz com as disfonias.

Este seção apresenta uma adaptação da técnica apresentada por esses pesquisadores, baseada na obtenção do sinal residual através de algoritmos adaptativos de modo a suportar as variações no tempo (mesmo que pequenas) do trato vocal.

• Estimação Paramétrica na Análise de Voz

A transcrição do trato vocal em equações, realizada no capítulo 2, fundamenta a análise por estimativas paramétricas do sinal de voz. Assim, um modelo linear para a produção da fala é dado por:

$$H(z) = R(z) \cdot V(z) \cdot G(z) \cdot E(z) \quad (58)$$

onde estão representados a excitação produzida pelos pulmões (principalmente), por $E(z)$; a vocalização na laringe e cordas vocais, por $G(z)$, e

assim, gerando o pulso glotal; a modulação das componente harmônicas do sinal glotal e assim, caracterização da estrutura formante pelo trato supraglotal, por $V(z)$; e os efeitos da radiação dos lábios (e outras perdas associadas à influências térmicas, por rugosidade e alteração das paredes do trato vocal) por $R(z)$.

Este modelo é composto então tanto de zeros quanto pólos, uma vez que a influência dos zeros capacita a reprodução de sinais vocálicos nasalados (pelo acoplamento do trato nasal através da movimentação do véu palatino). Desta feita, seria necessário utilizar-se de modelos ARMA para descrever corretamente este sistema.

Entretanto, considerando que um modelo AR com um número infinito de pólos pode ser empregado para estimar espectralmente qualquer sistema, pelo teorema de Kolmogorov, (na prática, isto é impossível, pois os requisitos computacionais e a quantidade de amostras da série temporal são irreais), alguns pólos adicionais à um modelo AR permite a este representar o sistema de produção de fala, corrigindo os erros advindos da falta de zeros (responsáveis pela atenuação as componentes espectrais do sinal de voz).

Assim, obtém-se um modelo AR a ser estimado que representa o espectro densidade de potências para um trecho de amostras, considerando-as um processo aleatório WSS:

$$H(z) = S(z) = \frac{\sigma}{\sum_{i=0}^P a_i z^{-i}} E(z) \quad (59)$$

onde P expressa a quantidade de pólos do sistema. Considerando que no modelo inicial ter-se-ia K formantes, com frequências centrais abaixo de 5000 Hz e comprimento de banda geralmente menor que 150 Hz, utiliza-se as três primeiras formantes, que são mais energéticas no espectro densidade de potência. Assim, P é expresso por:

$$P = 2K + m \quad (60)$$

uma vez que os pólos correspondentes aos formantes são conjugados, e o acréscimo de m pólos para correção dos erros devido à falta de zeros ($0 < m < 7$).

Para obtenção dos parâmetros a_i e σ , admite-se que o sinal de voz $s(n)$ é um processo AR com P pólos à serem determinados, e desta forma, o seu preditor é obtido por uma combinação linear de $P-1$ amostras passadas:

$$\hat{s} = -\sum_{i=1}^P a_i s[n-i] \quad (61)$$

Deseja-se então obter a_i 's que minimizem o erro médio quadrático entre a amostra predita e a amostra real, ou seja:

$$\rho = E[|s[n] - \hat{s}[n]|^2] \quad (62)$$

Através do princípio de ortogonalidade, em que o erro de estimativa é ortogonal à estimativa, obtém-se um sistema de equações não lineares, ou seja:

$$\begin{aligned} r_{ss}[k] &= -\sum_{i=1}^P a_i r_{ss}[k-i] & k = 1, 2, \dots, P \\ \sigma^2 &= \rho_{\min} = r_{ss}[0] + \sum_{i=1}^P a_i r_{ss}[-i] \end{aligned} \quad (63)$$

onde $r_{ss}[k]$ é a autocorrelação do sinal $s(n)$ considerando um deslocamento de k amostras.

Estas são equações de Yule-Walker para processos AR, ou equações de Wiener-Hopf, conforme a abordagem que se dê ao problema de predição linear.

Diversos algoritmos são encontrados na literatura para a resolução deste problema, como o algoritmo de Levinson (normalmente empregado nas análises acústicas por sua eficiência na determinação dos coeficientes, e pela recursividade que permite a estimação dos P parâmetros a partir dos $P-1$ prévios), método da covariância e covariância modificada, método de Burg, entre outros.

A caracterização deste modelo paramétrico para estimação do sistema de produção de fala é importante para a avaliação acústica da voz se o objetivo é avaliar o sinal glotal e desta forma, o comportamento (normal ou patológico) das estruturas da laringe. Retornando à eq. (59), pode-se reescrevê-la de forma a obter o modelo de excitação, ou seja:

$$E(z) = \frac{1}{\sigma} \sum_{i=0}^P a_i z^{-i} S(z) \quad (64)$$

que transcrito no domínio temporal, obtêm-se o gerador de pulsos de excitação, ou seja:

$$e(n) = \frac{1}{\sigma} \sum_{i=0}^P a_i s(n-i) \quad (65)$$

Este novo formato é conhecido por filtro preditor de erro, onde o sinal estimado de excitação é agora conhecido por erro de estimação. Considerando um estimador polarizado da função de autocorrelação r_{ss} na resolução das equações de Yule-Walker, uma das propriedades do filtro é a sua mínima fase, que garante a estabilidade do sistema uma vez que os zeros estão situados dentro do círculo unitário, e assim, assegura que o processo $s(n)$ é estacionário em sentido amplo (WSS).

O erro de predição (ou no caso dos sinais de voz, pulsos de estimação) caracteriza-se por aproximar-se muito de um ruído branco, devido a capacidade de maximização do “spectral flatness”. Uma vez que o sinal de excitação do modelo de produção de fala compõem-se de pulsos espaçados, estes serão obtidos na saída do filtro, com presença de ruído nos intervalos entre os pulsos.

O termo inverso advém do fato de estimar-se o espectro densidade de potência do trato supraglotal (com alta resolução) por meio de um modelo AR, e através da filtragem do sinal de voz usando o sistema inverso ao estimado, obtendo-se então os resíduos, ou seja, o sinal de excitação do sistema linear invariante no tempo.

É importante caracterizar este sinal de excitação, que é diferente do sinal glotal real. Isto advém das simplificações realizadas quando da caracterização do trato vocal, onde o efeito causado pelo subsistema que representa a laringe e cordas vocais produzirá o sinal glotal. Na determinação dos parâmetros do modelo AR, tanto $G(z)$, $R(z)$ e $V(z)$ estão aglutinadas, caracterizando o efeito conjunto destas estruturas na produção da voz.

A técnica de Davis (1979) baseia-se no modelamento apresentado nesta seção. Empregando o algoritmo de Levinson sobre o trecho mais estável do sinal de voz, determina-se um modelo AR, que aplicado sob a forma de filtro inverso fornece o sinal residual, ou de excitação o modelo paramétrico do trato vocal. Maiores detalhes deste trabalho são apresentados na seção 3.6.

• Filtros Adaptativos

Para se estimar o sinal de excitação do modelo apresentado, algumas considerações devem ser feitas. A principal é a suposição de estacionariedade do sinal, uma vez que a estimativa paramétrica estimará um sistema linear invariante no tempo e desta forma exige o conhecimento de informações estatísticas de primeira e segunda ordem. Desta forma, é necessário que se desloque a janela temporal para avaliar amostras supostamente invariantes.

Apesar das análises serem realizadas sobre fonemas vocálicos sustentados, a natureza patológica das vozes de pacientes pode exigir técnicas de estimação residual que considerem a alteração na estrutura formante. Desta forma, o emprego de algoritmos adaptativos torna-se interessante.

Avaliando esta metodologia de cálculo de Davis (1979), percebe-se que os pontos críticos estão na exigência de estacionariedade do sinal pela definição matemática da estimativa do filtro autoregressivo (DELLER JR *et al*, 1993), impondo assim um trato vocal invariante no tempo por alguns milissegundos.

Assim, movimentação da língua, por exemplo, mesmo que durante a fala sustentada, pode refletir no sinal residual como alguma disfonia ocorrendo ao nível glotal.

O segundo ponto importante é a definição do trecho inicial para estimação, ou seja, determinação do trecho mais estável do sinal de voz. As características de fluxo de ar iniciais combinadas com a constrição da laringe (“ataque vocal”) são diferentes na população (normal e patológica), exigindo assim que o início da estimativa seja feita de acordo com a pessoa.

São apresentados algoritmos adaptativos que consideram a não estacionariedade do sinal e estimam o sinal residual, além de medidas acústicas que representem este novo sinal, para posterior emprego em discriminador estatístico ou neural.

• Estimativa usando Gradiente Descendente

A primeira abordagem neste sentido é através da utilização de filtros causais ótimos em relação ao erro médio quadrático (filtros de Wiener), para a produção do sinal residual que, como já citado, apresenta informações sobre a estrutura glotal. O modelo adotado para estes filtros preditivos é (percebe-se

que σ da eq. (65) foi ignorado temporariamente uma vez que não se tem conhecimento estatístico sobre o sinal):

$$\begin{aligned} e(n) &= s(n) - \mathbf{a}_n^T \mathbf{s}_{n-1} \\ \mathbf{s}_n &= [s(n), s(n-1), \dots, s(n-P+1)] \end{aligned} \quad (66)$$

Considerando inicialmente a abordagem pela regra do gradiente, em que objetiva-se minimizar o erro médio quadrático entre a amostra real e a amostra obtida pelo preditor (operação de filtragem), tem-se a seguinte relação de recursão para atualização dos parâmetros do modelo paramétrico:

$$\mathbf{a}_{n+1} = \mathbf{a}_n + \tilde{\nu}[-\mathbf{g}] \quad (67)$$

onde $\tilde{\nu}$ é a taxa de convergência, para adaptação dos coeficientes do modelo e $[-\mathbf{g}]$ denota o gradiente negativo do erro, $E\{e^2(n)\}$, em relação à $\mathbf{a}_n$. Como não é de conhecimento esta última informação estatística, é necessário estimá-la, ou seja:

$$\hat{\mathbf{g}} = \nabla_{\mathbf{a}_n} \{e^2(n)\} = 2 \cdot e(n) \cdot \nabla_{\mathbf{a}_n} \{e(n)\} \quad (68)$$

Considerando agora a eq. (9) que rege a produção da fala, pode-se estimar o gradiente do erro com relação ao vetor de coeficientes $\mathbf{a}_n$, tem-se que:

$$\nabla_{\mathbf{a}_n} \{e(n)\} = -\mathbf{s}_{n-1} \quad (69)$$

E assim, da eq. (12) tem-se a recursão que adapta os parâmetros do filtro à cada nova amostra que é apresentada:

$$\mathbf{a}_{n+1} = \mathbf{a}_n + 2 \cdot \tilde{\nu} \cdot e(n) \cdot \mathbf{s}_{n-1} \quad (70)$$

Uma vez que se espera que os parâmetros estimados sejam estáveis e sua convergência siga para valores estáveis, é necessário que a taxa de convergência esteja no intervalo:

$$0 < \tilde{\nu} < \frac{2}{\lambda_{\max}} \quad (71)$$

onde $\lambda_{\max}$ corresponde ao maior autovalor da matriz de autocorrelação $\mathbf{R}_{ss}$. Como novamente as informações estatísticas não estão disponíveis, utiliza-se o somatório dos autovalores desta matriz, que corresponde à potência média do sinal $s(n)$. Sua estimativa pode ser realizada através de:

$$\sigma_s^2(n) = \beta \cdot \sigma_s^2(n-1) + (1-\beta) \cdot s^2(n) \quad (72)$$

Uma vez que β corresponde à uma constante de tempo nesta recursão, e considerando N o número de amostras sobre o qual se trabalhará, tem-se que N é aproximadamente igual ao inverso de $(1-\beta)$. Assim, a recursão final para adaptação dos parâmetros do filtro preditor de erro é dada por:

$$\mathbf{a}_{n+1} = \mathbf{a}_n + \frac{1-\beta}{\sigma_s^2(n)} \cdot e(n) \cdot \mathbf{s}_{n-1} \quad (73)$$

Percebe-se agora que a estimativa da variância aparece como condição básica para convergência do procedimento, mesmo que fosse inicialmente desconsiderada. Se o valor fosse conhecido, no desenvolvimento analítico da expressão recursiva ter-se-ia a equivalência:

$$\sigma_s^2(n) = \sigma \quad (74)$$

Naturalmente então, a cada estimativa do modelo do trato vocal (caracterizado pelos parâmetros $\mathbf{a}$), o sinal residual, caracterizando a excitação do sistema, é dada por $e(n)$.

