Campus de São Carlos

**ESTIMAÇÃO DO SINAL GLOTTAL PARA PADRÕES
ACÚSTICOS DE DOENÇAS DA LARINGE**

Autora: Aparecida de Cássia Guerra

Orientador: Prof. Dr. José Carlos Pereira

UNIVERSIDADE DE SÃO PAULO

**ESCOLA DE ENGENHARIA
DE SÃO CARLOS**

Universidade de São Paulo
Escola de Engenharia de São Carlos

Estimação do Sinal Glotal para Padrões Acústicos de Doenças da Laringe

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

Orientada: Aparecida de Cássia Guerra
Orientador: Prof. Dr. José Carlos Pereira

São Carlos - 2005

DEDALUS - Acervo - EESC

31100052278

Class.	TESE-EESC
Cott.	2635
Tombo	T242/05
Sysno	1461119

Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca – EESC/USP

G934e Guerra, Aparecida de Cássia
Estimação do sinal glotal para padrões acústicos
de doenças da laringe / Aparecida de Cássia Guerra. --
São Carlos, 2005.

Tese (Doutorado) -- Escola de Engenharia de São
Carlos-Universidade de São Paulo, 2005.
Área: Engenharia Elétrica.
Orientador: Prof. Dr. José Carlos Pereira.

1. Sinal de voz. 2. Pulso glotal. 3. Modelo LF.
4. Processamento digital de voz. 5. Pregas vocais.
6. Pulso derivativo global - modelo. I. Título.

FOLHA DE JULGAMENTO

Candidata: Licenciada **APARECIDA DE CASSIA GUERRA**

Tese defendida e julgada em 03-05-2005 perante a Comissão Julgadora:

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

APROVADA

Prof. Dr. **MARCOS GRELLET**
(Faculdade de Medicina de Ribeirão Preto/USP)

Aprovada

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

APROVADA

Prof. Dr. **CARLOS DIAS MACIEL**
(Escola de Engenharia de São Carlos/USP)

APROVADA

Prof. Dra. **MARIA STEIA VELUDO DE PAIVA**
(Escola de Engenharia de São Carlos/USP)

APROVADA

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

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

DEDICATÓRIA

*À SANTA CATARINA,
AOS MEUS PAIS (ARMANDO E ROSA)*

AGRADECIMENTO

Um agradecimento especial ao meu orientador e amigo, Prof. Tit. José Carlos Pereira, que sempre esteve a meu lado, incentivando e apoiando em todos os momentos.

Aos professores Maria Stela e Maciel que fizeram parte da minha banca de qualificação e me ajudaram muito.

À todos os funcionários do departamento de Engenharia Elétrica, em especial a Marisa e Denise que sempre se mostraram disponíveis, atenciosas e que com certeza, sem a ajuda delas, seria impossível o andamento e término do trabalho e ao Zé Carlos, pela simpatia e dedicação no decorrer destes anos.

À todos amigos do Lim, em especial ao Mori, que sempre esteve presente e a Giovana.

Aos amigos e funcionários da Bioengenharia., em especial Janete e Nelson.

Às minhas amigas fonoaudiólogas que contribuíram muito, em especial a Jussara, Luciana, Ivani, Raquel e Eugênia.

À minha família, meus irmãos, cunhadas, meus queridos sobrinhos (Rafael, Fernanda, Lívia, Priscila e Letícia), que são o incentivo, a alegria e a esperança para a concretização dos sonhos e meus pais em especial, que sempre me apoiaram e estiveram ao meu lado, pois sem eles minha vida não teria sentido.

À minhas amigas, Luciana, Andréa, Gláucia, Solange e Mara que trouxeram momentos especiais nestes anos.

Ao meu noivo Clodoaldo e sua família, principalmente pela paciência , apoio, pelo amor e pelos momentos especiais que foram um equilíbrio em minha vida.

Enfim, a todos que de alguma forma fizeram parte da minha vida nestes anos.

Este trabalho foi financiado pela FAPESP.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas.	III
Resumo	i
Abstract	iii
1 Introdução.....	1
2 Revisão da Literatura	3
2.1 O sub-sistema glotal	5
2.2 Modelo Acústico da Produção de fala e alguns aspectos espectrais.....	7
2.3 Patologias do Trato Vocal.....	12
3 Técnicas Acústicas para Análise de Voz.....	17
3.1 Sinais Diretos	17
3.2 Sinais Indiretos	17
3.3 Fonte.....	18
3.4 Modelo Fonte-Filtro	19
3.5 Modelos Glotais para Processamento de Voz	23
3.5.1 Modelos Glotais Paramétricos Não Interativos	24
3.5.1.1 Considerações Gerais.....	31
4 Modelo LF.....	32
4.1 Modelando o Pulso Derivativo Glotal.....	33
4.2 Métodos para Estimar o Pulso Glotal Derivativo.....	35
4.2.1 Método da Autocorrelação	35
4.2.2 Escolha da Ordem de Predição	36
4.2.3 Escolha da Janela	37
4.2.4 Pré-ênfase.....	37
4.2.5 Determinação da Região da Fase Fechada Através da Modulação Formante.....	37
4.3 Avaliando o Modelo LF a Partir do Pulso Glotal Derivativo	39

	4.3.1	Parâmetros Acústicos.....	39
	4.3.2	Parâmetros Acústicos Relacionados ao Pulso Derivativo Glotal.....	41
	4.3.3	Parâmetros Propostos.....	49
	4.4	Análise Estatística	51
5		Resultados	52
	5.1	Cálculo do Pulso Glotal Derivativo, Através da Filtragem Inversa.....	52
	5.2	Análise dos Parâmetros Acústicos	53
	5.2.1	Materiais e Métodos.....	54
6		Conclusões.....	89
7		Referências Bibliográficas	90

LISTA DE FIGURAS

Figura 2.1: Aparelho fonador humano	3
Figura 2.2: Modelo do trato vocal.....	4
Figura 2.3: Glote	5
Figura 2.4: (a) Localização da Laringe de perfil	6
Figura 2.4: (b) Localização da Laringe de frontal	6
Figura 2.5: a) partes da Laringe (frontal e perfil)	6
Figura 2.5: b) Imagens laringoscópicas.....	6
Figura 2.6: vibração das pregas vocais.....	7
Figura 2.7: Modelo para síntese da voz.....	7
Figura 2.8: a) Representação esquemática do perfil do sistema vocal de cada vogal	9
Figura 2.8: b) formas de ondas típicas no tempo	9
Figura 2.8: c) espectro correspondente para cada vogal	9
Figura 2.9: Espectro do som da laringe.	10
Figura 2.10: a) tipos de fonações e correspondentes ondas glotais referentes a voz tensa	11
Figura 2.10: b) tipos de fonações e correspondentes ondas glotais referentes a voz normal	11
Figura 2.10: c) tipos de fonações e correspondentes ondas glotais referentes a voz sopreosa.....	11
Figura 2.11: Pregas vocais e patologias	14
Figura 3.1 Modelo de Produção de Fala.....	18
Figura 3.2: modelo representativo do trato vocal.....	20
Figura 3.3 : Trecho Periódico do Sinal de Voz.....	21
Figura 3.4: Esquema simplificado do sistema de fonação humano.....	22
Figura 3.5: Modelo fonte-filtro.....	22
Figura 3.6: Modelo polinomial de Rosenberg.....	25
Figura 3.7: Modelo trigonométrico de Rosenberg.....	25
Figura 3.8 : Modelo derivativo do pulso glotal LF.....	27
Figura 3.9 : Modelo derivativo glotal Fujisaki e Ljungqvist.....	29
Figura 3.10: Modelo glotal de Hedelin.....	30
Figura 3.11: Pulso glotal encontrado pelo filtro inverso.....	31
Figura 3.12: Pulso glotal derivativo real.....	31
Figura 4.1: Modelo fonte-filtro com sinal derivativo modelo pelo modelo LF.....	32
Figura 4.2: Pulso Glotal e Sinal Derivativo Glotal (Modelo LF)	34
Figura 4.3: Valores de EE: a) pulso derivativo glotal com um pequeno valor EE.....	41
Figura 4.3: Valores de EE: b) pulso glotal derivativo com um grande valor de EE.	41
Figura 4.4: a) Valores esquematizados de RK: pulso derivativo glotal com um pequeno valor de RK. .	42
Figura 4.4: b) Valores esquematizados de RK: pulso derivativo glotal com um valor RK grande.....	43
Figura 4.5: a)Valores de OQ esquematizados: pulso derivativo glotal com um pequeno valor de OQ..	45
Figura 4.5: b)Valores de OQ esquematizados: pulso glotal derivativo com um maior valor para OQ..	45
Figura 4.6: a) Pulso Glotal	47
Figura 4.6: b) Pulso Glotal Derivativo.....	47
Figura 4.7: a) Relações esquematizadas TRp.....	50
Figura 4.7: a) Relações esquematizadas TRs.....	50
Figura 5.1: Cálculo do pulso glotal derivativo.....	53
figura 5. 2: Vozes femininas normais.....	56
Figura 5.3: Vozes masculinas normais.....	57
figura 5.4 : Pulsos derivativos femininos.....	58
Figura 5.5: Pulsos derivativos masculinos.....	59
Figura 5.6: Parâmetros relacionados a vozes femininas e masculinas normais.....	63
Figura 5.7: Comparação das médias dos parâmetros de forma para vozes normais, com nódulo e edema de Reinke.	68
Figura 5.8: Comparação das médias dos parâmetros de banda para vozes normais, com nódulo e edema de Reinke.	74
Figura 5.9: Comparação das médias do parâmetro ece vozes normais, com nódulo edema de Reinke..	78

Figuras 5.10: Comparação das médias dos parâmetros TR_p e TR_s , para vozes normais, com nódulo e edema de Reinke. 82

A

B

C

D

E

F

LISTA DE TABELAS

Tabela 2.1: Patologias e sintomas acústicos.....	13
Tabela 3.1: Parâmetros referentes aos modelos polinomial e trigonométrico.....	25
Tabela 3.2: Valores de alguns parâmetros do modelo LF.....	27
Tabela 3.3: Valores para os modelos Fujisaki e Ljungqvist.....	29
Tabela 3.4: Parâmetros do modelo definido por Hedelin.....	30
Tabela 3.5: Resumo dos modelos paramétricos não-interativos.....	31
Tabela 4.1: Parâmetros do modelo LF para o sinal derivativo da glote.....	33
Tabela 5.1: Parâmetros Acústicos.....	61
Tabela 5.2: Médias dos parâmetros para vozes normais.....	62
Tabela 5.3: Relação dos tipos de vozes e quantidade de casos.....	65
Tabela 5.4: Parâmetros relacionados as vozes normais, nódulo e Reinke.	66
Tabela 5.5: ANOVA para o parâmetro e_0	69
Tabela 5.6: ANOVA para o parâmetro α	70
Tabela 5.7: ANOVA para o parâmetro β	70
Tabela 5.8: ANOVA para o parâmetro w_g	71
Tabela 5.9: Teste-t para o parâmetro w_g	72
Tabela 5.10: Comparação dos resultados dos parâmetros e_0 , α , β , w_g	73
Tabela 5.11: ANOVA para o parâmetro B_1	75
Tabela 5.12: ANOVA para o parâmetro B_2	75
Tabela 5.13: ANOVA para o parâmetro B_1B_2	76
Tabela 5.14: ANOVA para o parâmetro B_2B_1	77
Tabela 5.15: Comparação dos resultados dos parâmetros B_1 , B_2 , B_1/B_2 e B_2/B_1	78
Tabela 5.16: ANOVA para o parâmetro e_{ce}	79
Tabela 5.17: Teste-t para o parâmetro w_g	80
Tabela 5.18: Comparação dos resultados do parâmetro e_{ce}	81
Tabela 5.19: ANOVA para o parâmetro TR_p	83
Tabela 5.20: ANOVA para o parâmetro TR_s	83
Tabela 5.21: Teste-t para o parâmetro TR_p	85
Tabela 5.22: Teste-t para o parâmetro TR_s	86
Tabela 5.23: Comparação dos resultados dos parâmetros TR_p e TR_s . normais, com nódulo e com edema de Reinke.	87
Tabela 5.24: Possíveis resultados para a discriminação entre vozes normais, com nódulo e com Edema de Reinke.	88

RESUMO

Muitas pesquisas tem sido feitas em processamento digital de sinais (PDS) na tentativa de se avaliar o sinal de fala para diagnosticar doenças da laringe. Medidas acústicas têm sido propostas de forma a avaliar indiretamente o trato glotal por meio do sinal de voz coletado através de microfone convencional. Para isso, o modelo paramétrico *Liljencrants-Fant* (LF) foi desenvolvido para representar o sinal glotal em condições normais e patológicas. Tais parâmetros apresentam vantagens sobre medidas acústicas por possuírem características fisiológicas reais das pregas vocais. Assim, podendo ser empregados para identificação de doenças da laringe.