• Estimativa Recursiva por Mínimos Quadrados – RLS

Uma segunda abordagem ao problema de estimação é o emprego da técnica de estimação recursiva por mínimos quadrados. Considerando que a obtenção de uma estimativa de $s(n)$ pode ser obtida a partir de uma combinação linear de amostras anteriores, a solução direta para este problema é:

$$\begin{aligned} \hat{\mathbf{a}}_n &= [\mathbf{S}_{n-1}^T \mathbf{S}_{n-1}]^{-1} \cdot [\mathbf{S}_{n-1}^T \mathbf{S}_n] \\ \mathbf{S}_n &= \text{col}[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \\ \mathbf{s}_n &= [s(n-1), s(n-2), \dots, s(n-P-1)] \end{aligned} \quad (75)$$

Entretanto, devido ao não conhecimento das amostras, o emprego da recursividade se faz necessário, para a estimação adaptativa do erro de predição. Considerando a autocovariância das amostras passadas denotada por $\mathbf{P}_n$, tem-se que

$$\begin{aligned} \mathbf{P}_n &= [\mathbf{s}_{n-1}^T \mathbf{s}_{n-1}]^{-1} \\ \mathbf{P}_{n+1} &= [\mathbf{s}_n^T \mathbf{s}_n]^{-1} = [\mathbf{P}_n^{-1} + \mathbf{s}_{n+1}^T \mathbf{s}_{n+1}]^{-1} \end{aligned} \quad (76)$$

Empregando o lema de inversão de matrizes, obtem-se então a equação para atualização da matriz de covariância, ou seja:

$$\mathbf{P}_{n+1} = \mathbf{P}_n - \frac{\mathbf{P}_n \mathbf{s}_{n+1} \mathbf{s}_{n+1}^T \mathbf{P}_n}{1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1}} \quad (77)$$

com $\mathbf{P}_0 = \rho \mathbf{I}$, com ρ normalmente elevado para acelerar a convergência do algoritmo.

Desta forma, retornando à eq. (20), obtém-se a forma de recursiva de predição linear para o sinal de excitação do trato vocal, ou seja:

$$\hat{\mathbf{a}}_{n+1} = \hat{\mathbf{a}}_n + \kappa_{n+1} [s(n+1) - \mathbf{s}_{n+1} \hat{\mathbf{a}}_n] \quad (78)$$

onde

$$\begin{aligned} \delta_{n+1} &= 1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1} \\ \kappa_{n+1} &= \mathbf{P}_n \mathbf{s}_{n+1} \delta_{n+1}^{-1} \end{aligned} \quad (79)$$

Assim, o sinal de excitação do modelo de produção de fala é obtido à cada interação, através da eq (21), ou seja:

$$e_{n+1} = s(n+1) - \mathbf{s}_{n+1} \hat{\mathbf{a}}_n \quad (80)$$

• Estimativa por Filtro de Kalman

A abordagem de estimação do sinal residual, pela determinação paramétrica do sistema, através de filtros de Kalman foi desenvolvida considerando a existência de ruído, possivelmente correlacionado, no sinal residual.

Considerando esta possibilidade, poder-se-ia pensar que uma filtragem mais robusta pudesse ser realizada de modo a se obter um sinal excitatório mais condizente realístico, sem efeitos mascaráveis advindos do trato supraglotal (por ser estimado).

O desenvolvimento desta técnica é realizada em duas etapas. Em primeiro lugar deve-se estimar as matrizes de estado do sistema de produção de fala. Isto porque na segunda etapa, será estimado o resíduo propriamente dito através de filtro de Kalman e, a operação deste se dá através do conhecimento destas informações.

A estimativa do espaço de estados foi realizada a partir de um modelo ARX, utilizando o método dos mínimos quadrados. Assim, os coeficientes do sistema abaixo serão determinados:

$$\mathbf{A}(z^{-1})s_i = \mathbf{C}(z^{-1})e_i \quad i = 0, 1, \dots, P-1 \quad (81)$$

Conhecendo-se amostras passadas, busca-se determinar os coeficientes $\mathbf{A}$ e $\mathbf{C}$, representados por $\mathbf{a}_n$, a equação do erro de estimação é dada por:

$$\begin{aligned} \hat{e}(n) &= s(n) - \mathbf{x}_n \hat{\mathbf{a}}_n \\ \mathbf{x}_n &= [-s(n-1), \dots, -s(n-P-1), e(n-1), \dots, e(n-N-1)] \\ \hat{\mathbf{a}}_n &= [a_1, \dots, a_{P-1}, c_1, \dots, c_{N-1}]^T \end{aligned} \quad (82)$$

Minimizando o quadrado da equação acima, tem-se que os parâmetros do modelo ARX, considerando o conhecimento prévio de $s(n)$ e $e(n)$, são dados por:

$$\begin{aligned} \hat{\mathbf{a}}_n &= [\mathbf{X}_{n-1}^T \mathbf{X}_{n-1}]^{-1} \cdot [\mathbf{X}_{n-1}^T \mathbf{X}_n] \\ \mathbf{X}_n &= \text{col}[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \end{aligned} \quad (83)$$

Para uso destas equações de forma adaptativa (uma vez que o trato vocal é um sistema variante no tempo), e considerando o lema de inversão de matrizes, as equações que regem a identificação paramétrica é:

$$\begin{aligned} \hat{\mathbf{a}}_{n+1} &= \hat{\mathbf{a}}_n + \kappa_{n+1} [s(n+1) - \mathbf{x}_{n+1} \hat{\mathbf{a}}_n] \\ \kappa_{n+1} &= \frac{\mathbf{P}_n \hat{\mathbf{x}}_{n+1}^T}{1 - \hat{\mathbf{x}}_{n+1} \mathbf{P}_n \hat{\mathbf{x}}_{n+1}^T} \\ \mathbf{P}_{n+1} &= [\mathbf{I} - \kappa_{n+1} \hat{\mathbf{x}}_{n+1}^T] \mathbf{P}_n \\ \hat{\mathbf{x}}_n &= [-s(n-1), \dots, -s(n-P-1), -e(n-1), \dots, -e(n-N-1)] \cdot \hat{\mathbf{C}}(z^{-1}) \end{aligned} \quad (84)$$

com $\mathbf{P}_0 = \rho \mathbf{I}$, com ρ normalmente elevado para acelerar a convergência do algoritmo.

A cada iteração, as matrizes $\mathbf{A}$ e $\mathbf{C}$ são utilizadas no processo de filtragem propriamente dito, através do emprego de filtros de Kalman. Primeiramente é realizado uma transformação da equação à diferença eq. (81), com os parâmetros já estimados, para a representação em espaço de estados, obtendo-se assim o seguinte sistema:

$$\begin{cases} \mathbf{z}_{n+1} = \mathbf{M}_n \mathbf{z}_n + \mathbf{N}_n w(n) \\ s(n) = \mathbf{H}_n \mathbf{z}_n + w(n) \end{cases} \quad (85)$$

onde $\mathbf{w}_n$ é um ruído branco com variância unitária.

A partir deste sistema (variante no tempo), determina-se então a melhor estimativa linear para o sinal de voz, baseado na consideração estocástica do sistema eq. (85). Determina-se primeiramente, no filtro de Kalman, a estimativa do espaço de estado à posterior, através das equações abaixo:

$$\begin{aligned}\hat{\mathbf{z}}_{n+1|n} &= \mathbf{M}_n \hat{\mathbf{z}}_{n|n-1} + \kappa'_n [s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1}] \\ \hat{\mathbf{z}}_{0|-1} &= 0 \\ \kappa'_n &= \frac{\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n}{1 + \mathbf{H}_n \mathbf{P}'_n \mathbf{H}_n^T}\end{aligned}\quad (86)$$

A matriz de covariância $\mathbf{P}'_n$ é obtida, na forma recursiva, através da equação de Riccati, ou seja:

$$\begin{aligned}\mathbf{P}'_{n+1} &= \mathbf{M}_n \mathbf{P}'_n \mathbf{M}_n^T + \mathbf{N}_n \mathbf{N}_n^T - \kappa'_n [\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n] \\ \mathbf{P}'_0 &= \rho \mathbf{I}\end{aligned}\quad (87)$$

A partir destas iterações, a cada nova amostra no instante n , surge que o erro de estimação, que corresponde ao sinal de excitação do trato vocal, ou seja:

$$e'(n) = s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1} \quad (88)$$

Desta forma, tem-se a sequência residual que espera-se corresponder à melhor estimativa do sinal de excitação do modelo de produção de fala, segundo a abordagem de Kalman.

• Considerações sobre Convergência

Diferente da técnica convencional de produção do sinal residual, através da suposição de estacionariedade de uma janela de dados sobre a qual se aplica o método da autocorrelação, os métodos adaptativos necessitam de uma quantidade de amostras (tempo) para atingirem estabilidade assintótica dos coeficientes dos filtros.

Considerando que as medidas acústicas são calculadas sobre todo o sinal residual, estas sofrem influência direta desta característica dos filtros adaptativos pois as primeiras amostras de resíduo ainda não condizem com o sinal de excitação do modelo do trato vocal. Assim, medidas errôneas podem ser geradas se o tempo de convergência não for levado em conta.

A primeira iniciativa neste sentido é desconsiderar o conjunto de amostras iniciais do sinal residual correspondente à um tempo mínimo de influência da convergência nas medidas acústicas. Para isto, deve-se levar em conta o padrão de voz que será analisado. Verificou-se que para vozes normais, os algoritmos adaptativos estudados tem um tempo de convergência inferior à um período de frequência fundamental, principalmente quando aplica-se os algoritmos por RLS e filtros de Kalman usando valores elevados para ρ (empregou-se $\rho = 1000$).

Entretanto, vozes perceptualmente ásperas ou roucas apresentam grandes variações na frequência fundamental (conforme extensão da patologia) e foi necessário avaliar uma quantidade de amostras residuais que seriam desconsideradas na determinação dos parâmetros acústicos. Concluiu-se que desconsiderando o equivalente à um período de 50 Hz (20 ms) do sinal residual, que corresponde à uma voz extremamente grave é suficiente para os casos patológicos considerados.

Mesmo assim, é necessário empregar uma segunda técnica, conjuntamente à apresentada, para aumentar a robustez dos parâmetros acústicos. Isto porque o aumento do ruído de alta frequência e modificações na estrutura espectral (redução energética a partir do 2. formante) das vozes roucas (carcinomas e pacientes corpectomizados, por exemplo) retardam a convergência dos coeficientes dos filtros, principalmente no algoritmo de gradiente descendente.

É realizada uma estimação para os filtros que empregam o modelo AR (filtros de Wiener, nesta pesquisa), utilizando-se o método de autocorrelação sobre o conjunto das 1024 primeiras amostras do sinal de voz. Utilizando o algoritmo de Levinson na resolução das equações de Yule-Walker, obtém-se um conjunto de coeficientes para emprego nos filtros adaptativos, reduzindo-se assim o tempo de convergência destes. Note que as amostras utilizadas neste cálculo são reutilizadas pelos algoritmos adaptativos.