Além da estimação dos parâmetros LF, no domínio do tempo (parâmetros T), a forma de onda da derivativa glotal também pôde ser quantificada através dos parâmetros identificados na literatura por parâmetros R (R_d , R_a , R_k , e R_g), parâmetros quocientes Q (SQ, OQ, CQ, AQ, e NAQ), parâmetros B1 e B2 que são as extensões de bandas do pulso derivativo LF, e o parâmetro *ece*, que relaciona os parâmetros β e T_a . Os parâmetros B1 e B2 e *ece* apesar de serem propostos na literatura, não são encontrados resultados referentes a essas duas medidas. Os resultados mostraram que os parâmetros B não foram confiáveis na discriminação entre as vozes, por outro lado, o parâmetro *ece* mostrou-se ser opção na discriminação entre as vozes normais, nódulo e Reinke.

O objetivo deste trabalho é direcionar a atenção sobre o sinal glotal, estimando-o automaticamente mediante técnicas de PDS aplicadas ao sinal de fala, visando extrair parâmetros que identifiquem as condições normais e patológicas da laringe.

Por fim foram propostos os parâmetros TR_p e TR_s , visando dissociar os efeitos de primeira ordem dos de ordem superior na fase de retorno do pulso glotal com a finalidade de estimar a real não-linearidade do sub-sistema glotal, retratando as condições fisiológicas do movimento das pregas vocais. Com um nível de confiança de 95%, o parâmetro de primeira ordem (TR_p) é efetivo na discriminação do Edema de Reinke, porém mostrou-se ineficaz na detecção do nódulo. Em relação ao parâmetro de ordem superior, conclui-se que o TR_s é um excelente detetor de vozes patológicas (nódulo e Edema de Reinke), porém não é capaz de discriminar as patologias.

Palavras-chaves: sinal de voz, pulso glotal, modelo LF, processamento digital de voz, pregas vocais, pulso derivativo glotal - modelo.

ABSTRACT

Many researches has been conducted in digital signal processing (DSP) attempting to evaluate the physiological conditions of larynx. Acoustical parameters have been proposed to evaluate the glotal tract from voice signal. One technique proposed is the *Liljencrants-Fant* model (LF) developed to represent normal and pathologic conditions of the larynx. Those parameters compare favourably as far as real physiologic characteristic of vocal folds is concerned. So, a primary use of the model is the larynx pathologic identification.

Beyond LF parameters estimation, (T parameters in the time domain), the waveform of glotal pulse derivative also can be quantified through, R parameters (R_d , R_a , R_k and R_g), quotient parameters (SQ, OQ, CQ, AQ and NAQ), B parameters (B_1 and B_2) that are band extension of the LF glotal pulse derivative and the *ece* parameter that in fact, is a relationship between β and T. Although proposed in the literature, no results are found, related to B and *ece* parameters. Our founds show that B parameters do not present good results in voice discrimination, however, *ece* parameter seems to be good option to discriminate normal voice, nodulo and Reinke edema.

The main purpose of this work is to estimate the glotal signal from the voice signal using DSP techniques in order to obtain parameters that identifies the physiological larynx condition.

In order to estimated the shape of return phase of glotal pulse, two parameters have been proposed in this work. The first one evaluates the pulse (TR_p), in other words, the first order component of the return phase. The second is responsible to evaluate superior orders components of the return phase (TR_s), i e, the non-linear component of the glotal pulse.

With 95% of confidence level, TR_p is effective in Reinke Edema discrimination however it is inefficient for nodule e dection.

By the other hand, the TR_s parameter works well to detect pathologic voice however is unable to discriminated them.

Key words: voice signal, glottal pulse, LF model, digital signal processing, vocal folds, glottal derivative pulse - model .

1 INTRODUÇÃO

Os sinais de fala são compostos por uma seqüência de sons ou segmentos fonéticos, regulados pelas regras da linguagem e pelas características do orador. Para entender, sintetizar, reconhecer ou de um modo geral, processar os sinais de fala, é necessário entender o mecanismo da sua produção.

Muitas pesquisas tem sido feitas em processamento digital de sinais na tentativa de se avaliar o sinal de fala para diagnosticar doenças da laringe. Medidas acústicas tem sido propostas de forma a avaliar indiretamente o sistema fonador, constituído do gerador glotal e do trato vocal, por meio do sinal de voz coletado através de microfone convencional. O gerador glotal é responsável pelos pulsos de ar que excitam o trato vocal resultando na emissão de som pelos lábios. Estes pulsos constituem o sinal glotal. Segundo [ANANTHPADMANABHA *et al*/82], o sinal glotal $g(t)$ é melhor analisado quando sua derivada $[g'(t)]$ é estimada e pode ser caracterizada por um conjunto de sete parâmetros. Com estes parâmetros calculados para o sinal derivativo $g'(t)$ algumas considerações devem ser aplicadas para emprestar maior robustez a estas estimativas paramétricas. O acoplamento, por exemplo, entre os tratos glotal e supraglotal (trato vocal) produz pequenas ondulações (ripple) no sinal, que devem ser modeladas adequadamente. Alguns fonemas sustentados, particularmente o / i / provocando um aumento da pressão supraglotal que afeta diretamente a vibração das pregas vocais e, desta forma, modifica o decaimento da forma de onda de $g'(t)$.

O ponto mais importante na estimativa do sinal derivativo está na suposição de que durante o fechamento das pregas vocais, o sinal de voz somente é influenciado pelo trato vocal. Portanto, a característica do filtro pode ser determinada otimamente na fase de fechamento glotal. Deste modo, a filtragem inversa permite estimar o sinal derivativo glotal de forma robusta. Sem esta referência, tais estimativas exigem um estudo mais aprofundado do acoplamento entre estes dois sistemas

Uma vez determinado o modelo, são estimados parâmetros característicos de $g'(t)$ visando a comparação com os parâmetros do modelo Liljencrants-Fant (LF). Como as doenças da laringe impõem não-linearidades que devem ser investigadas, a inclusão de

mais parâmetros, se torna necessária afim de definir melhor a curva paramétrica do sinal derivativo.

Determinados estes parâmetros para as doenças da laringe, pode-se então sintetizar o sinal de voz uma vez que a função de transferência do trato supraglotal já é conhecida. Esta síntese permite que se valide inclusive modelos biomecânicos da laringe [ISHIZAKA/1972], [QI/1994].

Devido à correlação destes parâmetros com a dinâmica do movimento das pregas vocais, estes podem ser empregados, inclusive, para determinação automática de doenças da laringe.

2 REVISÃO DA LITERATURA

O som é uma onda mecânica que se movimenta em sentido longitudinal à direção de propagação. As sucessivas compressões e dilatações do ar, ou seja, alterações de pressão, se propagam até influenciar o ouvido (ou um transdutor de pressão), onde é adequadamente decodificado (traduzido para impulsos elétricos) e interpretado.

Uma forma inteligível do som é a voz, que é produzida a partir de um conjunto de órgãos que apresentam como função primária possibilitar a respiração humana e auxiliar nos mecanismos da deglutição (Fig.2.1). A produção de fonemas (representação dos sons produzidos, que formam as linguagens propriamente ditas) está sujeita a limitações fisiológicas impostas por estes órgãos dada as exigências respiratórias do ser humano.

Figura 2.1: Aparelho fonador humano (Adaptado de [DELLER *et al*/1993]).

Para facilitar o estudo do sistema de produção de fala, o mesmo é normalmente reduzido à três subgrupos distintos, a saber: subglotal, glotal e supraglotal. Dependendo do tipo de som emitido cada subgrupo contribui para a formação dos sinais específicos. Além

disto, permite um modelamento convenientemente localizado das diversas características que contribuem na emissão sonora. Por exemplo, o primeiro grupo fornece a energia para estabelecimento do fluxo aéreo, este iniciando nos pulmões e encerrando no final da traquéia (*conus elasticus*), recebe o nome de subgrupo de respiração. Situados na laringe, encontram-se os órgãos do segundo grupo, responsáveis pela produção de uma onda quase-periódica mediante vibração das pregas vocais em um movimento por variação de pressão (segundo a lei de Bernoulli) e pela elasticidade do músculo vocálico. Um esquema simplificado do sistema vocal é apresentado na Fig. 2.2, [FLANAGAN *et al.*/1970], onde o trato vocal é excitado pelo ar expelido dos pulmões por ação de uma força muscular, e modulado pelo sistema massa-mola correspondente às pregas vocais.

Figura 2.2: Modelo do trato vocal. [RABINER *et al.*/1978]

O período entre duas aberturas (ou fechamento) sucessivos das pregas vocais, determina a frequência fundamental (ou *período pitch*). Com a responsabilidade de produzir os fonemas de uma língua encontra-se o sistema ressonador, constituindo assim o último subgrupo de órgãos do aparelho fonador.

Apesar da aparente separação induzir a total independência entre os subgrupos, deve-se manter em mente que existe o acoplamento entre estes elementos pois a inexistência de um dos grupos automaticamente impede a produção adequada da fala. Além disto, a dinâmica da fala envolve a movimentação simultânea de diversas estruturas do trato vocal (Fig. 2.4); isto leva a se considerar as redundâncias do trato vocal quanto à geração de

voz, pois existe um conjunto de configurações dos órgãos que produzem os mesmos fonemas e coarticulações.

2.1 O SUB-SISTEMA GLOTAL

A voz ou vocalização é produzida pelo fluxo de ar expiratório advindo dos pulmões passando pela laringe (a Fig. 2.4 (a) e (b) mostra a localização da laringe, perfil e frontal), movimentando as pregas vocais aproximadas e colocando-as em vibração. Sua frequência de vibração é determinada por quão elásticas ou tensas estas estão, o que é indicado por seu alongamento e espessura, quanto mais rápido elas vibram (frequência mais alta), mais aguda é a altura percebida da voz. Quão vigorosamente as pregas vocais vibram, originado pela pressão de ar abaixo delas e pela velocidade do ar passando entre elas, determina a intensidade da vocalização ou a intensidade percebida da voz. Embora problemas de respiração possam contribuir para uma voz inadequada, a maioria dos problemas vocais parece resultar da aproximação falha das pregas vocais.

Figura 2.3: Glote

O espaço entre elas é conhecido como *glote* (Fig. 2.3) e sua abertura pode ser controlada movimentando-se as cartilagens tireóide e aritenóide, que podem ser visualizadas nas Fig. 2.5 a) e b) que mostram as partes da laringe e seu funcionamento.

(a)

(b)

Figura 2.4: Localização da Laringe (a) perfil e (b) frontal

a) Partes que compõem a Laringe (vista frontal e perfil)

b) 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;

6) Comissura posterior

Figura 2.5: a) partes da Laringe (frontal e perfil) , b) Imagens laringoscópicas

Nas cartilagens aritenóide e tiróide o fluxo contínuo de ar dos pulmões é geralmente transformado em vibrações rápidas e audíveis quando falamos. Isso é feito pelo fechamento das pregas vocais, que causa um aumento gradativo da pressão na região infra-glótica, que acaba por fazer com que estas se abram repentinamente, liberando a pressão, para então tornarem a se fechar. Este processo produz uma sequência de pulsos cuja

freqüência é controlada pela pressão do ar e pela tensão e comprimento das pregas vocais, Fig. 2.6.