Apesar da possibilidade de redução da quantidade de amostras residuais descartadas no cálculo das medidas acústicas para 10 ms ou menos, de acordo com a patologia, manteve-se a primeira abordagem sem alteração, pois seria necessário avaliar para cada patologia a quantidade ótima de resíduos para desconsideração.

Para os coeficientes do modelo ARX do filtro de Kalman não foi realizada a sua estimação prévia, motivado pela sua velocidade de convergência, apesar da implementação não ser complexa.

4.3.6 Medidas do Sinal Residual

Considerando agora os três algoritmos implementados, é importante desenvolver um procedimento para se obter informações relevantes sobre as características do sinal de resíduo de estimação, de modo a relacionar este com as patologias, objetivo pelo qual buscou-se uma maior precisão na sua derivação.

A primeira medida considerada é SNR (relação sinal-ruído). Considerando que o sinal residual é o erro de estimativa entre as amostras estimadas para um modelo paramétrico e o sinal real, achou-se conveniente determinar esta medida pois a voz proveniente de algumas patologias, como cistos, pólipos e carcinomas caracterizam-se por soprosidade, e assim, ruído em alguns intervalos de frequência no espectro de potência.

O seu cálculo é feito sobre o trecho mais estável (considerando aspectos de convergência) e corresponde ao valor médio da soma do quadrado das amostras de resíduo.

A segunda medida é EX (Coeficiente de excesso) e corresponde matematicamente à medida de curtose da distribuição das amplitudes do sinal residual. Verifica-se em pessoas normais que a curva é leptocúrtica, enquanto que para padrões de voz soprosos, a distribuição é platicúrtica, pois o padrão de pulsos periódicos existente neste sinal é mais definido em vozes normais do que em patológicas. Este parâmetro é determinado pela razão entre o momento de quarta ordem pelo quadrado do momento de segunda ordem, dos resíduos, ou seja:

$$EX = \frac{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^4}{\left(\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^2 \right)^2} - 3 \quad (89)$$

onde N é a quantidade de resíduos, $\bar{s}$ é o valor médio dos resíduos neste segmento e s_i é o i -ésimo resíduo do trecho mais estável da filtragem inversa.

A determinação deste, e dos próximos parâmetros, é realizada a cada trecho de 4096 amostras, sem sobreposição, para se determinar um valor mediano final com maior correlação entre a disfonia e o sinal de voz.

Para se avaliar o comportamento espectral do sinal de voz, definiu-se medidas para avaliação do nivelamento (*flattness*) do sinal de voz, pois sabe-se que quanto maior a quantidade de ruído, maior é o valor obtido para as medidas de nivelamento espectral, onde os limites teóricos são zero à menos infinito.

Primeiramente, SFR ("*spectral flattness*" do ruído) permite verificar o grau de mascaramento dos harmônicos por ruído (proveniente do fechamento incorreto da glote). É obtido através da razão entre a média geométrica pela média aritmética do sinal residual (em escala logarítmica), ou seja:

$$SFR = 10 \cdot \log \left(\frac{\frac{1}{N} \prod_{i=1}^N s_i}{\frac{1}{N} \sum_{i=1}^N s_i} \right) \quad (90)$$

onde N é a quantidade de resíduos e s_i é o i -ésimo resíduo do trecho mais estável da filtragem inversa.

Semelhante à esta medida, tem-se SFF ("*spectral flattness*" do filtro inverso), para avaliação do mascaramento das formantes por ruído. Sabe-se que soproidade é um fenômeno que acarreta em um aumento da energia das componentes não harmônicas da voz e simultânea redução das amplitudes nas frequências formantes. Esta relação de energia é acompanhada por esta medida. Matematicamente é definida pela diferença entre a medida de nivelamento espectral do sinal de voz pela medida SFR.

Vozes ásperas apresentam elevado valor para SFR, pois perturbação na frequência fundamental e harmônicos caracteriza patologias como hipercinesia, avaliando o ruído proporcionado por estas alterações no espectro de frequências. Já vozes soprosas têm valores elevados para SFF pois uma quantidade de ruído elevado, proveniente das glote (adução incorreta), reduz a energia nas frequência formantes com aumento das amplitude laterais. Juntamente com SNR, estas medidas tem por objetivo determinar a quantidade de ruído no espectro.

A medida *Jitter* mede a perturbação da frequência fundamental. Sabe-se que patologias como carcinomas, laringite crônica e paralisia unilateral, por

exemplo, produzem sinais de voz com alta variação ciclo-a-ciclo do período de pitch, enquanto que outras como cistos ou nódulos tem oscilações reduzidas. Assim, é importante avaliar esta característica da voz.

Para pessoas normais, o sinal residual consiste de uma série de picos periódicos bem definidos, com baixa amplitude nas amostras laterais. Assim, pode-se determinar o período da frequência fundamental como diferença entre picos sucessivos. A estimativa destes períodos foi realizada de forma síncrona para garantir maior precisão, determinando-se as maiores diferenças absolutas de amplitudes (picos bem definidos), em sucessivas regiões de existência destes picos.

O centro destas regiões foi calculado através de uma estimação assíncrona da frequência fundamental, utilizando o método da autocorrelação sobre segmentos de 4096 amostras, sem sobreposição. O intervalo de busca corresponde à $\frac{1}{2}$ período de F_0 estimado. Uma vez determinado precisamente o instante da ocorrência do pico, realiza-se uma interpolação parabólica para garantir precisão de 0.05% (TITZE *et al*, 1987).

O parâmetro Jitter é então definido como a mediana das perturbações de primeira ordem, ou seja, diferenças entre sucessivos períodos da fundamental. A mediana é empregada para garantir robustez na medição.

Definiu-se a medida Pico para avaliar a relação entre as patologias e as estimativas de picos periódicos desenvolvidas para o parâmetro anterior.

Considerando a normalização do sinal de voz, espera-se que valores elevados para este parâmetro resultem na discriminação de pessoas normais, e proporcionalmente à diferença absoluta entre os períodos de pico (menor valor para este parâmetro), obtém identificação das patologias.

O cálculo desta medida baseia-se na mediana das máximas diferenças absolutas entre amplitudes, que caracteriza os picos periódicos.

Estas duas medidas acústicas (Jitter e Pico) foram desenvolvidas para realizar a mesma tarefa de APQ e PPQ de Koike (1973), Koike *et al* (1977) e Davis (1979), com menor exigência de processamento e empregando medida estatística robusta à oscilações aleatórias com baixa probabilidade de ocorrência.

A última medida desenvolvida também avalia a periodicidade do sinal residual. Através da determinação da mediana das amplitudes do segundo pico

da autocorrelação normalizada do sinal residual (já calculada previamente, na estimação dos períodos de frequência fundamental), a medida **P/A** (*Pico na autocorrelação*) possibilita determinar o nível de perturbação na periodicidade do sinal de voz pois a redução na amplitude do segundo pico da autocorrelação e aumento das amplitudes na sua vizinhança. Patologias de origem soprosa (nódulos) ou rouca (paralisia unilateral) apresentam baixos valores para este parâmetro.

As medidas acústicas baseadas no sinal residual que foram apresentadas devem ser capazes de representar adequadamente o sinal residual para que durante o emprego da rede neurais artificiais para classificação e discriminação das patologias, estas estruturas possam “aprender” adequadamente este tipo de sinal para poder reconhecer o padrão vocálico das disfonias.

As medidas Jitter, Pico e P/A informam o nível de perturbação do sinal residual pela importância deste tipo de medida na avaliação de patologias (BAKEN, 1996). SFF, SFR e SNR por sua vez, avaliam a quantidade de ruído turbulento proveniente do inadequado funcionamento das pregas vocais na atividade de abertura e constrição da glote. O parâmetro EX permite avaliação tanto de perturbação da frequência fundamental quanto avaliação de oscilações aleatórias, pois a distribuição das amostras é afetada por estas duas características patológicas, e o grau de achatamento desta curva é um indicio da extensão da disfonia.

Capítulo 5

Teoria de Decisão e Resultados

5.1. Teoria de Decisão

O trabalho estatístico é iniciado no momento da descrição dos níveis de incerteza de um cientista sobre as quantidades de interesse. A ferramenta usada para a descrição do nível de incerteza é a Probabilidade. A probabilidade de um estado de natureza específico, digamos θ , é um índice que indica o nível de incerteza (ou conhecimento) sobre a veracidade da afirmação “o valor do estado, θ é θ_0 ”. Para este importante índice de incerteza, existem regras específicas e bem estabelecidas que devem ser obedecidas.

Probabilidades devem ser atribuídas a todos os estados de natureza possíveis. Ao conjunto de todas as afirmações probabilísticas de um problema usamos o termo modelo probabilístico ou, equivalentemente, distribuição de probabilidades.

Assim, o estatístico deve definir o modelo probabilístico que, em sua opinião, melhor representa o estado de incerteza sobre θ e procurar fontes adicionais de informação que possam diminuir esse nível de incerteza.

Na procura pela diminuição da incerteza, após definir a distribuição de probabilidades de θ , inicia-se a busca por quantidades observáveis, X por exemplo, que estejam associadas a θ . O objetivo é observar o valor x de X e transformar o modelo de probabilidade de θ no modelo de probabilidade (condicional) de θ dado que $X = x$.

O mecanismo que transforma uma quantidade não observada, X , em observada, x , é aqui denominado experimento. O conjunto de quantidades obtidas após a realização de experimentos é denominado conjunto dos resultados experimentais ou banco de dados.

O modelo probabilístico de θ , definido antes (depois) da realização do experimento X , é denominado distribuição a priori (a posteriori) de θ .

Efetuar a operação priori/posteriori é, na verdade, construir um processo indutivo de ganho de conhecimento. O que foi chamado de distribuição a posteriori em um dado momento, pode no momento seguinte passar a ser uma distribuição a priori para um novo experimento, digamos Y .

Muitas vezes enfrenta-se um dilema: Que variáveis aleatórias observar para diminuir a incerteza sobre θ . Ele pode ter a opção de realizar um dentre um conjunto de experimentos, digamos X, Y, Z , etc.

Pode haver custos ou restrições associados à realização dos experimentos. Assim, faz-se necessário selecionar quais serão realizados. Para uma escolha adequada, ele deve descrever o tipo de associação que, em sua opinião, existe entre θ e cada um das variáveis aleatórias dos experimentos.

Essa descrição também é feita por meio de modelos probabilísticos. Por exemplo, considere que o experimento X vai ser realizado. Para cada valor θ do parâmetro, represente por $p(x|\theta)$ a função de probabilidade que avalia, para cada x , a probabilidade de observarmos $X = x$. Se existem pelo menos dois valores de θ , digamos θ_1 e θ_2 , tal que as funções (em x), $p(x | \theta_1) \neq p(x | \theta_2)$, então θ e X são dependentes ou associados. Neste caso é razoável realizar-se o experimento X com o intuito de diminuir a incerteza sobre θ .

O modelo $p(x|\theta)$ é uma função de duas componentes, x e θ . Para um valor fixado de θ , p é uma função de probabilidades, isto é $p(x|\theta) = p(X = x|\theta)$.

Por outro lado, após observar $X = x$, p é uma função apenas de θ . Neste caso p passa a ser denominada função de verossimilhança de θ com respeito a observação x . Para evitar mal entendidos, usaremos a notação $L(\theta|x)$ para a verossimilhança e $p(x|\theta)$ para a função de probabilidades.