¹Figura 2.6: vibração das pregas vocais

1. <http://www.umanitoba.ca/faculties/arts/linguistics/russell/138/sec5/phonatio.htm>

2.2 MODELO ACÚSTICO DA PRODUÇÃO DE FALA E ALGUNS ASPECTOS ESPECTRAIS

Há dois tipos elementares de sons na fala: sons vozeados e não-vozeados, os sons vozeados são produzidos pelo ar atravessando a glote ou a abertura das pregas vocais. O som produzido pela laringe é chamado de vozeado ou fonação. Ao contrário, os sons não-vozeados são gerados por uma constrição em algum ponto ao longo do sistema vocal. Ao se forçar o ar através da constrição produz-se turbulência [DELLER/93]. A Fig. 2.7 mostra o diagrama da produção dos sons vozeados e não vozeados.

Figura 2.7: Modelo para síntese da voz

A característica espectral da fala é variante no tempo (ou não-estacionária), pois o sistema físico muda no tempo. No espectro das vogais existem regiões de pico (ou de ênfase) que são chamados ressonantes. Estes picos ressonantes são conseqüências do subsistema articulatório que forma diversas cavidades acústicas no sistema vocal. A localização dos picos ressonantes no domínio frequência, depende das dimensões físicas e da forma do sistema vocal. Cada sistema vocal é caracterizado por um conjunto de frequências ressonantes. Do ponto de vista de modelagem do sistema, o subsistema articulatório determina as propriedades do filtro da fala. Como estes picos ressonantes “formam” o espectro de uma forma geral, cientistas da fala referem-se a eles como formantes. Os Formantes do espectro são referidos como F1, F2 F3..., começando pela menor frequência. Em princípio, há infinitos formantes, mas na prática, encontra-se de três a cinco após a conversão analógica/digital apropriada do sinal [DEL/93]. As vogais podem ser distinguidas pela localização das formantes no espectro. Na Fig. 2.8, encontram-se as cinco vogais /a/, /e/, /i/, /o/ e /u/, sendo que a) corresponde a representação esquemática do perfil do sistema vocal de cada vogal b) as formas de ondas típicas no tempo e c) o espectro correspondente do trato vocal para cada vogal.

Figura 2.8: a) Representação esquemática do perfil do sistema vocal de cada vogal b) formas de ondas típicas no tempo c) espectro correspondente para cada vogal [DELLER/93].

Em geral, o formante F1 varia muito com a altura da língua (acima ou abaixo da boca) ou a abertura da mandíbula e F2 varia com a posição da língua (frente ou atrás). 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 F1) mantendo, porém, a frequência fundamental. A vogal que apresenta a maior frequência F1 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/92].

A fonte de energia acústica das vogais é tipicamente gerada pela vibração das pregas vocais. A forma de onda do som da laringe idealizada é uma série de pulsos triangulares espaçados por um período fundamental T_0 . O espectro de potência do pulso glotal produzido pela vibração do sistema vocal é um espectro de harmônicas no qual as componentes diminuem a uma taxa de 12dB/oitava, como mostra a Fig. 2.9. O resultado é chamado de espectro de linha, ou espectro no qual a distribuição de energia está na forma de linhas. O espectro de energia da voz pode ser idealizado como um espectro de linha no qual cada linha individual é um múltiplo inteiro da frequência fundamental (f_0). A média da frequência fundamental dos homens é de 120Hz e para a mulher é de 220Hz. Porém, o ser humano pode produzir frequências maiores ou menores que a média. Praticamente falando, o espectro mostra que a energia da voz é predominantemente de componentes harmônicas de baixa frequência.

Figura 2.9: Espectro do som da laringe.

O subsistema articulatório é o responsável pelas modificações do som glotal, atuando como uma cavidade ressonante variável, atenuando certas frequências e amplificando outras do sinal glotal gerado pela fonte [KENT92].

Do ponto de vista de modelagem, o subsistema articulatório pode ser representado por um filtro, com determinada função de transferência que altera o espectro do som glotal. As relações entre as amplitudes das harmônicas variam muito das configurações da laringe e de cada falante, incluindo certas patologias. Variações na frequência fundamental apresentam alterações no espectro do pulso (ou glotal).

As vibrações produzidas em um ponto do trato vocal, atravessam o restante deste, que inclui laringe, faringe, cavidade bucal, cavidade nasal, língua, lábios, palato e dentes.

Os espaços formados entre esses componentes funcionam como cavidades ressonantes, modificando a excitação de ar provenientes da glote.

Muitos problemas vocais originam-se neste nível glótico, particularmente, ao longo da borda interna membranosa de cada prega vocal.

A postura glótica das pregas vocais, sua espessura, elasticidade e grau de aproximação, apresentam muita relação com a qualidade dos sons da vocalização. Muitos distúrbios funcionais da voz são resultantes da fisiologia inadequada dos músculos laríngenos, particularmente, na aproximação ou colisão incorreta das duas pregas vocais.

Às vezes, as pregas vocais são aproximadas de forma excessivamente comprimida. As membranas que as cobrem são unidas tão firmemente que a laringe funciona como uma válvula, impedindo o fluxo normal de ar que coloca as pregas em vibração (Fig. 2.10a). A voz resultante é tensa e soa, às vezes, quase como a voz cansada da “gagueira” laríngea ou disfonia espasmódica. Um problema oposto pode ser observado quando as pregas vocais são aproximadas de uma maneira tão frouxa que excesso de ar escapa entre elas, o que produz uma voz sussurrada ou muito soprosa (Fig. 2.10c).

Figura 2.10: Tipos de fonações e correspondentes ondas glotais referentes a: a) voz tensa b) voz normal e c) voz soprosa

Como é inviável observar as pregas vocais a partir do pulmão, ou seja, de baixo para cima, crescimentos anormais ocorrendo na face inferior das pregas não podem ser

visualizados pelos métodos atualmente disponíveis, caracterizando-se assim mais uma dificuldade no diagnóstico de patologias da laringe.

Com o intuito de suplantar estas dificuldades e prover auxílio aos especialistas no diagnóstico de doenças do trato vocal, particularmente as que afetam as pregas vocais, tem sido considerado o emprego do sinal de voz na avaliação das estruturas. Através da análise da fala, indiretamente 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 anestésicos.

Dada a natureza numérica da análise acústica, os diagnósticos são quantitativos e portanto objetivos. Permitem que o especialista de voz possa identificar o tratamento mais adequado ao paciente, uma vez que terá subsídios que lhe possibilitam identificar a patologia. Além disso auxilia em dirimir dúvidas eventuais que exames por laringoscopia indireta possam apresentar.

A própria evolução do tratamento pode ser avaliada, pois medidas acústica indicam precisamente a evolução do quadro clínico do paciente, permitindo que o especialista mantenha ou altere a medicação ou a forma de conduzir o tratamento.

Também nos estudos que empregam medidas acústicas, verifica-se que a grande maioria busca relacionar estes valores obtidos do sinal com características perceptuais da voz.

2.3 PATOLOGIAS DO TRATO VOCAL

A produção de fala exige o bom funcionamento e condições das estruturas do trato vocal. Qualquer interferência na dinâmica implica na geração de um sinal acústico diferente do que a pessoa emite normalmente, geralmente exigindo algum esforço adicional da mesma na sua produção. As patologias podem ocorrer em qualquer um dos três grupos de órgãos do aparelho fonador (respiratório, vibratório ou ressonador), sendo convencionalmente classificadas quanto à origem das mesmas: origem funcional, onde por um mal uso ou funcionamento de parte do sistema, origina-se uma fala anormal; e origem orgânica, em que a patologia é proveniente de anomalias físicas no sistema de produção de fala. Algumas patologias da laringe e sintoma acústico que elas produzem são apresentadas

na Tab. 2.1, e as pregas vocais com as patologias, cisto, nódulo, edema de Reinke e pólipos podem ser visualizadas nas figuras 2.11 a), b), c) e d) respectivamente.

Tabela 2.1: Patologias e sintomas acústicos

Origem Funcional	Sintoma Acústico
Edema de Reinke	Redução da frequência fundamental devido ao aumento de massa e presença de soprosidade.
Diplofonia	Presença de duas frequências não múltiplas, indicando haver duas fontes de sinal.
Disfonia Espasmódica	Em instantes não periódicos, há condições de aspereza ou soprosidade em sons sustentados.
Disfonia Ventricular	Pouca variação de altura e baixa frequência fundamental. As vezes percebe-se padrão diplofônico.
Espessamento	Variação da frequência fundamental.
Falsete	Elevada altura associada com escape de ar (ruído).
Laringite traumática	Variação de frequência fundamental e baixa potência (intensidade) do sinal.
Nódulos	Os nódulos interferem na vibração da mucosa, alterando a frequência fundamental e impedem a adução completa das pregas gerando soprosidade.
Pólipos	Apresenta soprosidade e dependendo de tamanho e localização pode abafar a vibração da mucosa gerando rouquidão.
Quebras de altura	Mudança repentina da frequência fundamental (uma oitava acima ou abaixo)
Quebras de vocalização	Aparecimento de trecho não vocálico, com uma rouquidão intensa durante emissão de um sinal vocálico.
Úlceras de contato	Primeiro aspereza seguida de rouquidão.

² Figura 2.11: pregas vocais patológicas

2. www.voiceproblem.org

Acusticamente, as patologias são diferenciadas em relação à padrões perceptuais como rouquidão, aspereza e sopro, e à escala destas características. Entretanto, uma quantidade de variações destes termos são encontrados na literatura e diferentes especialistas podem fornecer diferentes diagnósticos baseado no uso da sua própria percepção. Em alguns casos, este problema de discriminação patológica afeta a identificação das patologias até mesmo em exames como a videolaringoscopia e nasofibrosopia devido à carência de uma padronização e compreensão completa sobre o que é um aparelho vocal normal e sua dinâmica.

Uma voz áspera caracteriza-se por um elevado estiramento das pregas vocais, causada pelos músculos que a aduzem, exigindo assim elevados níveis de fluxo de ar vindos do pulmão. A voz soa estridente, e é caracterizada acusticamente por variações da frequência fundamental e amplitude do sinal glotal. Em patologias leves, pode-se identificar que, de modo geral, as pregas vocais não apresentam lesão. [ISSHIKI et al /66], menciona que um baixo fluxo de ar durante a fonação ocorre durante o comportamento hiperfuncional da laringe, acarretando a presença de pouco ruído nas altas frequências.

A soprosidade por sua vez caracteriza-se por um funcionamento adequado da mucosa e musculatura da laringe; entretanto, seja por alguma alteração orgânica nas pregas vocais ou mal funcionamento dos músculos adutores, ocorre escape de ar na região glotal, que origina ruído com ampla faixa de frequências, predominando principalmente as altas frequências (acima do segundo formante). Isto advém do fato de que as mucosas de ambas as pregas não estão totalmente aduzindo a prega vocal nos instantes necessários; ou seja, o efeito de Bernoulli associado à elasticidade do músculo vocal não são capazes de manter o espaço glotal fechado durante os instantes de elevação da pressão subglotal.

Existem estudos relacionando a quantidade de ruído com fatores que provocam turbulência por fechamento incompleto da glote, pois a velocidade do ar pelas pregas vocais excede o valor crítico de Reynolds; o número crítico encontrado para um modelo físico, tamanho real, de laringe foi de 2000 [ISSHIKI/78] para a percepção de soprosidade.

O último elemento perceptual da voz é a rouquidão, que nada mais é que a combinação dos efeitos tanto de soprosidade quanto de aspereza. A prega não vibra mais tão periodicamente e ocorrem escapes de ar, produzindo assim ruído turbulento e variações acentuadas na frequência fundamental.

Avaliando-se acusticamente a voz em termos perceptuais, pode-se chegar ao conjunto patológico que originou tal modificação acústica. Cabe lembrar que pode ocorrer situações em que “ouve-se” um determinado padrão acústico e nenhuma patologia é verificada em exames de videolaringoscopia e videoestroboscopia. [ISSHIKI/69] alerta para o fato que devido a complexa estrutura de funcionamento do sistema produtor de fala, patologias com características soprosas podem ser confundidas com as perceptualmente definidas como roucas.

Algumas patologias de origem funcional apresentam este problema, uma vez que as estruturas não possuem qualquer problema orgânico, o fluxo respiratório é normal e a movimentação das estruturas é adequada. Entretanto, o mal funcionamento de algum músculo ou ligamento pode incorrer na incompleta adução das pregas vocais, caracterizando assim uma soprosidade, pela existência de ruído turbulento.

O próprio mal uso das estruturas de fonação podem acarretar no aparecimento de alterações orgânicas pelo uso excessivo de determinadas estruturas, principalmente as pregas vocais e musculatura anexa. Nódulos e úlceras de contato podem-se desenvolver

pela produção constante de fala em alturas elevadas, acarretando, acusticamente sopro e/ou rouquidão.

O adequado estado das pregas vocais, sua densidade muscular e o estado da mucosa que a reveste permitem que ambas sejam aduzidas otimamente, possibilitando a criação de uma condição ideal de pressão e velocidade do fluxo de ar, gerando assim o sinal quase-periódico que excitará a estrutura supraglotal.

Um mal funcionamento de qualquer músculo intrínseco (principalmente) acarretará uma vocalização soprosa. Além disso, a voz naturalmente apresenta uma aspereza, pois as exigências de fluxo de ar aumentam (para permitir a vibração da mucosa) com a maior distensão das pregas vocais. Qualquer interferência nas bordas de aproximação das pregas vocais resultará em disфонia e ruído turbulento, com conseqüente sopro e/ou rouquidão na avaliação perceptual.

A estrutura de ressonância (órgãos supra-glota) é responsável pela caracterização fonética, transformando o sinal de excitação em fonemas, mediante movimento articulatorio de músculos e cartilagens, estabelecendo cavidades responsáveis pela ênfase de determinadas frequências e atenuação de outras.

Os componentes espectrais correspondentes às frequências naturais de ressonância são denominadas *formantes*. Teoricamente pode-se determinar infinitas frequências formantes, porém na prática, utiliza-se de três a cinco formantes. Devido às características particulares (anatômicas e fisiológicas) das pessoas, as frequências formantes apresentam certa variabilidade, tanto na frequência central como na largura de banda.

A cavidade supraglotal quando não funcionando adequadamente pode produzir efeitos sobre a fala, dada sua capacidade de produzir ênfase ou atenuação sobre as componentes espectrais do sinal glotal. O mal acoplamento das cavidades em regiões como das falsas pregas vocais, pregas ariepiglóticas, constrições da faringe e língua podem bloquear a passagem de ar, produzindo disфонia. Além disto, a ênfase inadequada pode produzir fonemas não desejados, efeitos coarticulatórios impróprios com as ansiedades da pessoa, modificando-se assim a envoltória espectral e as frequências formantes.

3 TÉCNICAS ACÚSTICAS PARA ANÁLISE DE VOZ

Muitas pesquisas são feitas em processamento digital de sinais na tentativa de se avaliar o sinal de fala para diagnosticar doenças da laringe [ROSA/00]. Medidas acústicas são propostas de forma a avaliar indiretamente o trato vocal por meio do sinal de voz coletado através de microfone convencional.

As técnicas acústicas para análise de voz utilizam a fonte de sinal e os métodos de análise numéricos. Uma fonte possível é o sinal propriamente dito, isto é, a pressão de som irradiada pela boca ou o sinal de contato na garganta. A outra fonte é o sinal indireto, isto é, o sinal glotal ou sinal residual obtidos por técnicas de filtragem inversa. Para cada uma destas fontes, os sinais podem ser analisados no domínio do tempo ou no domínio da frequência dependendo das vantagens e limitações que apresentam [PEREIRA/99].

3.1 SINAIS DIRETOS

A pressão de som irradiada pela boca é o sinal mais imediato para 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 supraglotal, cujas características ressonantes presumivelmente não serão afetadas pela patologia. As possíveis mudanças no pulso glotal devido a patologias, alterarão a forma de onda de pressão irradiada, podendo reforçar/atenuar componentes ou até mesmo eliminá-las. Alguns pesquisadores sugerem o uso de microfone de contato na garganta como forma de minimizar efeitos supra-glotaais, porém informação através destes sinais de contato é limitada pela ação de filtragem do tecido.

3.2 SINAIS INDIRETOS

O som vocálico é modelado como uma fonte (glote) excitando um tubo acústico. A técnica de filtragem inversa no sinal sonoro retira os efeitos do tubo acústico e, o sinal resultante aproxima-se do sinal periódico da fonte. Se a estrutura supraglotal não é afetada pela patologia, e a causa da mudança da voz está na glote, então esta aproximação do sinal

periódico da fonte contém informação suficiente para se analisar os efeitos acústicos da patologia.

Há dois métodos de filtragem inversa: a filtragem inversa glotal e a filtragem inversa residual:

a) FILTRAGEM INVERSA GLOTTAL

Neste procedimento os componentes espectrais relativas ao trato vocal e à radiação dos lábios são extraídas do sinal para se estimar a forma de onda da velocidade volumétrica como função do tempo.

b) FILTRAGEM INVERSA RESIDUAL

Esta técnica é baseada no modelo linear de produção da fala. O filtro residual inverso é o inverso das contribuições espectrais dos estimadores da radiação labial, do trato vocal e da fonte glotal. O resultado da filtragem do sinal de voz é o sinal residual. [KOIKE *et al*/75] usando o sinal residual para análise qualitativa de vozes normais e patológicas concluíram que, para casos patológicos intermediário e avançado, não há nenhuma informação adicional sobre os resultados que utilizam o sinal. Entretanto, para alguns casos em estágio inicial, o sinal residual mostrou desempenho superior ao sinal direto.

3.3 FONTE

A produção de fala, passa por duas etapas: a fonação e a articulação (Fig. 3.1). O som originado quando da fonação passa pelo trato vocal que atua como um filtro, mudando as características do som (volume e velocidade do ar).

Figura 3.1 Modelo de Produção de Fala

A fonte sonora é uma sucessão quase periódica de impulsos de ar emitidos através da glote. Durante a abertura e o fechamento das pregas vocais a função de transferência do filtro é considerada linear e invariante no tempo.

As características da fonte e do filtro estão associados a fenômenos de fonação e articulação regulados por interações mecânicas e acústicas. Um estreitamento articulatorio perturba a característica aerodinâmica do fluxo de ar e muda muito a fonte. O movimento de abertura das pregas vocais altera a fonte e a impedância glotal (parte integrante da função de transferência do filtro).

Na fase de abertura da glote as partes subglotal e supraglotal, do trato vocal, estão acopladas acusticamente através de uma impedância glotal não linear e variável no tempo. Durante o fechamento da glote os sistemas subglotal e supraglotal produzem oscilações livres e independentes um do outro. As frequências de ressonância e largura de banda são diferentes nas duas fases.

3.4 MODELO FONTE-FILTRO

O princípio básico para determinação do sinal glotal é a aplicação de métodos de filtragem inversa. Em processamento digital de sinais basicamente tem-se o método de correlação (eficientemente resolvido pelo algoritmo de Levinson) e o método da covariância (através do método de Cholesky).

A suposição para a aplicação destes métodos está na separação linear da função de transferência do trato supraglotal ($V(z)$) e efeitos de radiação dos lábios ($R(z)$) da função de transferência da glote ($G(z)$). Também é considerado que o sistema é invariante no tempo pois, a dinâmica do trato vocal é lenta se comparada à frequência fundamental do sinal de voz e o sinal adquirido é um som vocálico sustentado. Para excitar este modelo, um trem de pulsos (onde o intervalo entre pulsos caracteriza a frequência fundamental da fala) representa a função do pulmão como gerador de energia para excitar a glote. A Fig. 3.2 representa esse sistema:

Figura 3.2: modelo representativo do trato vocal

Os valores de ganho (A_v e A_n) consideram os efeitos da pressão exercida pelos pulmões durante a fonação. O modelo do pulso glotal conforma o pulso, de modo a torná-lo similar àquele produzido pela vibração das pregas vocais relativo ao som vocálico. Considerando-se que para sinais vocálicos tem-se uma onda quase-periódica, utiliza-se um gerador de trem de pulso, separados por um período de tempo denominado período de pitch (frequência fundamental), o sistema que produz o sinal advindo das pregas vocais é dado por:

$$g(n) = \begin{cases} \frac{1}{2} \left[1 - \cos \left(\frac{\pi n}{N_1} \right) \right], & 0 \leq n \leq N_1 \\ \cos \left[\frac{\pi(n - N_1)}{2(N_1 - N_2)} \right], & N_1 \leq n \leq N_2 \\ 0, & \text{caso contrário} \end{cases} \quad (3.1)$$

Este sistema não é único e definitivo. Foi construído com base em propriedades espectrais e similaridade 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 deste modelo e no efeito de filtragem passa-baixa característico do sistema. A Fig. 3.3 representa um exemplo de sinal glotal obtido a partir da equação 3.1, definida empiricamente:

Figura 3.3 : Trecho Periódico do Sinal de Voz

O modelo para o sinal não-vocálico é obtido através de um gerador de ruído 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.

No domínio do tempo, o sinal de voz $u(t)$ é dado pela soma de convolução:

$$u(t) = g(t) * v(t) \quad (3.2)$$

Como a radiação pode ser aproximada por um diferenciador [PLUMBE/99] uma vez que seu efeito é amplificar os componentes de alta frequência do sinal relativos da cavidade bucal, o sinal da equação (3.2) após a radiação labial pode ser representado por:

$$u'(t) = \frac{d[g(t) * v(t)]}{dt} \quad (3.3)$$

onde o termo entre colchetes representa o sinal na boca e $u'(t)$ é o sinal já irradiado nos lábios. O diagrama em blocos deste modelo está mostrado na Fig. 3.4.

Considerando os efeitos de radiação na função de transferência $v(t)$, denominando-a $h(t)$ [PLUMBE/99], pode-se simplificar a equação (3.3) por:

$$u'(t) = \left\{ \frac{d[g(t)]}{dt} \right\} * h(t) \quad (3.4)$$

Fica evidenciado, portanto, que a derivada do sinal de voz é uma convolução entre a derivada do sinal glotal $g'(t) = \frac{d[g(t)]}{dt}$ e a resposta ao impulso do trato supraglotal e de radiação dos lábios (Fig. 3.5). Desta forma demonstra-se a importância do sinal derivativo da onda glotal e a equação (3.4) permite que, uma vez determinado $g'(t)$, o sinal de voz $u'(t)$ pode ser estimado sem a necessidade de integrar a onda glotal.

Figura 3.4: esquema simplificado do sistema de fonação humano

Figura 3.5: modelo fonte-filtro

Para se estimar $g'(t)$ é necessário buscar algumas referências temporais no sinal de voz. O mais importante e confiável é o intervalo de fechamento das pregas vocais. Durante este tempo, o trato supraglotal não é excitado pois, não existe escoamento de ar e portanto, os efeitos de acoplamento (geralmente não-lineares) entre o trato supraglotal e a laringe podem ser considerados nulos. Assim, uma das técnicas básicas envolve a avaliação instantânea das frequências formantes (constantes, pois o sistema é suposto ser estacionário) cujos módulos apresentam valores mínimos de amplitude durante a fase de fechamento. O intervalo de fechamento é então definido pelo intervalo de tempo na qual as amplitudes formantes permaneçam mínimas, garantidas as condições de estacionariedade do sistema. Assim, a porção do sinal de voz neste intervalo de tempo é caracterizado no modelo matemático (Equação 3.2) apenas pela convolução entre o trem de pulsos (excitação) e a resposta impulso. Assim "invertendo-se" este sistema, pode-se avaliar uma estimativa do sinal derivativo glotal $g'(t)$.