Seja θ o parâmetro, e X uma variável aleatória, que está relacionada a θ na opinião do cientista. O conjunto de todos os valores possíveis de θ , Θ , é denominado espaço paramétrico. O conjunto de todos os valores possíveis de X , X , é denominado espaço amostral. Os seguintes conjuntos são os principais elementos do trabalho estatístico.

- ✓ Distribuição a priori : $f(\theta) : \Theta \rightarrow \mathbb{R}$
- ✓ Distribuição Amostral : $p(x|\theta) : X \rightarrow \mathbb{R}, \forall \theta \in \Theta$
- ✓ Função de Verossimilhança : $L(\theta|x) = p(x|\theta) : \Theta \rightarrow \mathbb{R}, \forall x \in X$
- ✓ Distribuição a Posteriori : $f(\theta|x) : \Theta \rightarrow \mathbb{R}, \forall x \in X$

Como vimos, deve-se realizar dentro das possibilidades, os experimentos que podem proporcionar um ganho maior de informação. Entretanto, a escolha de qual experimento deve ser realizado é feita, evidentemente, antes de sua realização. Assim, sua escolha é feita com base em suas expectativas sobre os possíveis e prováveis resultados daqueles experimentos que se apresentam

como alternativas. Essa análise de expectativas, para cada um dos experimentos concorrentes, pode ser identificada como o clássico Planejamento de Experimentos. Essa área é caracterizada pela procura dos planejamentos ótimos. Em muitas situações, esses são experimentos não realizáveis devido à restrições operacionais ou de custos. No contexto Bayesiano (devido ao reverendo Thomas Bayes), a procura do que realizar e observar é chamado de análise pré-posteriori. Esse trabalho inicial do estatístico Bayes consiste em comparar, para cada experimento possível, priori e posteriori esperada.

Dentre os conceitos fundamentais relacionados ao processo de decisão destaca-se o Teorema de Bayes, que é uma ferramenta útil para testes de hipóteses, relacionando probabilidades condicionais e conjuntos de eventos A e B [SKLAR, 1997]:

$$P(A|B) P(B) = P(B|A) P(A) = P(A, B)$$

Um formato mais frequentemente encontrado para a expressão de Bayes é :

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

No processo de discriminação das 4 patologias da laringe (Edema, Fenda, Nódulo e Pólipo) e voz normal, a expressão de Bayes toma sua versão contínua dada por :

$$P(d = i | x) = \frac{p(x | d = i) P(d = i)}{p(x)}, i = 1, \dots, M$$

onde $P(d=i|x)$ representa a probabilidade a posteriori de d ter assumido o valor i ($d=i$, ou seja, igual a umas das patologias ou voz normal), a partir do parâmetro acústico x (lembrando que estes são de um total de 38) associado. A função $p(x|d=i)$ é a função de verossimilhança (*likelihood function*) do parâmetro x , dada patologia ou voz normal, $d=i$. O valor $P(d=i)$ é a probabilidade a priori de ocorrência do evento $d=i$ (probabilidade das patologias e voz normal, $d=i$). Finalmente, $p(x)$ é a função de densidade de probabilidade total sobre todo o espaço das M patologias e voz normal, ou seja :

$$p(x) = \sum_{i=1}^M p(x | d = i) P(d = i)$$

Os mais utilizados critérios de decisão aplicados são baseados no teste de hipóteses instrumentalizado pelo Teorema de Bayes. São eles : o critério de máximo posteriori (MAP, *Maximum A-Posteriori*) e o critério de máxima verossimilhança (ML, *Maximum Likelihood*) . Dado o exemplo $\{d = (\text{Patologia 1}, \text{Patologia 2}) ; M=2 \text{ e } i=1,2\}$, o critério MAP estabelece :

$$P(d = \text{Patologia 1} | x) \underset{< (H2)}{>^{(H1)}} P(d = \text{Patologia 2} | x)$$

o que significa: escolha a hipótese 1 (patologia 1) se a probabilidade a-posteriori $P(d=\text{Patologia 1}|x)$ for maior que $P(d=\text{Patologia 2}|x)$ e escolha a hipótese 2 (Patologia 2) em caso contrário.

Utilizando o teorema de Bayes pode-se escrever a expressão anterior na forma equivalente :

$$p(x | d = \text{Patologia 1}) P(d = \text{Patologia 1}) \underset{< (H2)}{>^{(H1)}} p(x | d = \text{Patologia 2}) P(d = \text{Patologia 2})$$

ou

$$\frac{p(x | d = \text{Patologia 1}) P(d = \text{Patologia 1})}{p(x | d = \text{Patologia 2}) P(d = \text{Patologia 2})} \underset{< (H2)}{>^{(H1)}} 1$$

Na expressão anterior tem-se, no termo da esquerda, a conhecida razão de verossimilhança. Para patologias com mesma probabilidade (equiprováveis), a expressão reduz-se a :

$$\frac{p(x | d = \text{Patologia 1})}{p(x | d = \text{Patologia 2})} \underset{< (H2)}{>^{(H1)}} 1$$

o que corresponde ao critério de máxima verossimilhança (ML). Verifica-se, portanto, que para patologias equiprováveis, o critério ótimo MAP reduz-se ao critério ML (também ótimo nesse caso).

Esse procedimento descrito acima, foi aplicado para 4 patologias (Edema, Fenda, Nódulo e Pólipo) e a voz normal, em etapas univariada e multivariada, que serão melhor detalhadas em resultados.

5.2. Resultados

Neste tópico são apresentados os resultados da classificação das patologias em análise baseado na aplicação da teoria Bayesiana descrita acima. A principal informação resultante desse trabalho é a de destacar os parâmetros mais convenientes para a discriminação de patologias da laringe, no caso a população patológica em questão é composta por edema, fenda, nódulo e pólipos e para completar o estudo a voz normal. O procedimento foi dividido em duas etapas : univariada e multivariada.

5.2.1. Etapa univariada

Nesta etapa testa-se o percentual discriminatório de cada uma das 38 medidas acústicas individualmente com relação as 4 patologias em questão mais a voz normal. Para cada medida, reserva-se 50% do número de casos de cada patologia para a fase de treinamento, onde obtém-se a média e a variância. Os outros 50% destinam-se a fase de teste, onde aplica-se a teoria Bayesiana para a obtenção da discriminação. A tabela 5.2 mostra os percentuais de acerto das 38 medidas individualmente na discriminação das 4 patologias e da voz normal. A figura 5.1 mostra o gráfico da variação percentual discriminatória das 38 medidas com relação a patologia edema. A tabela 5.1 destaca as medidas acústicas com melhores resultados discriminatórios, ou seja, com pelo menos 50% de acerto discriminatório da patologia edema.

Figura 5.1. Variação percentual discriminatória para a patologia edema.

Tabela 5.1. Medidas que melhor discriminaram a patologia edema.

Exec	Parâmetro	Edema
11	Hfpr_>6000Hz	92
8	Shimmer_JR	89
34	lne_400-2000Hz	89
2	Jitter_PFF	76
37	Pvi	73
13	Hfpr_400-2000Hz	63

Tabela 5.2. Percentual discriminatório das 38 medidas acústicas para as 4 patologias e voz normal na etapa univariada.

Exec	Parâmetro	Edema	Fenda	Nódulo	Normal	Pólipo
1	Fo	0	5	70	50	17
2	Jitter_PFF	75	0	0	67	7
3	Jitter_DPF	5	35	6	71	10
4	Jitter_JR	10	0	16	91	21
5	Jitter_JF	0	0	83	2	25
6	Shimmer_PFF	0	25	86	0	32
7	Shimmer_DPF	1	0	33	82	35
8	Shimmer_JR	89	0	3	17	0
9	Shimmer_JF	30	0	53	63	17
10	Br	0	20	0	0	64
11	Hfpr_>6000Hz	92	0	0	52	0
12	Hfpr_60-400Hz	0	10	3	86	3
13	Hfpr_400-2000Hz	53	0	20	50	32
14	Hfpr_2000-5000Hz	5	10	36	63	7
15	Snl_Média_100-2600Hz	14	85	26	0	0
16	Snl_Mediana_100-2600Hz	12	85	33	0	0
17	Snl_Média_2600-5100Hz	12	40	03	2	0
18	Snl_Mediana_2600-5100Hz	12	45	56	0	0
19	Snl_Média_5100-8000Hz	12	50	53	4	0
20	Snl_Mediana_5100-8000Hz	12	70	56	13	0
21	Snl_Média_100-5100Hz	8	60	33	17	7
22	Snl_Mediana_100-5100Hz	12	45	46	4	0
23	Snl_Média_100-8000Hz	0	55	53	4	32
24	Snl_Mediana_100-8000Hz	0	0	0	73	89
25	Ra_Média	0	0	33	82	0
26	Ra_Mediana	14	0	33	69	28
27	Hr_Média	44	50	0	13	3
28	Hr_Mediana	42	20	53	43	0
29	Ha_Média	0	5	80	0	17
30	Ha_Mediana	0	10	76	0	7
31	Hnr_Média	44	0	20	67	0
32	Hnr_Mediana	42	20	53	43	0
33	Hne_60-400Hz	14	5	66	26	0
34	Hne_400-2000Hz	89	5	13	28	0
35	Hne_2000-5000Hz	5	80	30	0	3
36	Hne_5000-8000Hz	0	0	20	0	100
37	Pvl	73	0	0	54	25
38	Avl	0	0	10	91	0

A figura 5.2 a seguir mostra o gráfico da variação percentual discriminatória das 38 medidas acústicas individualmente com relação a patologia fenda. A tabela 5.3 destaca as medidas acústicas com melhores resultados discriminatórios, ou seja, com pelo menos 50% de acerto discriminatório da patologia fenda.

Figura 5.2. Variação percentual discriminatória para a patologia fenda.

Tabela 5.3. Medidas que melhor discriminaram fenda.

Exec	Parâmetro	Fenda
15	Snl_Media_100-2600Hz	85
16	Snl_Mediana_100-2600Hz	85
35	lne_2000-5000Hz	80
20	Snl_Mediana_5100-8000Hz	70
21	Snl_Media_100-5100Hz	60
23	Snl_Media_100-8000Hz	55
19	Snl_Media_5100-8000Hz	50
27	Hr_Media	50

A figura 5.3 mostra o gráfico da variação percentual discriminatória das 38 medidas acústicas com relação a patologia nódulo. A tabela 5.4 destaca as medidas acústicas com melhores resultados discriminatórios, ou seja, com pelo menos 50% de acerto discriminatório da patologia nódulo.

Figura 5.3. Variação percentual discriminatória para a patologia nódulo.

Tabela 5.4. Medidas que melhor discriminaram nódulo.

Exec	Parâmetro	Nódulo
6	Shimmer_PFF	86
5	Jitter_JF	83
29	Hs_Media	80
30	Hs_Mediana	76
1	Fo	70
33	Hne_60-400Hz	66
17	Snl_Media_2600-5100Hz	63
18	Snl_Mediana_2600-5100Hz	66
20	Snl_Mediana_5100-8000Hz	66
9	Shimmer_JF	63
19	Snl_Media_5100-8000Hz	63
23	Snl_Media_100-8000Hz	63
28	Hr_Mediana	63
32	Hnr_Mediana	63

A figura 5.4 mostra o gráfico da variação percentual discriminatória das medidas acústicas com relação a voz normal. A tabela 5.5 destaca as medidas acústicas com melhores resultados discriminatórios, ou seja, com pelo menos 50% de acerto discriminatório da voz normal.

Figura 5.4. Variação percentual discriminatória para voz normal.

Tabela 5.5. Medidas que melhor discriminaram voz normal.