3.5 MODELOS GLOTAIS PARA PROCESSAMENTO DE VOZ

Há vários modelos glotais que diferem em forma e complexidade tornando-os capazes de modelar diferentes estilos de vozes e consequentemente sua aplicabilidade em modelar de forma mais adequada as propriedades fisiológicas da glote. Os modelos glotais podem ser divididos em três categorias:

- a) *modelo glotal paramétrico não interativo*: este modelo assume a não interação entre a fonte e o trato glotal possibilitando dessa forma serem gerados através de funções paramétricas capazes de descrever o pulso glotal ou sua derivada.
- b) *modelo glotal paramétrico interativo*: neste modelo há a interação entre o glote e o trato vocal, são modelados por funções paramétricas mais sofisticadas.
- c) *modelo glotal fisiológico*: este tipo de modelo é o mais complexo pois envolve teorias relacionadas a interação entre a fonte glotal e as cavidades sub e supra-glotaais que resultam na vibração das pregas vocais.

A seguir uma descrição dos modelos paramétricos não interativos, baseados na teoria fonte-filtro da produção de fala, os detalhes dos demais modelos, poderão ser encontrados em [CUMMINGS *et al*/95].

3.5.1 MODELOS GLOTAIS PARAMÉTRICOS NÃO INTERATIVOS

Os modelos glotais paramétricos não interativos assumem que a fonte glotal e o trato vocal são linearmente separáveis e que não há interação entre eles. A base deste modelo é que a forma de onda glotal age como a excitação para um filtro linear que representa as ressonâncias do trato vocal e radiação nos lábios ou nariz. Para propósito de análise, síntese e codificação de voz, este modelo tem-se mostrado adequado devido a sua simplicidade, onde seus parâmetros são ajustados a partir da forma de onda glotal da voz encontrada através da filtragem inversa, ou seja, o pulso glotal é derivado a partir da filtragem do sinal de fala com o inverso do trato vocal.

Alguns autores, como [Rosenberg/71] e [Holmes/73], demonstraram a partir de vários resultados experimentais, com diferentes formas glotais, a importância de encontrar uma forma do pulso glotal adequada para a qualidade da fala sintética. Os pulsos glotais simulados pelos modelos polinomial (equação 3.5) e trigonométrico (equação 3.6) foram o que mais demonstraram semelhança com um pulso glotal natural proporcionando conseqüentemente uma fala sintética com melhor qualidade.

$$u_g(t) = \begin{cases} \alpha \left(3 \left(\frac{t}{T_p} \right)^2 - 2 \left(\frac{t}{T_p} \right)^3 \right) & \text{para } 0 \leq t \leq T_p \\ \alpha \left(1 - \left(\frac{t - T_p}{T_N} \right)^2 \right) & \text{para } T_p \leq t \leq T_p + T_N \end{cases} \quad (3.5)$$

$$u_g(t) = \begin{cases} \frac{\alpha}{2} \left(1 - \cos\left(\frac{t\pi}{T_p}\right) \right) & \text{para } 0 \leq t \leq T_p \\ \alpha \cos\left(\frac{\pi}{2} \frac{t - T_p}{T_N}\right) & \text{para } T_p \leq t \leq T_p + T_N \end{cases} \quad (3.6)$$

Estes modelos são definidos por três parâmetros, sendo um deles referente a amplitude (α) e os demais ao tempo (T_p : intervalo entre o tempo onde há o começo da abertura da glote e o tempo onde o pulso glotal atinge o máximo positivo e T_N : intervalo entre o tempo de máximo valor positivo do pulso glotal e o tempo onde há o começo do fechamento da glote). A tabela 3.1, mostra os valores dos parâmetros que foram usados para descrever os modelos polinomial e paramétrico respectivamente [CUMMINGS/1995].

Tabela 3.1: Parâmetros referentes aos modelos polinomial e trigonométrico

	modelos	
parâmetros	polinomial	trigonométrico
α	55503.763	56383.798
T_p	18.072	18.784
T_N	13.928	13.216

Esses modelos podem ser visualizados através da Fig. 3.6 e 3.7.

Figura 3.6: modelo polinomial de Rosenberg

Figura 3.7: modelo trigonométrico de Rosenberg

Durante o fechamento glotal $u_g(t)$ é considerado zero, isto devido a hipótese de que não há passagem de ar nesta fase. Os resultados dos modelos de Rosenberg foram confirmados e estendidos por [Holmes/1973], que comparou duas vozes sintéticas sendo que uma usou a excitação glotal estimada pelo filtro inverso e a outra pelo pulso glotal trigonométrico. Apesar da semelhança dos resultados, apresentada pelas duas vozes, especialistas confirmaram maior naturalidade no pulso originado pelo filtro inverso. Estes modelos além de apresentarem uma base para o desenvolvimento de outros modelos paramétricos glotais, também mostraram sua utilidade para a síntese de formantes [HOLMES/1973], [HOLMES/1983] e codificação de voz. Holmes incorporou a radiação labial para a implementação das formantes, usando a segunda derivada do pulso glotal extraído pelo filtro inverso. [MATSUI *et al*/1991], também usou os pulsos originados pelo filtro inverso de uma grande variedade de tipos de voz, como excitação para um sintetizador de formantes. Os pulsos gerados para a fala sintética foram testados visualmente baseando-se na fala sintética produzida.

Outro modelo utilizado para sintetizar formantes, foi desenvolvido por [Klatt and Klatt/1990], onde o pulso glotal é especificado em termos de sete parâmetros: amplitude, frequência fundamental, quociente de abertura, *tilt* espectral, *flutter* (flutuações quase-aleatórias na frequência fundamental), grau de diplofonia (irregularidades formando duplo pulso) e amplitude do ruído de aspiração. Estes três últimos parâmetros foram adicionados no modelo polinomial de Rosenberg, na intenção de encontrar uma fala sintética mais natural, apesar de aumentar a complexidade do modelo. Klatt também usou o modelo derivativo de [FANT *et al*/1985] (LF) como uma fonte opcional para o seu sintetizador. O modelo LF, é o modelo paramétrico mais simples, sendo sua estrutura modelada como um filtro de primeira derivada do pulso glotal, isto devido aos efeitos de radiação labial que são incorporados no modelo glotal. Apesar de sua simplicidade, este modelo tem sido utilizados por muito pesquisadores na intenção de sintetizar, analisar e codificar vozes, inclusive com características patológicas. O modelo LF utilizado pelo sintetizador de Klatt, foi especificado por cinco parâmetros, e pode ser descrito pela equação (3.7) e visualizado

na Fig. (3.8) quando os valores estabelecidos na tabela 3.2 [CUMMINGS/1995] são adotados.

$$\begin{cases} u'_g(t) = E_0 e^{\alpha t} \sin(\omega_g(t)) & \text{para } 0 \leq t \leq T_e \\ u'_g(t) = -\left(\frac{E_e}{\varepsilon T_a}\right) \left(e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_e-T_e)}\right) & \text{para } T_e \leq t \leq T_e \end{cases} \quad (3.7)$$

onde os parâmetros que são derivados a partir do pulso encontrado pelo filtro inverso são:

E_e : amplitude máxima negativa do pulso derivativo;
 T_p : tempo onde o pulso derivativo encontra o máximo positivo;
 T_e : tempo de máxima descontinuidade glotal;
 T_a : tempo referente à fase de retorno e;
 T_o : período de *pitch*.

Tabela3.2: valores de alguns parâmetro modelo LF

parâmetros	Modelo LF
E_0	2622.799
α	0.032
ε	2.000

Figura 3.8 : modelo derivativo do pulso glotal LF

Outro modelo derivativo glotal, com base polinomial, foi desenvolvido por [Fujisaki *et al*/1986], [Fujisaki *et al*/1987]. Este modelo conta com 6 parâmetros: duração da fase de abertura (W), inclinação do pulso (S), intervalo de tempo a partir do fechamento glotal para o máximo negativo do pulso derivativo (D), o declive na abertura glotal (A), o declive antes do fechamento glotal (B) e o declive após o fechamento glotal (C). Este modelo é mostrado na Fig. (3.9). W e S (equações 3.8 e 3.9) são definidos em termos R,

que é duração da fase de abertura antes do pico e F , que é duração da fase de abertura depois do pico:

$$W = R + F \quad (3.8)$$

e

$$S = \frac{R + F}{R - F} \quad (3.9)$$

O período de *pitch* (T) é especificado separadamente. As equações que especificam os segmentos polinomiais são dados pela equação 3.10:

$$\begin{aligned} u'_g(t) &= A - \frac{2A + R\alpha}{R}t + \frac{A + R\alpha}{R^2}t^2 && \text{para } 0 < t \leq R \\ u'_g(t) &= \alpha(t - R) + \frac{3B - 2F\alpha}{F^2}(t - R)^2 - \frac{2B - F\alpha}{F^3}(t - R)^3 && \text{para } R < t \leq W \\ u'_g(t) &= C - \frac{2(C - \beta)}{D}(t - W) + \frac{C - \beta}{D^2}(t - W)^2 && \text{para } W < t \leq W + D \\ u'_g(t) &= \beta && \text{para } W + D < t \leq T \end{aligned} \quad (3.10)$$

onde

$$\alpha = \frac{4AR - 6FB}{F^2 - 2R^2} \quad \text{e} \quad \beta = \frac{CD}{D - 3(T - W)}$$

Figura 3.9 : modelo derivativo glotal Fujisaki e Ljungqvist

A tabela 3.3, mostra o valores especificados para o modelo acima, adotados por [CUMMINGS/1995].

Tabela 3.3: valores para os modelos Fujisaki e Ljungqvist

parâmetros	Modelo Fujisaki e Ljungqvist
A	2589.715
B	-7233.658
C	-2487.484
D	8.761
F	8.951
R	22.049
W	31.000
α	-462.648
β	20.317

No caso de codificação da fala, um dos modelos mais conhecidos foi projetado por [Hedelin/1984]. O autor modelou a excitação glotal com funções trigonométricas similares ao modelo trigonométrico de Rosenberg. A equação 3.11 mostra a definição do modelo:

$$\left\{ \begin{array}{ll} u_g(t) = y_1 + (y_2 - y_1) \left(\frac{t - t_1}{t_2 - t_1} \right) & \text{para } t_1 < t \leq t_2 \\ u_g(t) = y_2 + (y_3 - y_2) \sin^2 \left(\frac{\pi}{2} \frac{t - t_2}{t_3 - t_2} \right) & \text{para } t_2 < t \leq t_3 \\ u_g(t) = y_4 + (y_3 - y_4) \cos^2 \left(\frac{\pi}{2} \frac{t - t_3}{t_4 - t_3} \right) & \text{para } t_3 < t \leq t_4 \end{array} \right. \quad (3.11)$$

O fechamento glotal, que é definido como o tempo entre t_1 e t_2 , é um sinal não zero. Este modelo é definido por 6 parâmetros sendo 3 referentes ao tempo e 3 referentes a amplitude. Assumindo a continuidade dos pulsos, tem-se que $y_4 = y_1$. A Fig. 3.10 mostra o ajuste do modelo em relação ao pulso glotal encontrado pelo filtro inverso LPC, onde os parâmetros segundo [CUMMINGS/1995] são especificados na tabela 3.4:

Tabela 3.4: parâmetros do modelo definido por Hedelin

Parâmetros	Modelo Hedelin
y_1	-2519.922
y_2	-384.262
y_3	53848.860
y_4	-586.117
t_1	0
t_2	12.963
t_3	39.049
t_4	50.000

Figura 3.10: modelo glotal de Hedelin

Para concluir este capítulo, a tabela 3.5, mostra um resumo dos modelos paramétricos não-iterativos definidos pelas equações 3.5 à 3.11 e visualizados pelas Fig. s 3.5 a 3.9. Eles foram ajustados pelo pulso determinado pelo filtro inverso de uma voz real, com a vogal sustentada /i/, segundo [CUMMINGS/1992], (Fig. 3.11) ou por sua derivativa (Fig. 3.12).

Figura 3.11: pulso glotal encontrado pelo filtro inverso

Figura 3.12: pulso glotal derivativo real

Tabela 3.5: resumo dos modelos paramétricos não-interativos

pesquisador	Base funcional	Número de parâmetros	Tipo de modelo
Rosenberg	Polinomial	3	Velocidade volumétrica
Holmes	Manipulado natural	-	Segunda derivativa
Matsui	Manipulado natural	-	Velocidade volumétrica
Klatt	Polinomial	7	Velocidade volumétrica
Liljencrants e Fant	Trigonométrico	5	Primeira derivativa
Fujisaki e Ljungqvist	Polinomial	6	Primeira derivativa
Hedelin	Trigonométrico	6	Velocidade volumétrica

3.5.1.1 CONSIDERAÇÕES GERAIS

Dentre os modelos citados, todos se mostraram úteis em aplicações como análise, síntese e codificação de voz. Sendo que os mais simples de serem gerados, foram os modelos polinomial e trigonométrico, desenvolvidos por Rosenberg, por outro lado o mais complexo é o modelo definido por Fujisaki e Ljungqvist. E por fim, o modelo derivativo, LF tem a característica adicional de que a radiação labial é incorporada no modelo glotal. Este modelo será descrito em detalhes no próximo tópico por ser o modelo de ajuste usado neste trabalho.

4 MODELO LF

Após a estimativa dos pulsos glotais pela filtragem inversa, o próximo passo consiste em quantificar e interpretar as características relevantes deste sinal. Como a medida exata de alguns parâmetros a partir do sinal filtrado envolve um nível de complexidade que às vezes impossibilita a robustez dos resultados, modelos paramétricos são usados, com a finalidade de estimar estes parâmetros, com a vantagem que requer menos tempo na estimação. Porém, a confiabilidade do modelo, depende da intervenção de um profissional da área para um controle exato dos detalhes para ajuste.

Neste trabalho, é feita a parametrização da primeira derivativa do pulso glotal baseada no ajuste do modelo LF, onde suas propriedades tem sido muito discutida na literatura desde de sua publicação original [FANT et al/1985]. Este modelo, além de assegurar um ajuste às formas de ondas encontradas através de um número mínimo de parâmetros é flexível na adaptação a fonações bruscas.

Nesta sessão serão apresentadas as funções que caracterizam o modelo LF.

A partir do sinal derivativo glotal, estima-se um conjunto de sete parâmetros (tabela 4.1 e figura 4.1) que defina $g'(t)$ em termos de intervalo de tempo (abertura- t_o , fechamento- t_c e “fase de retorno”- t_a) e amplitudes. O modelo mais tradicional é o modelo Liljencrants-Fant (LF) [FANT/1986]. A figura 4.1 mostra o sistema fonte-filtro com a derivativa do sinal glotal modelada pelo modelo LF:

Figura 4.1: modelo fonte-filtro com sinal derivativo modelado por LF

4.1 MODELANDO O PULSO DERIVATIVO GLOTAI

A representação do sinal glotal é baseada na quantificação da primeira derivada do pulso glotal, considerando os parâmetros propostos pelo modelo LF. Tais parâmetros apresentam vantagens sobre medidas acústicas por estarem relacionados com características fisiológicas reais das pregas vocais.

O modelo matemático que caracteriza um sinal de voz num intervalo de tempo no qual as frequências formantes são consideradas estacionárias, é dado pela convolução entre o trem de pulsos e a função de transferência do trato vocal onde com a inversão do sistema, obtêm-se a estimativa do sinal glotal $g(t)$ e conseqüentemente a derivada deste, $g'(t)$.

Tabela 4.1: Parâmetros do modelo LF para o sinal derivativo da glote

T_0	Tempo de abertura da glote
T_e	O tempo de máximo valor negativo do pulso derivativo glotal
E_e	O valor do sinal derivativo no tempo T_e
α	Razão entre E_e e a amplitude de pico da porção positiva do sinal derivativo glotal
ω_g	Frequência contida no sinal derivativo no início do pulso glotal; também determina o tempo transcorrido entre zero e T_e
β	Constante de tempo de amortecimento do sinal
T_c	Tempo de fechamento da glote

Os parâmetros, $E_0 = \frac{E_e}{e^{\alpha(T_e - T_0) + \sin \omega_g (T_e - T_0)}}$, ω_g e α descrevem a forma da onda glotal

durante a fase de abertura. Os parâmetros E_e e β descrevem a forma da fase de retorno.

A fase de retorno determina a quantidade de energia de altas frequências nos sinais de excitação e fala. Quanto mais rapidamente as pregas vocais se fecham, menor a fase de retorno, resultando em maior energia de altas frequências e menor inclinação espectral.

As equações que definem o modelo LF são [FANT/1986] :

$$\begin{aligned}
 g'(t) &= 0 & 0 \leq t \leq T_0 \\
 g'(t) &= E_0 e^{\alpha(t-T_0)} \sin[\omega_g(t-T_0)] & T_0 \leq t \leq T_c \\
 g'(t) &= -E_1 [e^{-\beta(t-T_c)} - e^{-\beta(T_c-T_c)}] & T_c \leq t \leq T_c
 \end{aligned} \tag{4.1}$$

onde $E_1 = \frac{E_c}{1 - e^{-(T_c-T_c)}}$ e o tempo na origem é $t = 0$.

Figura 4.2: Pulso Glotal e Sinal Derivativo Glotal (Modelo LF)

O ponto máximo de abertura em U_0 , ocorre em T_p , e máxima excitação com amplitude E_c ocorre em T_c , que corresponde ao instante onde há a colisão entre as pregas vocais.

Integrando as equações 4.1, o modelo pode ser expresso em termos do pulso glotal, o que é visualizado na figura 4.2 (a).

4.2 MÉTODOS PARA ESTIMAR O PULSO GLOTTAL DERIVATIVO

4.2.1 MÉTODO DA AUTOCORRELAÇÃO

Seja $s[n]$ a saída do filtro do trato vocal com a resposta impulsiva $h[n]$, excitado pela derivativa da corrente glotal $u[n]=g'[n]$, incluindo a radiação labial.

Para um filtro tudo pólo $h[n]$, a função sistema é dada por:

$$H(z) = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (4.2)$$

No domínio do tempo, tem-se:

$$s[n] = \sum_{i=1}^p a_i s[n-i] + v_g[n] \quad (4.3)$$

Para estimar os coeficientes a_i a partir do sinal de fala $s[n]$, usa-se o método da predição linear. Este método tornou-se dominante para estimar parâmetros do sinal de fala considerando-se o sinal estacionário, devido à variação lenta do trato vocal. A idéia básica da predição linear é a de que o valor de uma amostra pode ser aproximado (predito) por combinação linear dos valores das amostras anteriores, tirando correlação entre estas, resultando em um modelo autoregressivo. Os coeficientes LPC (*Linear Predictive Coding*) são estimados por minimização do erro quadrático entre a amostra atual e sua predição, resultando em um filtro que simula o trato vocal. As ressonâncias do trato vocal correspondem a picos (formantes) na função sistema do mesmo. Esta função de sistema corresponde à envoltória do espectro da voz, portanto o método de detecção de formantes consiste em estimar esta envoltória, e detectar os picos correspondentes aos mesmos. Esta estimação LPC apresenta as seguintes características:

- Capacidade de obter uma estimativa do espectro com um número menor de pontos que a transformada discreta de Fourier. Isto é de grande importância, uma vez que a fala só pode ser considerada estacionária em trechos muito curtos (dezenas de milissegundos).

- Se usada com uma ordem relativamente baixa, fornece um espectro já alisado, o que facilita a identificação dos picos dos formantes, que também podem ser obtidos através da identificação dos zeros do polinômio correspondente ao inverso da função de transferência do filtro obtido.

- A envoltória espectral da voz, moldada pelas cavidades ressonantes do trato vocal, em geral, apresenta picos semelhantes a um espectro autoregressivo, o que torna mais simples a modelagem através da predição linear.

- Apesar de seu algoritmo ser em geral mais lento que o da FFT (*Fast Fourier Transform*), pode ser agilizado através do algoritmo de Levinson-Durbin.

- O sinal de erro pode ser usado para identificação dos pulsos glotais, uma vez que o erro é grande no início de cada pulso. Portanto, a excitação é representada pelo erro de predição, ou seja, pela diferença entre o sinal de voz e o sinal estimado por predição linear. O sinal de erro é elevado ao quadrado, para que corresponda à energia do erro; então, é passado por um filtro "passa-baixas" escolhido de modo a suavizar o resultado, tornando-o melhor para efeito de visualização, e mais parecido com os pulsos de pressão provenientes da glote no caso de sinais vozeados.

- O método da covariância geralmente apresenta um erro médio quadrático ligeiramente inferior ao da autocorrelação, mas, para o tamanho do trecho utilizado, esta diferença é insignificante [RABINER *et al*/ 1978]

4.2.2 ESCOLHA DA ORDEM DE PREDIÇÃO

Segundo [RABINER *et al*/ 1978], são necessários, em média, 2 pólos por kHz de amostragem para representar o trato vocal, mais 3 a 4 pólos para o sinal de excitação e o efeito de irradiação nos lábios. Como exemplo, para uma frequência de amostragem de 10

kHz, isso nos dá um total de 13 a 14 pólos. Ainda assim, para algumas vogais, onde o número de pólos é capaz de discriminar os formantes, aumenta-se a ordem do sistema.

4.2.3 ESCOLHA DA JANELA

A função janela é aplicada a cada trecho do sinal, antes da análise por predição linear. A janela escolhida é a Hamming, para minimizar o efeito das bordas do bloco, uma vez que o método de cálculo de LPC é o da autocorrelação. O tamanho da janela mínimo normalmente utilizado na análise de voz por LPC assíncrona em relação aos pulsos glotais é 10 a 20 ms [RABINER *et al* 1978]. Como as características do sinal variam pouco durante a pronúncia de vogais, a janela utilizada foi de 30 ms.

4.2.4 PRÉ-ÊNFASE

A pré-ênfase visa minimizar os efeitos da irradiação nos lábios e da forma de onda dos pulsos glotais na envoltória espectral e consiste de um filtro derivador, realçando as altas frequências. Para isto utiliza-se um filtro média móvel (MA) com função de transferência $1 - 0.98.z^{-1}$.

4.2.5 DETERMINAÇÃO DA REGIÃO DA FASE FECHADA ATRAVÉS DA MODULAÇÃO FORMANTE

Durante a fase fechada, a ausência de interação filtro/fonte, resultará em nenhuma ou em pequena modulação formante, enquanto na fase aberta a mudança da área glotal resultará em formantes não-estacionárias, portanto, a fase fechada glotal é identificada como a região em que a modulação formante é estacionária. Conforme descrito no tópico anterior, o efeito dos formantes pode ser atenuado calculando a função de covariância sobre o resíduo de predição em vez do próprio sinal, que mantém a mesma periodicidade do sinal mas em que é retirada a contribuição dos formantes dada pelo filtro LPC. Neste caso é usado um modelo tudo pólo de ordem p , para o método de covariância. A estimação da fase

fechada glotal é obtida, usando uma estimacão da modulacão formante, calculada entre sucessivos pulsos glotais. Para medir as frequências formantes dentro do ciclo glotal, uma análise de predicão linear, baseada na covariância móvel com mudançã de uma-amostra, é usada. O comprimento da janela de covariância móvel, denotado por N_w , é escolhido por cada “frame” como $N_w = N/4$, para $p+3 \leq N_w \leq 2p$, onde N é o comprimento do período de “pitch” e p é ordem do preditor linear [PLUMPLE *et al*/1999].