Exec	Parâmetro	Normal
4	Jitter_JR	91
38	Avi	91
12	Hfpr_60-400Hz	86
7	Shimmer_DPF	82
25	Ra_Media	82
24	Snl_Mediana_100-2600Hz	73
3	Jitter_DPF	71
26	Ra_Mediana	69
2	Jitter_PFF	67
31	Hnr_Media	67
9	Shimmer_JF	63
14	Hfpr_2000-5000Hz	63
37	Pvi	54
11	Hfpr_>6000Hz	52
1	Fo	50
13	Hfpr_400-2000Hz	50

A figura 5.5 mostra o gráfico da variação percentual discriminatória das 38 medidas acústicas com relação a patologia pólipos. A tabela 5.6 destaca as medidas acústicas com melhores resultados discriminatórios, ou seja, com pelo menos 50% de acerto discriminatório da patologia pólipos.

Figura 5.5. Variação percentual discriminatória para patologia pólipos.

Tabela 5.6. Medidas que melhor discriminaram pólipos.

Exec	Parâmetro	Pólipos
36	Line_5000-8000Hz	100
24	Snl_Mediana_100-2600Hz	89
10	Br	64

5.2.2. Etapa multivariada

Nessa etapa testa-se o percentual discriminatório das combinações entre as 38 medidas acústicas, onde combinamos 2a2, 3a3, 4a4, e não continuamos por questões de tamanho de arquivo para análise e de tempo de processamento, e principalmente pelo fato de conseguirmos resultados de 100% de acerto até 4a4. O resultado obtido nessa etapa é muito importante pois fornece a informação das melhores combinações de medidas acústicas para discriminação das 4 patologias analisadas. Na sequência desse tópico, ilustraremos com tabelas as melhores combinações discriminatórias e com gráficos a variação percentual discriminatória das combinações.

• Combinação 2a2

A tabela 5.7 destaca as medidas acústicas combinadas 2a2 que resultaram em pelo menos 80% de acerto discriminatório da patologia edema. O total de combinações foi de 703 e com aproximadamente 10 minutos de processamento.

A figura 5.6 mostra o gráfico da variação percentual discriminatória das combinações 2a2 das medidas acústicas com relação a patologia edema.

Figura 5.6. Combinação 2a2 : Variação percentual discriminatória para patologia edema.

Tabela 4.7. Combinação 2a2: medidas que melhor discriminaram edema.

Combinação	Edema	Parâmetro
695	90	Nne_400-2000Hz--Nne_5000-8000Hz--
696	90	Nne_400-2000Hz--Pvi--
446	89	Snl_Media_100-2800Hz--Nne_400-2000Hz--
633	89	Ra_Mediana--Nne_400-2000Hz--
234	87	Shimmer_DPF--Nne_400-2000Hz--
346	87	Hfpr_>6000Hz--Nne_400-2000Hz--
620	87	Ra_Media--Nne_60-400Hz--
621	87	Ra_Media--Nne_400-2000Hz--
684	87	Hnr_Mediana--Nne_400-2000Hz--
233	85	Shimmer_DPF--Nne_60-400Hz--
263	85	Shimmer_JR--Nne_60-400Hz--
398	85	Hfpr_400-2000Hz--Nne_400-2000Hz--
264	83	Shimmer_JR--Nne_400-2000Hz--
339	83	Hfpr_>6000Hz--Ra_Media--
347	83	Hfpr_>6000Hz--Nne_60-400Hz--
352	83	Hfpr_>6000Hz--Avl--
663	83	Ns_Media--Nne_400-2000Hz--
671	83	Ns_Mediana--Nne_400-2000Hz--
69	81	Jitter_PFF--Nne_400-2000Hz--
340	81	Hfpr_>6000Hz--Ra_Mediana--
343	81	Hfpr_>6000Hz--Ns_Media--

Analisando as tabelas 5.2 e 5.7 referentes a patologia edema, podemos concluir que dentre as medidas acústicas com melhores percentuais de acerto discriminatório em ambas etapas univariada e multivariada, tem-se que as medidas acústicas Nne e Pvi devem figurar nas combinações quando o objetivo for discriminar a patologia Edema, e além disso, Hfpr e Shimmer também mostraram-se importantes.

A tabela 5.8 mostra as medidas acústicas combinados 2a2 que resultaram em pelo menos 90% de acerto discriminatório da patologia fenda. A figura 5.7 mostra o gráfico da variação percentual discriminatória das combinações 2a2 das medidas acústicas com relação a patologia fenda.

Figura 5.7. Combinação 2a2 : Variação percentual discriminatória para patologia fenda.

Tabela 5.8. Combinação 2a2: medidas que melhor discriminaram fenda.

Combinação	Fenda	Parâmetro
47	100	Jitter_PFF--Hfpr_60-400Hz--
48	100	Jitter_PFF--Hfpr_400-2000Hz--
51	100	Jitter_PFF--Snl_Mediana_100-2600Hz--
52	100	Jitter_PFF--Snl_Mediana_2600-6100Hz--
53	100	Jitter_PFF--Snl_Mediana_2600-6100Hz--
54	100	Jitter_PFF--Snl_Mediana_6100-8000Hz--
55	100	Jitter_PFF--Snl_Mediana_6100-8000Hz--
116	100	Jitter_JR--Hfpr_60-400Hz--
116	100	Jitter_JR--Hfpr_2000-6000Hz--
120	100	Jitter_JR--Snl_Mediana_100-2600Hz--
121	100	Jitter_JR--Snl_Mediana_2600-6100Hz--
122	100	Jitter_JR--Snl_Mediana_2000-6100Hz--
123	100	Jitter_JR--Snl_Mediana_6100-8000Hz--
124	100	Jitter_JR--Snl_Mediana_6100-8000Hz--
155	100	Jitter_JF--Snl_Mediana_2600-6100Hz--
212	100	Shimmer_DPF--Hfpr_60-400Hz--
216	100	Shimmer_DPF--Snl_Mediana_100-2600Hz--
217	100	Shimmer_DPF--Snl_Mediana_2600-6100Hz--
218	100	Shimmer_DPF--Snl_Mediana_2600-6100Hz--
219	100	Shimmer_DPF--Snl_Mediana_6100-8000Hz--
220	100	Shimmer_DPF--Snl_Mediana_6100-8000Hz--
242	100	Shimmer_JR--Hfpr_60-400Hz--
247	100	Shimmer_JR--Snl_Mediana_2600-6100Hz--
248	100	Shimmer_JR--Snl_Mediana_2600-6100Hz--
249	100	Shimmer_JR--Snl_Mediana_6100-8000Hz--
250	100	Shimmer_JR--Snl_Mediana_6100-8000Hz--
251	100	Shimmer_JR--Snl_Mediana_100-6100Hz--
276	100	Shimmer_JF--Snl_Mediana_2600-6100Hz--
277	100	Shimmer_JF--Snl_Mediana_2600-6100Hz--
279	100	Shimmer_JF--Snl_Mediana_6100-8000Hz--
320	100	Hfpr_>6000Hz--Hfpr_60-400Hz--
331	100	Hfpr_>6000Hz--Snl_Mediana_2600-6100Hz--
332	100	Hfpr_>6000Hz--Snl_Mediana_2600-6100Hz--
333	100	Hfpr_>6000Hz--Snl_Mediana_6100-8000Hz--
334	100	Hfpr_>6000Hz--Snl_Mediana_6100-8000Hz--
335	100	Hfpr_>6000Hz--Snl_Mediana_100-6100Hz--
336	100	Hfpr_>6000Hz--Snl_Mediana_100-6100Hz--
56	94	Jitter_PFF--Snl_Mediana_100-6100Hz--
90	94	Jitter_DPF--Snl_Mediana_6100-8000Hz--
116	94	Jitter_JR--Hfpr_>6000Hz--
117	94	Jitter_JR--Hfpr_400-2000Hz--
119	94	Jitter_JR--Snl_Mediana_100-2600Hz--
148	94	Jitter_JF--Hfpr_>6000Hz--
149	94	Jitter_JF--Hfpr_60-400Hz--
150	94	Jitter_JF--Hfpr_400-2000Hz--
151	94	Jitter_JF--Hfpr_2000-6000Hz--
154	94	Jitter_JF--Snl_Mediana_2600-6100Hz--
156	94	Jitter_JF--Snl_Mediana_6100-8000Hz--
180	94	Shimmer_PFF--Hfpr_>6000Hz--
181	94	Shimmer_PFF--Hfpr_60-400Hz--
182	94	Shimmer_PFF--Hfpr_400-2000Hz--
183	94	Shimmer_PFF--Hfpr_2000-6000Hz--
186	94	Shimmer_PFF--Snl_Mediana_2600-6100Hz--
187	94	Shimmer_PFF--Snl_Mediana_2600-6100Hz--
188	94	Shimmer_PFF--Snl_Mediana_6100-8000Hz--
189	94	Shimmer_PFF--Snl_Mediana_6100-8000Hz--
216	94	Shimmer_DPF--Snl_Mediana_100-2600Hz--
221	94	Shimmer_DPF--Snl_Mediana_100-6100Hz--
243	94	Shimmer_JR--Hfpr_400-2000Hz--
244	94	Shimmer_JR--Hfpr_2000-6000Hz--
278	94	Shimmer_JF--Snl_Mediana_6100-8000Hz--
280	94	Shimmer_JF--Snl_Mediana_100-6100Hz--
328	94	Hfpr_>6000Hz--Hfpr_2000-6000Hz--

Analisando as tabelas 5.3 e 5.8 referentes a patologia fenda, podemos concluir que dentre as medidas acústicas com melhores percentuais discriminatórios em ambas etapas univariada e multivariada, Snl deve figurar nas combinações das medidas acústicas quando o objetivo for discriminar Fenda, e as medidas acústicas Jitter e Shimmer mostram-se importantes.

A tabela 5.9 mostra as medidas acústicas combinados 2a2 que resultaram em pelo menos 90% de acerto discriminatório para a patologia nódulo. A figura 5.8 mostra o gráfico da variação percentual discriminatória para a patologia nódulo.

Figura 4.8. Combinação 2a2 : Variação percentual discriminatória para patologia nódulo.

Tabela 4.9. Combinação 2a2: medidas que melhor discriminaram nódulo.

Combinação	Nódulo	Parâmetro
41	100	Jitter_PFF--Shimmer_PFF--
178	100	Shimmer_PFF--Shimmer_JF--
193	100	Shimmer_PFF--Snl_Mediana_100-8000Hz--
195	100	Shimmer_PFF--Ra_Mediana--
205	100	Shimmer_PFF--Nne_5000-8000Hz--
206	100	Shimmer_PFF--Pvl--
5	98	Fo--Shimmer_PFF--
21	98	Fo--Snl_Mediana_100-5100Hz--
76	98	Jitter_DPF--Shimmer_PFF--
191	98	Shimmer_PFF--Snl_Mediana_100-5100Hz--
196	98	Shimmer_PFF--Hr_Media--
203	98	Shimmer_PFF--Nne_400-2000Hz--
204	98	Shimmer_PFF--Nne_2000-5000Hz--
22	93	Fo--Snl_Media_100-8000Hz--
489	93	Snl_Media_2600-5100Hz--Nne_400-2000Hz--

Analisando as tabelas 5.4, 5.9 referentes a patologia nódulo, podemos concluir que dentre as medidas acústicas com melhores percentuais discriminatório em ambas etapas univariada e multivariada, Jitter e principalmente Shimmer mostram-se eficientes na discriminação da patologia Nódulo. Além disso, Fo, Snl e Nne também são medidas acústicas importantes para discriminação da patologia nódulo.

A tabela 5.10 abaixo mostra as medidas acústicas combinados 2a2 que resultaram em pelo menos 90% de acerto discriminatório para a voz normal. A figura 5.9 mostra o gráfico da variação percentual discriminatória para voz normal.