Devido à possibilidade de que as pregas vocais nunca se fecham completamente, o “grau” de modulacão formante durante a fase fechada varia de pessoa para pessoa. Uma aproximacão estatística para identificar a fase fechada glotal, consiste em achar uma pequena região de amostras de formantes seqüenciais, em que a modulacão formante é mínima, para isto calcula-se a soma absoluta da diferençã entre sucessivas formantes estimadas sobre um intervalo de amostras, ou seja:

$$D(n_0) = \sum_{i=n_0}^{n_0+4} |F(i) - F(i-1)| \quad \text{para} \quad 1 \leq n_0 < N - N_w - 5 \quad (4.4)$$

o argumento n_0 é a primeira amostra da região de cinco-amostras e é variado para minimizar $D(n)$; $F(n)$ representa os valores das formantes para cada amostra no ciclo glotal e N é o número de amostras do ciclo glotal. O tamanho de cinco amostras é selecionado para assegurar estatísticas significativas para determinar a região formante estacionária inicial.

Após identificada uma região estacionária inicial, a média e o desvio padrão da primeira formante dentro desta pequena região, são calculados. Em seguida haverá uma extensão da região para o lado direito se a próxima amostra for menor que dois desvios padrões a partir da média. Isto é incluído na região estacionária e a média e o desvio padrão são recalculados antes de continuar o teste para a próxima amostra. A média e o desvio padrão finais da extensão da região para o lado direito são guardadas e serão usadas no cálculo para a extensão para o lado esquerdo que acontecerá até que a amostra seja maior do que dois destes desvios padrões a partir da média.

A fase fechada considerada inclui cada amostra da fala que foi usada para calcular os valores formantes estáveis.

O resultado deste algoritmo é uma região $[N_1, N_2]$, em que a modulação formante é considerada estacionária e em que a estimação do trato vocal é obtida pelo filtro inverso, onde o resultado do filtro inverso é ajustado com o modelo teórico LF.

4.3 AVALIANDO O MODELO LF A PARTIR DO PULSO GLOTAI DERIVATIVO GLOTAI

Muitos parâmetros da fonte são definidos no domínio do tempo e relacionados com o padrão vibratório das pregas vocais, no entanto, aplicação do modelo LF e simplificações oriundas deste, têm sido estudadas por diversos autores correlacionando a parametrização em outros domínios, como por exemplo no domínio da amplitude do sinal (parâmetros AQ [ALKU/1996], e NAQ [ALKU/2002]). A motivação para extração de parâmetros em domínios que não o do tempo, se dá principalmente pela razão de que no processo do filtro inverso (para obtenção do pulso glotal) há um cancelamento incompleto de formantes, tendo como consequência, formas de ondas distorcidas pelas flutuações resultantes, dificultando a extração do pulso glotal, e consequentemente, a determinação precisa dos parâmetros glotais (quociente de abertura, de fechamento e índice de velocidade). Deste modo, com a intenção de salientar características importantes do pulso derivativo glotal, neste tópico, serão apresentados alguns parâmetros relevantes para a quantificação acústica e perceptual.

4.3.1 PARÂMETROS ACÚSTICOS

Houve-se a necessidade de discorrer sobre alguns parâmetros acústicos, para uma possível corroboração dos resultados preliminares. Para essa análise, usou o programa Análise de Voz, desenvolvido por [MONTAGNOLI/1998]

A seguir é dada uma descrição sucinta de cada um deles:

a) *Frequência fundamental – F_0* : Este parâmetro avalia a ocorrência do número de ciclos do pulso glotal que ocorrem por segundo. O limiar masculino, segundo avaliação estatística, foi estipulado ser 160 Hz, enquanto o feminino é de 250 Hz.

b) *Jitter*: Este parâmetro avalia a perturbação da frequência do sinal. É calculado pela razão entre as variações do período e o período médio constantes da amostra selecionada. Uma amostra com jitter igual a zero, significa que os períodos e, portanto a frequência, são iguais para qualquer ciclo da amostra do sinal. O valor limite de 8% foi estabelecido sobre uma população sem anomalias na laringe através de técnicas estatísticas.

c) *Shimmer*: Similarmente ao jitter, o shimmer avalia a perturbação da amplitude do sinal. Através dos mesmos conceitos matemáticos, calcula-se o parâmetro pela substituição do período pela amplitude do sinal. Usando-se as mesmas técnicas estatísticas, determinou-se o valor de limiar igual a 15%.

d) *Nivelamento Espectral - SF (Spectral Flatness)*: Este parâmetro é uma medida da variação espectral do sinal, isto é, está relacionado à variação das amplitudes das componentes espectrais do sinal. Os parâmetros avaliados dentro desta medida são: *Nivelamento Espectral do Sinal Residual (SFR)* e o *Nivelamento Espectral do Filtro Inverso (SFF)*, os quais quantificam a suavidade das componentes espectrais relativas à glote e ao trato vocal respectivamente.

e) *Amplitude de Pico – PA*: A quantificação desta medida corresponde à mediana da amplitude do segundo pico da autocorrelação do sinal residual. A variação deste parâmetro indica os efeitos da patologia sobre as estruturas da vocalização, pois a existência perceptual de aspereza ou sopro implica redução da magnitude deste parâmetro.

f) *Coefficiente de Excesso – Ex*: Este coeficiente é uma medida da forma da distribuição de amplitudes do sinal residual. Vozes normais apresentam distribuições mais

concentradas e de maior magnitude, enquanto vozes patológicas tem distribuições com menores amplitudes e são mais esparsas.

4.3.2 PARÂMETROS RELACIONADOS AO PULSO DERIVATIVO GLOTAL

- **Intensidade Espectral/Comprimento da excitação – EE:**

Esta medida é a amplitude do pico negativo do pulso derivativo glotal. Este pico ocorre no tempo de máxima descontinuidade da forma de onda (figura 4.3). Acusticamente, EE está correlacionado com a intensidade global do sinal. A mudança de EE resulta em uma mudança correspondente da amplitude em todas as componentes da fonte, com exceção das menores, que são relacionadas à primeira harmônica. A amplitude destas componentes é determinada pela forma do pulso e deste modo varia menos com a mudança em EE.

Figura 4.3: Valores de EE: a) pulso derivativo glotal com um pequeno valor EE;
b) pulso glotal derivativo com um grande valor de EE.

- **Simetria Glotal/Inclinação do pulso – R_k :**

Este parâmetro captura a assimetria glotal (figura 4.4). O pulso glotal, normalmente é inclinado para a direita pelo fato da fase de abertura ser maior do que a fase de fechamento. Esta medida é expressa entre os tempos de abertura e fechamento do pulso

glotal, ou seja,
$$\frac{\left((T_e - T_0) - (T_p - T_0) \right)}{(T_p - T_0)}.$$

Acusticamente, a inclinação do pulso afeta a amplitude das harmônicas das baixas frequências, quanto maior R_k , mais assimétrico a forma do pulso, diminuindo a excitação. Por outro lado, um aumento na inclinação do pulso, resulta em uma forte excitação provocando um acréscimo em EE e deste modo, um aumento nas componentes de alta frequência. Consequentemente, quanto mais simétrico o pulso há a tendência de um enfraquecimento de certas harmônicas [Ní Chasaide *et al*/1997], isto provavelmente é devido a um decréscimo da excitação.

Figura 4.4: Valores esquematizados de RK: a) pulso derivativo glotal com um pequeno valor de RK; b) pulso derivativo glotal com um valor RK grande.

- **Inclinação Espectral – R_a :**

Este parâmetro é dado pela razão entre o tempo em que ocorre a fase de retorno,

T_a , pelo período do pulso, ou seja, $\frac{T_a}{T_0}$. Esta medida relaciona-se com a maneira com que

as pregas vocais fazem contato durante a fonação, isto é, se instantaneamente ou gradualmente.

- **Parâmetro - R_g :**

Esta medida é dada pela razão entre o período do pulso pelo dobro do tempo da

duração da fase aberta, matematicamente é dado por: $\frac{T_0}{2T_p}$. Os componentes de baixa

frequência do espectro da fonte são geralmente reforçadas devido à forma do pulso. Geralmente, esta característica está presente na primeira harmônica. No entanto,

dependendo de algumas qualidades fonatórias, onde há um aumento no grau de tensão, a máxima elevação pode estar muito próxima da segunda harmônica. O parâmetro que captura o efeito deste "*reforço*" é F_g (frequência glotal)[FANT/1979 a] , que é definido como o inverso do dobro da duração da fase aberta.. O parâmetro R_g pode ser visto como uma versão da frequência glotal F_g .

- **Parâmetro SQ:**

Este parâmetro determina a razão entre o tempo no qual a glote permanece aberta (T_p) e o tempo no qual permanece fechada ($T_c - T_p$), matematicamente é indicada por:

$$SQ = \frac{T_p}{T_c - T_p} .$$

Acusticamente, esta medida relaciona a rapidez com que a pregas vocais se

fecham, indicando um fechamento mais abrupto ou mais suave.

- **Quociente de fechamento CQ:**

Este parâmetro é dado pela razão entre o tempo no qual a glote permanece fechada

e o período, ou seja, $CQ = \frac{T_c - T_p}{T_o}$.

- **Quociente de abertura - OQ:**

Este parâmetro mostra a proporção do ciclo do pulso onde a glote está aberta. Esta medida predita os valores das primeiras harmônicas, um aumento no valor de OQ está correlacionado com um aumento da amplitude dessas harmônicas no espectro vocal. O quociente de abertura, pode ser calculado alternativamente, como a razão da fase aberta

pelo período do pulso $\frac{T_c}{T_o}$ (figura 4.5).

A relação entre OQ e CQ é: $OQ = 1 - CQ$ e a relação entre R_k , R_g e OQ é dada por: $OQ = \left(\frac{1 + R_k}{2R_g} \right)$, implicando numa correlação positiva com R_k e uma correlação negativa com R_g .

Figura 4.5: Valores de OQ esquematizados: a) pulso derivativo glotal com um pequeno valor de OQ; b) pulso glotal derivativo com um maior valor para OQ

Percebe-se que R_a e R_k estão correlacionados negativamente com EE e há uma correlação positiva entre R_a e R_k . Através de correlações entre parâmetros, pode-se aumentar a flexibilidade e a precisão do modelo da forma de onda, apesar de dificultar a interpretação. Através disso, um parâmetro baseado na forma de onda R_d , é proposto, onde correlaciona parâmetros que definem o pulso derivativo glotal. Matematicamente é dado por $R_d = 1000 \cdot \frac{U_0}{EE} \cdot \frac{f_0}{110} = \frac{1}{0.11} \frac{T_d}{T_0}$, onde T_d é o tempo de declive em milisegundos.