Figura 5.9. Combinação 2a2 : Variação percentual discriminatória para voz normal.

Tabela 5.10. Combinação 2a2: medidas que melhor discriminaram voz normal.

Combinação	Normal	Parâmetro
19	100	Fo--Snl_Mediana_5100-8000Hz--
451	100	Snl_Mediana_100-2600Hz--Snl_Media_2600-
659	100	Ns_Media--Ns_Mediana--
689	100	Nne_60-400Hz--Nne_400-2000Hz--
701	100	Nne_5000-8000Hz--Pvt--
2	97	Fo--Jitter_DPF--
514	97	Snl_Media_5100-8000Hz--Snl_Mediana_5100
4	95	Fo--Jitter_JF--
76	95	Jitter_DPF--Jitter_JF--
143	95	Jitter_JF--Shimmer_PFF--
157	95	Jitter_JF--Snl_Mediana_5100-8000Hz--
379	95	Hfpr_400-2000Hz--Hfpr_2000-5000Hz--
666	95	Ns_Mediana--Hnr_Media--
40	93	Jitter_PFF--Jitter_JF--
158	93	Jitter_JF--Snl_Media_100-5100Hz--
484	93	Snl_Mediana_2600-5100Hz--Snl_Media_5100
37	91	Fo--Avi--
38	91	Jitter_PFF--Jitter_DPF--
74	91	Jitter_DPF--Jitter_JR--
108	91	Jitter_DPF--Avi--
109	91	Jitter_JR--Jitter_JF--
110	91	Jitter_JR--Shimmer_PFF--
112	91	Jitter_JR--Shimmer_JR--
113	91	Jitter_JR--Shimmer_JF--
133	91	Jitter_JR--Ns_Media--
134	91	Jitter_JR--Ns_Mediana--
138	91	Jitter_JR--Nne_400-2000Hz--
140	91	Jitter_JR--Nne_5000-8000Hz--
141	91	Jitter_JR--Pvt--
450	91	Snl_Media_100-2600Hz--Avi--
472	91	Snl_Mediana_100-2600Hz--Avi--
493	91	Snl_Media_2600-5100Hz--Avi--
532	91	Snl_Media_5100-8000Hz--Avi--
598	91	Snl_Media_100-8000Hz--Avi--
612	91	Snl_Mediana_100-8000Hz--Avi--
637	91	Ra_Mediana--Avi--
648	91	Hr_Media--Avi--
667	91	Ns_Media--Avi--
675	91	Ns_Mediana--Avi--
693	91	Nne_60-400Hz--Avi--
697	91	Nne_400-2000Hz--Avi--

Analisando as tabelas 5.5 e 5.10 referentes a voz normal, podemos concluir que dentre as medidas acústicas com melhores percentuais discriminatórios em ambas etapas univariada e multivariada, Jitter mostrou-se o mais eficiente seguido de Snl e Nne na discriminação da voz normal.

A tabela 5.11 destaca as medidas acústicas combinadas 2a2 que resultaram em pelo menos 90% de acerto discriminatório para a patologia pólipso. A figura 5.10 mostra o gráfico da variação percentual discriminatória da patologia pólipso.

Figura 5.10. Combinação 2a2 ; Variação percentual discriminatória da patologia pólipso.

Tabela 5.11. Combinação 2 a 2: medidas que melhor discriminaram pólipso.

Combinação	Pólipso	Parâmetro
9	100	Fo--Br--
10	100	Fo--Hfpr_>6000Hz--
11	100	Fo--Hfpr_80-400Hz--
12	100	Fo--Hfpr_400-2000Hz--
14	100	Fo--Snl_Media_100-2600Hz--
15	100	Fo--Snl_Mediana_100-2600Hz--
16	100	Fo--Snl_Media_2600-5100Hz--
241	96	Shimmer_JR--Hfpr_>6000Hz--
321	96	Br--Nne_400-2000Hz--
327	92	Hfpr_>6000Hz--Hfpr_400-2000Hz--

Analisando as tabelas 5.6. e 5.11 referentes a patologia pólipso, concluímos que em ambas etapas univariada e multivariada a medida acústica Br mostrou-se decisiva na discriminação da patologia pólipso, além de Fo, Hfpr e Snl.

• Combinação 3a3

O total de combinações 3a3 para as 38 medidas acústicas foi de 8436, com processamento de aproximadamente 34 minutos. Diante de uma quantidade grande de combinações resultantes, informaremos simplesmente o número de combinações resultante em 100% de acerto discriminatório, dessa forma, não listaremos os resultados como anteriormente.

O número de combinações 3a3 com discriminação em 100% de acerto para as 4 patologias mais a voz normal é o seguinte:

- ✓ 1 para a patologia edema ;
- ✓ 90 para a patologia fenda;
- ✓ 7 para a patologia nódulo ;
- ✓ 17 para voz normal e
- ✓ 60 para a patologia pólipos.

As medidas acústicas decisivas na discriminação em 100% para as 4 patologias mais a voz normal foram :

- ✓ *edema* : S_{nl}_Media_100-2600Hz--S_{nl}_Mediana_100-2600Hz--S_{nl}_Mediana_100-5100Hz
- ✓ *fenda* : Jitter, Shimmer, S_{nl} e Hfpr
- ✓ *nódulo* : Jitter, Shimmer e S_{nl}
- ✓ *normal* : S_{nl}
- ✓ *pólipos* : F₀, Jitter, Shimmer, S_{nl} e Hfpr

• Combinação 4a4

O total de combinações 4a4 para as 38 medidas acústicas foi de 73815, com processamento de aproximadamente 5h08' minutos. Diante de uma quantidade grande de combinações resultantes, informaremos simplesmente o número de combinações resultante em 100% de acerto discriminatório, dessa forma, não listaremos os resultados como anteriormente.

O número de combinações 4a4 com discriminação em 100% de acerto para as 4 patologias mais a voz normal é o seguinte:

- ✓ 60 para a patologia edema ;
- ✓ 225 para a patologia fenda;
- ✓ 32 para a patologia nódulo ;
- ✓ 21 para voz normal e
- ✓ 429 para a patologia pólipos.

As medidas acústicas decisivas na discriminação em 100% para as 4 patologias mais a voz normal foram :

- ✓ *edema* : Snl
- ✓ *fenda* : Jitter, Shimmer, Snl e Hfpr
- ✓ *nódulo* : F_0 , Jitter, Shimmer e Snl
- ✓ *normal* : Shimmer e Snl
- ✓ *pólipo* : todos os parâmetros se mostraram decisivos

4.4. Conclusões

A etapa univariada mostrou que a tentativa de discriminar as patologias simplesmente utilizando a medida incorre num baixo percentual discriminatório em alguns casos.

Na etapa multivariada onde o processo não é mais individual e sim formado pela combinação das medidas acústicas, o percentual discriminatório aumenta consideravelmente. Dessa forma, essa etapa mostra que a combinação das medidas acústicas potencializa o processo discriminatório, sendo que na combinação 3a3 das 38 medidas acústicas foram alcançados 100% de acerto discriminatório para as 4 patologias e para a voz normal.

Assim, concluímos que estendendo o estudo para mais patologias que afetam a laringe, entendemos que os resultados serão proporcionais aos obtidos nesse trabalho.

4.5. Trabalhos futuros

Diante dos resultados alcançados com 100% de acerto discriminatório para as 4 patologias da laringe (edema, fenda, nódulo e pólipo) mais a voz normal (livre de qualquer patologia), o trabalho potencializa estudos no comportamento das medidas acústicas mediante as combinações, inclusive continuando o processo combinatório afim de mais análises. A teoria Bayesiana mostrou-se eficaz na discriminação das 4 patologias da laringe mais a voz normal, e a utilização de outras técnicas como lógica fuzzy, modelos de redes neurais como RBF e até mesmo utilizá-los juntos, também poderiam ser testadas.

O nosso próximo objetivo é o de estruturar esse sistema implementado alcançando um protótipo para a utilização em rede pública, ou seja, um sistema de prevenção, diagnóstico e tratamento de alterações vocais e laríngeas.

6. Bibliografia

Askenfelt, A.; Guffin, J.; Sundberg, J.; Kitzing, P. **A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency.** *Journal of Speech and Hearing Research*, 23, 258-273, 1980.

Awan, S. N.; Frenkel, M. L. **Improvements in estimating the harmonics-to-noise ratio of the voice.** *Journal of Voice*, 8, 255-262, 1994.

BAKEN, R. J. **Clinical Measurement of Speech and Voice.** Allyn and Bacon, a division of Simon and Schuster, Inc., Needham Heights, MA. 1987.

Boone, S.; Daniel, R.; McFarlane, C. "A voz e a terapia vocal". Artes médicas, Porto Alegre, 1994.

CHILDERS, D. G. ; BAE, K. S. **Detection of Laryngeal Function Using Speech and Electroglottographic Data.** IEEE Transactions on Biomedical Engineering, vol. 39, Issue 1, Pages 19-25, January 1992.]

Childers, D. G.; Hicks, D. M.; Moore, G. P.; Eskenazi, L.; Lalwani, A. L. **Electroglottography and vocal fold physiology.** *Journal of Speech and Hearing Research*, 33, 245-254, 1990.

Colton, R. H.; Casper, J. K. **Compreendendo os problemas de voz.** Artes médicas, Porto Alegre, 1996.

Davis, S. B. **Acoustic characteristics of normal and pathological voices.** In Lass, N. J. (Ed.) *Speech and Language: Advances in Basic Research and Practice*, vol. 1, New York Academic Press, 271-335, 1979.

Davis, S. B. **Preliminary results using inverse filtering of speech for automatic evaluation of laryngeal pathology.** *Journal of the Acoustical Society of America*, 58, suppl. 1, p.111-112, 1970.

Davis, S. B.. **Acoustic characteristics of normal and pathological voices.** In Lass, N. J. (Ed.) *Speech and Language: Advances in Basic Research and Practice*, vol. 1, New York Academic Press, p.271-335, 1979.

Deal, R. E.; Emanuel, F. W. **Some waveform and spectral features of vowel roughness.** *Journal of Speech and Hearing Research*, 21, 250-264, 1978.

Deller Jr., J. R.; Proakis, J. G.; Hansen, J. H. L. **Discrete-time Processing of Speech Signals.** Prentice Hall, 908 p., 1987.

Dubnowski, J. J. R.; Schaffer, R. W.; Rabiner, L. R. **Real-time digital hardware pitch detector.** *IEEE Transactions on Acoustics, Speech and Signal Processing*, v.24, p.2-8, 1976.

Emanuel, F. W.; Lively, M. A.; McCoy, J. F. **Spectral noise levels and roughness ratings for vowels produced by males and females.** *Folia Phoniatrica*, 25, 110-120, 1973.

Emanuel, F. W.; Whitehead, R. L. **Harmonic levels and vowel roughness.** *Journal of Speech and Hearing Research*, v.22, p.829-840, 1979.

- Eskenazi, L.; Childers, D. G.; Hicks, D. M. **Acoustic correlates of vocal quality.** *Journal of Speech and Hearing Research*, 33, 298-306, 1990.
- Eustace, C. S., Stemple, J. C. e Lee, L. **Objective measures of voice production in patients complaining of laryngeal fatigue.** *Journal of Voice*, 10, 146-154, 1996.
- Fant, G. **Acoustic Theory of Speech Production.** Mouton's Gravenhague, 1960.
- Fex, B.; Fex, S.; Shiromoto, O.; Hirano, M. **Acoustic analysis of functional dysphonia: before and after voice therapy (accent method).** *Journal of Voice*, 8, 163-167, 1994.
- FROHLICH, M.; MICHAELIS, D.; WERNER, S. H. **Acoustic breathiness measures' in the description of pathologic voices,** Proceedings of the 1998 IEEE International Conference on Acoustics Speech and Signal Processing, Vol.: 2 , Page(s): 937 -940 vol.2. 1998.
- Frokjaer-Jensen, B.; Prytz, S. **Registration of voice quality.** *B & K Technical Review*, 3, 3-17, 1976.
- Fukazawa, T.; El-Assuooty, A.; Honjo, I. **A new index for evaluation of the turbulent noise in pathological voice.** *Journal of the Acoustical Society of America*, v.83, p.1189-1193, 1988.
- Gelfer, M. P. **Fundamental frequency, intensity and vowel selection: effects on measures of phonatory stability.** *Journal of Speech and Hearing Research*, v.38, p.1189-1198, 1995.
- Gubrynowicz, R.; Mikiel, W.; Zarnecki, P. **An acoustic method for the evaluation of the state of the larynx source in case involving pathological changes in the vocal folds.** *Archives of Acoustics*, 5, 3-30, 1980.
- Hall, K. D. **Variations across time in acoustic and electroglottographic measures of phonatory function in women with and without vocal nodules.** *Journal of Speech and Hearing Research*, 38, 783-793, 1995.
- Hecker, M. H. L.; Kreul, E. J. **Description of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency.** *Journal of the Acoustical Society of America*, v.49, n. 4, parte 2, p.1275-1282, 1971.
- Holmes, J. N. **Low-frequency phase distortion of speech recordings.** *Journal of the Acoustical Society of America*, 1975, 747-749, 1975.
- Hillenbrand, J.; Cleveland, R. A.; Erickson, R. L. **Acoustic correlates of breathy vocal quality.** *Journal of Speech and Hearing Research*, 37, 769-778, (1994).
- HIRANO, M.; BLESS, D.M. **Vocal fold vibration.** In: Videostroboscopic examination of the larynx. San Diego, Singular publishing Group Inx., 1993, p.23-36.
- Hiraoka, N.; Kitazoe, Y.; Ueta, H.; Tanaka, S.; Tanabe, M. **Harmonic-intensity analysis of normal and hoarse voices.** *Journal of the Acoustical Society of America*, v.76, p.1648-1651, 1984.

- Honjo, I.; Isshiki, N. **Laryngoscopic and voice characteristics of aged persons.** *Archives of Otolaryngology*, 106, 149-150, 1980.
- Horii, Y. **Fundamental frequency perturbation observed in sustained phonation.** *Journal of Speech and Hearing Research*, v.22, p.5-19, 1979.
- Horii, Y. **Jitter and shimmer differences among sustained vowel phonations.** *Journal of Speech and Hearing Research*, v.25, p.12-14, 1982.
- Horii, Y.. **Fundamental frequency perturbation observed in sustained phonation.** *Journal of Speech and Hearing Research*, v.22, p.5-19, 1979.
- ISSHIKI, N. **Functional surgery of the larynx.** In: Kyoto University Medical School, Dept. of Otolaryngology., ed., 1977. 207 p. [Special report presented at 78th Congress of Japan Otolaryngology Society, Fukouka, 1977].
- ISSHIKI, N. **Phonosurgery: Theory and practice.** Tokyo, Springer-Verlag, 1989.
- ISSHIKI, N. **Recent advances in phonosurgery.** *Folia Phoniatr.* , v.32, p.119-154, 1980.
- ISSHIKI, N. TAIRA, T.; TANABE, M. **Surgical alteration of the vocal pitch.** *The Journal of Otolaryngology*, v.12, p.335-340, 1983.
- ISSHIKI, N.; MORITA, H.; OKAMURA, H. ; HIRAMOTO, M. **Thyroplasty as a new phonosurgical technique.** *Acta Otolaryngol.* V.78, p.451-457, 1974.
- Iwata, S. **Periodicities of pitch perturbations in normal and pathologic larynges.** *Laryngoscope*, v.82, p.87-96, 1972.
- JO, C.; KIM, K.; KIM, D.; WANG, S. **Screening of Pathological Voice from ARS using Neural Networks.** *Proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications*, Firenze-Italy, september 2001.
- JOHNSON, R. A. **Applied Multivariate Statistical Analysis.** Prentice Hall, New Jersey, 1998.
- Kasuya, H.; Ogawa, S.; Mashima, K.; Ebihara, S. **Normalized noise energy as a acoustic measure to evaluate pathologic voice.** *Journal of the Acoustical Society of America*, 80, 1329-1334, 1986.
- Kent, R.D.; Read, C. **The Acoustic Analysis of Speech.** Singular Publishing Group, Inc., San Diego, 1992.
- Keung-Chi, N. G.; Abramson, N. B. **Uncertainty Management In Expert Systems.** *IEEE Expert*, 1990.
- Klingholz, F.; Martin, F. **Quantitative spectral evaluation of shimmer and jitter.** *Journal of Speech and Hearing Research*, 28, 169-174, 1985.
- Klingholz, F; Martin, F. **Quantitative spectral evaluation of shimmer and jitter.** *Journal of Speech and Hearing Research*, 28, 169-174, 1985.
- Koike, Y. **Applications of some acoustic measures for the evaluation of laryngeal dysfunction.** *Studia Phonologica*, 7, 17-23, 1973.
- Koike, Y. **Vowel amplitude modulations in patients with laryngeal diseases.** *Journal of the Acoustical Society of America*, 45, 839-844, 1969.

- Koike, Y., Takahashi, H.; Calcaterra, T. C. **Acoustic measures for detecting laryngeal pathology**. *Journal of Phonetics*, 14, 517-52, 1977.
- Koike, Y.; Markel, J. **Application of inverse filtering for detecting laryngeal pathology**. *Annals of Otolaryngology*, v.84, p.117-124, 1975.
- Kojima, H.; Gould, W. J.; Lambiase, A.; Isshiki, N. **Computer analysis of hoarseness**. *Acta Otolaryngologica*, 89, 547-554, 1980.
- Krom, G. **A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals**. *Journal of Speech and Hearing Research*, v.36, p.254-266, 1993.
- Krom, G. **Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments**. *Journal of Speech and Hearing Research*, 38, 794-811, 1995.
- Kuhl I. **Manual Prático de Laringologia**. Editora da Universidade, Porto Alegre 1982.
- Lieberman, P. **Perturbation in vocal pitch**. *Journal of the Acoustical Society of America*, v.33, p.597-603, 1961.
- Lieberman, P. **Some acoustic measures of the fundamental periodicity of normal and pathologic larynges**. *Journal of the Acoustical Society of America*, v.35, p.344-353, 1963.
- Lively, M. A.; Emanuel, F. W. **Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females**. *Journal of Speech and Hearing Research*, v.13, p.498-502, 1970.
- Maksym, J. N. **Real-time pitch extraction by adaptive prediction of the speech waveform**. *IEEE Transactions on Audio and Electroacoustics*, v.21, p.149-160, 1973.
- MARTINEZ, C. E. ; RUFINER, H. L. **Acoustic Analysis of speech for detection of Laryngeal Pathologies**. Information Technology Applications in Biomedicine.Proceedings. 2000 IEEE EMBS International Conference on , 2000.
- Milenkovic, P.. **Least mean square measures of voice perturbation**. *Journal of Speech and Hearing Research*, v.30, p.529-538, 1987.
- Murry, T.; Doherty, E. T. **Selected acoustic characteristics of pathologic and normal speakers**. *Journal of Speech and Hearing Research*, v.23, p.361-369, 1980.
- Muta, H.; Baer, T.; Wagatsuma, K.; Muraoka, T.; Fukuda, H. **A pitch-synchronous analysis of hoarseness in running speech**. *Journal of the Acoustical Society of America*, 84, 1292-1301, 1988.
- Napagiannis, G.; Mlynarek, A; Massie, R. **Discover The Larynx** . Disponível em < <http://www.projects.mmi.mcgill.ca/larynx/Default.htm> >. Acesso em : outubro de 2003.

Naranjo, N. V.; Lara, E. M.; Rodríguez, I. M.; García, G. C. **High-frequency components of normal and dysphonic voice.** *Journal of Voice*, 8, n. 2, p.157-162, 1994.

Nassar, S.M.; Koehler, C.; Pires, M.M.S. **Uma Abordagem Probabilística para Sistemas Especialistas: Avaliação do Estado Nutricional em Crianças de 0 a 2 anos.** In: III Simpósio Nacional de Informática, Santa Maria, RS, Brasil, 1998.

Nittrouer, S.; McGowan, R.; Milenkovic, P. H.; Beehler, D. **Acoustic measurements of men's and women's voices: a study of context effects and covariation.** *Journal of Speech and Hearing Research*, 33, 761-775, 1990.

NISHIZAWA, N. ; SAWASHIMA, M. ; YONEMOTO, K. **Vocal fold in vocal pitch change.** In: FUJIMURA, O., ed. *Vocal physiology : voice production, mechanisms and fuctions.* New York, Raven Press, Ltd., 1988. V.2, p.75-82

Noll, A. M.. **Cepstrum pitch determination.** *Journal of the Acoustical Society of America*, v.41, p.293-309, 1967.

Noll, A. M.. **Short-time spectrum and "cepstrum" techniques for vocal-pitch detection.** *Journal of the Acoustical Society of America*, v.36, p.296-302, 1964.

Pearl, J. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.** San Mateo, Calif.: Morgan Kaufmann, 1988.

Pinto, N. B.; Titze, I. R. **Unification of perturbation measures in speech signals.** *Journal of the Acoustical Society of America*, 87, 1278-1289, 1990.

Proakis, J. G.; Manolakis, D. G. **Digital Signal Processing: principles, algoritms, and applications.** New Jersey: Prentice Hall, p.968, 1996.

Prosek, R. A.; Montgomery, A. A.; Walden, B. E.; Hawkins, D. B. **An evaluation of residue features as correlates of voice disorders.** *Journal of Communication Disorders*, 20, 105-107, 1987.

Rabiner, L. R.; Schafer, R. W. **Digital Processing of Speech Signals.** New Jersey: Prentice Hall, 512 p., 1978.

Remacle, M.; Trigaux, I. **Characteristics of nodules through the high-resolution frequency analyzer.** *Folia Phoniatrica*, 43, 53-59, 1991.

RITCHINGS, RT.; MCGILLION, M.; MOORE, CJ. **Pathological voice quality assesment using artificial neural networks.** *Proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications*, Firenze-Italy, september, 2001.

ROSA, M.O. **Análise acústica da voz para pré-diagnóstico de patologias da laringe.** Dissertação apresentada à Escola de Engenharia de São Carlos, da Universidade Estadual de São Paulo, como parte dos requisitos mínimos para obtenção do título de mestre em Engenharia Elétrica, 1998.

Ross, M. J.; Shaffer, H. L.; Cohen, A.; Freudberg, R.; Manley, H. J.. **Average magnitude difference function pitch extractor.** *IEEE Transactions on Acoustics, Speech and Signal Processing*, v.5, p.353-362, 1974.

- Rothenberg, M.; Mahshie, J. J. **Monitoring vocal fold abduction through vocal fold contact area.** *Journal of Speech and Hearing Research*, 31, 338-351, 1988.
- Rothenberg, M. **A new inverse-filtering technique for deriving the glottal air flow waveform during voicing.** *Journal of the Acoustical Society of America*, 53, n. 6, 1632-1645, 1973.
- Sanderson, R. J.; Maran, A. G. D. **The quantitative analysis of dysphonia.** *Clinical Otolaryngology and Allied Sciences*, 17, 440-443, 1992.
- Schoentgen, J.; Guchteneere, R. de **Time series analysis of jitter.** *Journal of Phonetics*, 23, 189-201, 1995.
- Schroeder, M. R.. **Period histogram and product spectrum: New methods for fundamental frequency measurement.** *Journal of the Acoustical Society of America*, 43, p. 829-834, 1968.
- Shoji, K.; Regenbogen, E.; Yu, J. D.; Blaugrund, S. M. **H-Index: A new measure of glottal efficiency for the pathologic voice.** *Laryngoscope*, v.102, p.1113-1117, 1992.
- Shoji, K.; Regenbogen, E.; Yu, J. D.; Blaugrund, S. M. **High-frequency power ratio of breathy voice.** *Laryngoscope*, v.102, p.267-271, 1992.
- Sondhi, M. M. **New methods of pitch extraction.** *IEEE Transactions on Audio and Electroacoustics*, 16, 262-266, 1968.
- Sorensen, D.; Horii, Y. **Directional perturbation factors for jitter and for shimmer.** *Journal of Communication Disorders*, v.17, p.143-151, 1984.
- Sorensen, D.; Horii, Y. **Frequency and amplitude perturbation in the voice of female speakers.** *Journal of Communication Disorders*, v.16, p.57-61, 1983.
- STEVENSON, W.J. **Estatística Aplicada à Administração.** São Paulo, Harper e Row do Brasil, 1981.
- SVEC, J.G. ; SCHUTTE, H.K. **Videokimography: high-speed line scanning of vocal fold vibration.** *Journal of voice*, v.10, n.2, p.201-205, 1996.
- TITZE, I. R. ; LIANG, H. **Comparasion of F_0 extraction methods for high-precision voice perturbation measurements.** *Journal Speech and Hearing Research*, 1120-1133, 1993.
- Titze, I. R., Horii, Y.; Scherer, R. C. **Some technical considerations in voice perturbation measurements.** *Journal of Speech and Hearing Research*, v.30, p. 252-260, 1987.
- TITZE, I. R.; HORII, Y.; SCHERER, R. C. **Some technical considerations in voice perturbation measurements.** *Journal of Speech and Hearing Research*, 30, 252-260, 1987
- Titze, I. R.; Liang, H. **Comparasion of F_0 extraction methods for high-precision voice perturbation measurements.** *Journal Speech and Hearing Research*, 1120-1133, 1993.

Trullinger, R. W.; Emanuel, F. W.; Skenes, L. L.; Malpass, J. C. **Spectral noise level measurements used to track voice improvement in one patient.** *Journal of Communication Disorders*, v.21, p.447-457, 1988.

Wendahl, R. W. **Laryngeal analog synthesis of harsh voice quality.** *Folia Phoniatica*, v.15, p.241-250, 1963.

Wendahl, R. W. **Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness.** *Folia Phoniatica*, v.18, p.98-108, 1966.

WILLIAMS, R.G.; ECCLES, R. **A new clinical measure of external laryngeal size which predicts the fundamental frequency of the larynx.** *Acta Otolaryngol.* (Stockh), v.110, p.141-148, 1990.

Winholtz, W. S.; Ramig, L. O. **Vocal tremor analysis with the vocal demodulator.** *Journal of Speech and Hearing Research*, v.35, p.562-573, 1992.

Wolfe, V.; Steinfatt, T. M. **Prediction of vocal severity within and across voice types.** *Journal of Speech and Hearing Research*, v.30, p.230-240, 1987.

YANAGI, E. ; SLAVIT, D.H. ; MCCAFFREY, T.V. **Study of phonation in the excised canine larynx.** *Otolaryngology – Head and Neck Surgery*, v.105, n.4, p.586-595, 1991.

Yumoto, E.; Gould, W. J.; Baer, T. **Harmonics-to-noise ratio as an index of the degree of hoarseness.** *Journal of the Acoustical Society of America*, v.71, p.1544-1549, 1982.

Accardo A. P. ; Enzo M. **An algorithm for the automatic differentiation between the speech of normals and patients with Friedreich's ataxia based on the shorttime fractal dimension** , *Computers in Biology and Medicine*, Vol. 28, Issue 1, Elsevier Pages 75-89, Janeiro 1998.

Andrade, P.J.N. **Sistemas especialistas de apoio ao diagnóstico em medicina. Relações com o Teorema de Bayes e com a Lógica do raciocínio diagnóstico,** *Arq. Bras. Cardilogia*, vol.73, nº6, 1999.

Childers D. G.; Bae K. S. **Detection of Laryngeal Function Using Speech and Electroglottographic Data,** *IEEE Transactions on Biomedical Engineering*, Vol. 39, Issue 1, Pages 19-25, Janeiro, 1992.

Deller Jr.; J. R. ; Anderson, D. J. **Automatic Classification of Laryngeal Dysfunction using the Roots of the Digital Inverse Filter,** *IEEE Trans. on Biom. Eng.* , Vol.BME 27, pp.714-721, 1980.

Elman, J. L., **Finding Structure in Time,** *Cognitive Science*, 14, pp.179-221, 1990.

R. Erenshateyn ; P. Laskov **Human Gesture Recognition using Neural Networks and Multi-Class Encoding,** *Proceedings of Artificial Neural Networks in Engineering, ANNIE'97, St.Louis*, pp. 549-554, 1997.

Frohlich M., Michaelis D., Werner S. H. **Acoustic 'breathiness measures' in the description of pathologic voices**, *Proceedings of the 1998 IEEE International Conference on Acoustics Speech and Signal Processing*, Vol.: 2 , Page(s): 937 -40

Fukunaga, K. **Introduction to Statistical Pattern Recognition**, Academic Press, 1990.

Guoxin K., Jiarui L., Qiang Y., Guoying C. **A method of spectra-analysis for abnormal voice signals**, *Engineering in Medicine and Biology Society*, 1989. *Images of the Twenty-First Century*, *Proceedings of the Annual International Conference of the IEEE Engineering in* , 1989.

Hadjitodorov S.; Boyanov B.; Teston B. **Laryngeal pathology detection by means of class-specific neural maps**, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 4 Issue: 1 , Page(s): 68 -73, March 2000.

Hopfield, J. J. ; W. Tank, **Computing with Neural Circuits: A Model**, *Science*, Vol. 233, pp.625-633, Aug. 1986.

Horii, Y. **Fundamental Frequency Perturbation Observed In Sustained Phonation**. *Journal Of Speech And Hearing Research*, V.22, P.5-19, 1979.

Imagawa, K.; Lu, S.; Igi, S. **Color-Based Hands Tracking System for Sign Language Recognition**, *Proceedings of 3th International Conference on Automatic Face and Gesture Recognition*, pp. 462-467, Nara, Japan, 1998.

Isshiki, N. **Functional Surgery of the Larynx**. In: *Kyoto University Medical School, Dept. Of Otolaryngology*, Ed., 1977. 207 P. [Special Report Presented At 78th Congress Of Japan Otolaryngology Society, Fukouka, 1977].

Isshiki, N. **Phonosurgery: Theory And Practice**. Tokyo, Springer-Verlag, 1989.

Isshiki, N. **Recent Advances In Phonosurgery**. *Folia Phoniatr.* , V.32, P.119-154, 1980.

Isshiki, N. Taira, T.; Tanabe, M. **Surgical Alteration Of The Vocal Pitch**. *The Journal Of Otolaryngology*, V.12, P.335-340, 1983.

Isshiki, N.; Morita, H.; Okamura, H. ; Hiramoto, M. **Thyroplasty As A New Phonosurgical Technique**. *Acta Otolaryngol.* V.78, P.451-457, 1974.

Jo C.; Kim K.; Kim D.; Wang S. **Screening of Pathological Voice from ARS using Neural Networks**, *proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications*, Firenze, Itália, Setembro 2001.

Koike, Y., Takahashi, H.; Calcaterra, T. C. **Acoustic measures for detecting laryngeal pathology**. *Journal of Phonetics*, 14, 517-52, 1977.

Lieberman, P. **Perturbation in vocal pitch**. *Journal of the Acoustical Society of America*, v.33, p.597-603, 1961.

Liang, R.H., M. Ouhyoung, **A Sign Language Recognition System Using Hidden Markov Models and Context Sensitive Search**, *Proc. of ACM International Symposium on Virtual Reality and Software Technology*, pp. 59-66, Hong-Kong, 1996.

Martinez C. E., Rufiner H. L. **Acoustic Analysis of Speech for Detection of Laryngeal Pathologies**, *Information Technology Applications in Biomedicine*, 2000.

Niedzielska G.; Glijer E. ; Niedzielski A. **Acoustic analysis of voice in children with noduli vocals**, *International Journal of Pediatric Otorhinolaryngology*, Vol. 60, Issue 2, Elsevier Pages 119-122, 20 August 2001.

Nishizawa, N. ; Sawashima, M. ; Yonemoto, K. **Vocal Fold In Vocal Pitch Change**. In: Fujimura, O., Ed. *Vocal Physiology : Voice Production, Mechanisms And Functions*. New York, Raven Press, Ltd., 1988. V.2, P.75-82

Ritchings RT., McGillion M., Moore CJ. **Pathological voice quality assesment using artificial neural networks**, *Proceedings of 2nd International workshop on Models and Analysis of vocal emissions for biomedical applications*, Firenze, Itália, 2001.

Rontal, E., Rontal, M. and Rolnick, M. I. **Objective Evaluation of Vocal Pathology using Voice Spectrography**. *Ann. Otol. Rhinol. Laryngol.*, Vol.84, pp.662-671. 1975.

Rosa M. de O.; Greller M., Carvalho C.P.L.F. A. **Signal processing and statistical procedures to identify laryngeal pathologies**, in *Electronics, Circuits and Systems*, 1999 6th IEEE International Conference on IEEE Circuits and Systems Society, 1999.

Rosa M. de O.; Pereira C. J. ;C.P.L.F. Carvalho A. **Evaluation of Neural Classifiers using Statistic Methods for Identification of Laryngeal Pathologies**. *Proceedings of the Vth Brazilian Symposium on Neural Networks* , 1998.

Rosa M. de O.; Pereira J. C.; Grellet M. **Adaptive estimation of residue signal for voice pathology diagnosis**, *IEEE Transactions on Biomedical Engineering*, Vol. 47, Issue 1, Pages 96-104, January 2000.

Shoji, K.; Regenbogen, E.; Yu, J. D; Blaugrund, S. M. **High-Frequency Power Ratio Of Breathly Voice**. *Laryngoscope*, V.102, P.267-271, 1992.

Sklar, B. **A primer on turbo code concepts**. *IEEE Communication Magazine*, pp. 94-101, 1997.

Svec, J.G. ; Schutte, H.K. **Videokimography: High-Speed Line Scanning Of Vocal Fold Vibration**. *Journal Of Voice*, V.10, N.2, P.201-205, 1996.

Starner, T. , Pentland, A. ***Visual Recognition of American Sign Language Using Hidden Markov Models***, International Workshop on Automatic Face and Gesture Recognition, Zurique, Suíça, 1995.

Titze, I. R. ; Liang, H. ***Comparasion Of F₀ Extraction Methods For High-Precision Voice Perturbation Measurements***. *Journal Speech And Hearing Research*, 1120-1133, 1993.

Titze, I. R.; Horii, Y.; Scherer, R. C. ***Some Technical Considerations In Voice Perturbation Measurements*** *Journal Of Speech And Hearing Research*, 30, 252-260, 1987.

Williams, R.G.; Eccles, R. ***A New Clinical Measure Of External Laryngeal Size Which Predicts The Fundamental Frequency Of The Larynx***. *Acta Otolaryngol.* (Stockh), V.110, P.141-148, 1990.

Yanagi, E. ; Slavit, D.H. ; Mccaffrey, T.V. ***Study Of Phonation In The Excised Canine Larynx***. *Otolaryngology – Head And Neck Surgery*, V.105, N.4, P.586-595, 1991.