Com isso o modelo LF pode ser determinado pelos parâmetros R_d , EE , R_a e R_k . Baseado na simplificação geométrica da forma do pulso do modelo LF, as relação entre R_d , R_a , R_g e R_k , pode ser aproximada por:

$$R_d = \frac{1}{0.11} (0.5 + 1.2 R_k) \left(\frac{R_k}{4 R_g} + R_a \right) \quad (4.5)$$

Caso o parâmetro R_d seja conhecido, os parâmetros R_a , R_g e R_k , são calculados pelas relações abaixo, proposto pelo algoritmo encontrado em [CHILDERS/1999]:

$$R_a = (-1 + 4.8 R_d) / 100 \quad (4.6)$$

$$R_k = (22.4 + 11.8 R_d) / 100 \quad (4.7)$$

$$R_g = (R_k / 4) / (0.11 R_d / (0.5 + 1.2 R_k) - R_a) \quad (4.8)$$

• Quociente no Domínio da Amplitude - AQ

Este parâmetro é determinado como a razão entre a amplitude do fluxo – AC da forma de onda glotal e a amplitude mínima da onda derivativa (equação 4.9 e figura 4.6). Através deste parâmetro, pode-se quantificar diferentes tipos de fonação [ALKU/1996].

$$AQ = \frac{a_{AC}}{d_{\min}} \quad (4.9)$$

Figura 4.6: a) Pulso Glotal e b) Pulso Glotal Derivativo

A principal justificativa para o uso deste parâmetro, está no fato de que a fonte glotal estimada não fornece valores precisos de a_{AC} e d_{min} (figura 4.6), ou seja, a onda estimada é uma versão distorcida do verdadeiro pulso glotal, onde o fluxo original tem sido multiplicado por um fator de atenuação desconhecido (C_M) e somado por um componente DC , desconhecido também (C_A). Consequentemente nenhuma dessas duas distorções afetam o valor de AQ , ou seja, multiplicando por C_M , há uma mudança na amplitude do pulso e igualmente em sua derivativa. Portanto a distorção causada por C_M não afeta o valor de AQ , já que AQ é uma relação de dois valores determinados a partir da onda e da sua derivativa. A distorção aditiva C_A não muda a amplitude- AC da onda. O efeito de C_A é também removido pela derivação, consequentemente, distorções causadas por C_A não afeta o valor de AQ .

- **Quociente de Amplitude Normalizada- NAQ:**

O parâmetro NAQ é apresentado por [ALKU/2002], como um método para parametrizar a fase de fechamento glotal, usando duas medidas de amplitude a partir da forma de onda estimada pelo filtro inverso. O autor garante que esse parâmetro apresenta maior robustez em vozes patológicas que geram pulsos glotais “distorcidos” e onde a extração de parâmetros temporais é totalmente dificultada pela presença de ruído.

$$NAQ = \frac{AQ}{T_o} \quad (4.10)$$

- **Parâmetros de extensões de Banda - B1 e B2:**

Levando-se em consideração a importância das extensões de banda dos pulsos glotais, os parâmetros, $B1 \cong \frac{\alpha}{\pi}$ e $B2 \cong \frac{1}{2\pi T_a}$ foram calculados, assim como as relações $B1/B2$ e $B2/B1$ na busca de um valor que pudesse discriminar diferentes patologias. Os parâmetros B1 e B2 são detalhados em [FANT/1986].

- **parâmetro ece :**

Este parâmetro relaciona os parâmetros β e T_a no intervalo entre T_e e T_c , é expresso por:

$$ece = e^{\left(-(\beta-1)(t_c - t_e)\right)} \quad (4.11)$$

4.3.3 PARÂMETROS PROPOSTOS

- **Parâmetros relacionados aos efeitos não-lineares do sub-sistema glotal, TR (TR_p e TR_s):**

Considerando a importância da fase de retorno, na classificação de diferentes tipos de vozes, principalmente naquelas onde ocorre abdução incompleta das pregas vocais ocasionando escape de ar, decidiu-se quantificar esta fase. É proposto neste trabalho dois novos parâmetros para a quantificação do tempo de retorno (TR) do pulso glotal. O primeiro parâmetro avalia o tempo de retorno de primeira ordem (TR_p), relacionando com as forças principais do sistema glotal e o tempo de retorno de ordem superior (TR_s) relacionando a onda da mucosa. Desta forma analisa-se separadamente os efeitos lineares e não-lineares do sistema glotal.

Os parâmetros são expressos matematicamente por:

$$TR_p = \frac{(T_e + T_a)}{T_e} \quad \text{e} \quad TR_s = \frac{T_c - (T_e + T_a)}{(T_e + T_a)} \quad (4.12)$$

A representação gráfica de TR_p e TR_s no pulso glotal derivativo estão mostrados nas figuras 4.7 a) e b).

Figura 4.7: relações esquematizadas (a) TRp e (b) TRs

Estas medidas estão relacionadas com a maneira na qual as pregas vocais se fecham durante a fonação, se instantaneamente, ou gradualmente, além de estimar a não-linearidade efetiva do sub-sistema glotal. Para uma voz soprosa, por exemplo, dependendo

do grau da patologia, ocorre um escape de ar que pode comprometer o fechamento da pregas vocais, desta maneira, pode-se esperar que este fechamento ocorra de maneira gradual, aumentando a fase de retorno do pulso e consequentemente um aumento em TR_p e diminuição de TR_s .

4.4 ANÁLISE ESTATÍSTICA

A análise estatística foi baseada na hipótese de que há uma diferença significativa em pelo menos um parâmetro de cada grupo, que possa discriminar as características das vozes. Para demonstrar essa hipótese, utilizou-se a análise de variância (ANOVA), onde seus parâmetros são ajustados pelo método dos mínimos quadrados[BOX/1978]. Quando o teste rejeita a hipótese de igualdade $H_0 = \theta_1 = \theta_2 = \dots = \theta_k$, afirma-se que ocorre pelo menos uma desigualdade, e só isso na maioria dos casos não é suficiente para uma decisão. Um modo de investigar a causa da rejeição é comparar os grupos dois a dois e para isso utilizou-se o teste t. Em ambas análises utilizou-se um nível de significância (α) de 0,05 .

5 RESULTADOS

5.1 CÁLCULO DO PULSO GLOTAI DERIVATIVO, ATRAVÉS DA FILTRAGEM INVERSA

Para a análise, pulsos glotais individuais foram selecionados a partir do filtro inverso e após a filtragem inversa, pôde-se realizar o ajuste do modelo LF por meio dos parâmetros temporais retirados do sinal derivativo real.

Através das técnicas descritas na fundamentação teórica, foram testadas com versões modificadas, diversos algoritmos de filtragem inversa existentes na literatura para a obtenção das características do pulso derivativo e dentre eles os descritos em [GABELMAN/2002] e [CHILDERS/1999]. Buscou-se uma comparação entre os resultados obtidos, na procura de um algoritmo que proporcionasse maior confiabilidade quantitativa nos resultados.

O período foi calculado através do sinal residual encontrado a partir da análise LPC, onde os coeficientes foram estimados pelo método da autocorrelação, com uma janela de comprimento de 256 ou 512 amostras, dependendo da voz. Para o cálculo dos formantes para o filtro inverso, a análise LPC foi usada tendo seus coeficientes estimados pelo método da covariância usando uma janela de 35 a 56 amostras, dependendo da voz. Esta análise, apesar de ser menos estável devido ao comprimento da janela ser menor que um período, para vozes não modais, particularmente vozes patológicas, onde os períodos variam muito, se torna mais adequado. E por fim o pulso derivativo resultante do filtro inverso, foi então ajustado com o modelo LF usando mínimos quadrados. A figura 5.1 mostra este processo:

²Figura 5.1: Cálculo do pulso glotal derivativo

2. http://www.surgery.medsch.ucla.edu/glottalaffairs/software_of_the_boga.htm

5.2 ANÁLISE DOS PARÂMETROS ACÚSTICOS

Alguns parâmetros usados no modelo LF, em nível de teste, foram ajustados manualmente, apesar da possibilidade de erros serem introduzidos no filtro inverso [EPSTEIN et al./1999b], o ajuste manual proporciona características visuais mais próximas do pulso original, o que colabora qualitativamente para classificação entre vozes normais e patológicas além de facilitar a geração de uma voz sintética mais semelhante à voz original. Como o objetivo do trabalho é a busca de um padrão de modelo, baseado inclusive

na forma de onda do pulso derivativo, o ajuste manual pode colaborar para o desenvolvimento do modelo.

A análise dos parâmetros foi baseada primeiramente em vozes normais procurando um ajuste paramétrico que posteriormente, pudesse proporcionar uma possível discriminação em vozes patológicas e desta maneira obter um modelo mais coerente dentro dos padrões fisiológicos reais das pregas vocais, sendo que, os parâmetros foram discriminados baseados no estado físico que caracteriza a patologia e não nas características de qualidade vocal.

Dentre os parâmetros analisados, uma atenção especial foi dedicada aos valores dos parâmetros que descrevem a forma de onda glotal (ω_0 , α , β , E_0), aos parâmetros B1 e B2 [CHILDERS/1999] que relacionam as extensões de bandas dos segmentos 1 (intervalo entre T_0 e T_p) e 2 (intervalo entre T_p e T_c) respectivamente do pulso derivativo LF, ao parâmetro *ece* [CHILDERS/1995], que relaciona os parâmetros β e T_a , no intervalo T_e e T_c . Pode-se observar que apesar desses parâmetros apresentarem características evidentes na diferenciação das patologias, a avaliação dessas medidas não estão reportadas na literatura.

Parâmetros adicionais foram propostos neste trabalho na busca de um valor efetivo na discriminação de vozes normais e patológicas. Como são parâmetros que buscam uma relação com a fase de retorno, foram “batizados” por parâmetros TR (TR_p e TR_s). E por fim, as análises para a discriminação paramétrica de vozes normais e vozes com Edema de Reinke e Nódulos. Nesta etapa, foi feita uma análise rigorosa das vozes com Edema de Reinke e Nódulos.

5.2.1 MATERIAIS E MÉTODOS

● ETAPA 1

Nesta etapa do trabalho, foi dado prioridade a vozes normais (vogal /a/ sustentada) na busca de uma maior robustez nos algoritmos e determinação de padrões normais em sinais reais.

Foram selecionadas 8 pessoas, diagnosticadas através da videolaringostroboscopia realizada no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto – Usp, sendo 4 do sexo feminino (F1 à F4) e 4 do sexo masculino (M1 à M4), uma porção do sinal da vogal /a/ sustentada pode ser visualizada através das figuras 5.2 (a-d) e figuras 5.3 (a-d), onde o eixo x está relacionado ao tempo e o eixo y, relaciona-se a amplitude do sinal.

- VOZES FEMININAS ($F_1 - F_4$)

(a) sinal vocálico /a/ - F_1

(b) sinal vocálico /a/ - F_2

(c) sinal vocálico /a/ - F_3

(d) sinal vocálico /a/ - F_4

figuras 5. 2 (a-d) – vozes femininas normais

Pode-se perceber que apesar de serem diagnosticadas como vozes normais, o sinal difere para cada voz, isto talvez em consequência da alteração da mucosa que reveste as pregas vocais, ocasionando deste modo um padrão diferente para cada sinal.

- VOZES MASCULINAS ($M_1 - M_4$)

Figuras 5.3 (a-d): vozes masculinas normais

Para as vozes masculinas, segue a mesma justificativa dada para as vozes femininas devido à diferença encontrada no padrão do sinal vocálico.

A seguir, através da filtragem inversa, foi encontrado o pulso derivativo e consequentemente encontrado os parâmetros que modelam esse pulso, podendo então

através desses parâmetros gerar o pulso derivativo determinado pelo modelo LF e assim um ajuste comparativo dos sinais. Esse ajuste pode ser visualizado através das figuras 5.4 (a-d) e 5.5 (a-d).

- PULSOS DERIVATIVOS FEMININOS ($F_1 - F_2$)

figuras 5.4 (a-d): pulsos derivativos femininos

- PULSOS DERIVATIVOS MASCULINOS ($M_1 - M_2$)

Figuras 5.5 (a-d): pulsos derivativos masculinos

Verifica-se através das figuras acima uma ponderável parcela da oscilação (“ripple”), ocasionado, talvez pelo acoplamento não linear dos sistemas glotal e supra-glotal, o que não ocorre no modelo LF, já que este é ajustado linearmente não levando em consideração os acoplamentos. Uma possível origem deste acoplamento pode ser devido às falsas pregas vocais que estariam contribuindo com uma modulação na fonte glotal. Outra possível explicação é a formação da complexa onda mucosa durante a fase de abertura da glote.

Uma especial atenção foi dada à voz F2, já que seguindo os mesmos padrões de análise das restantes, observa-se que no pulso derivativo encontrado, há uma quantidade expressiva de componentes ruidosas que pode ser explicado por um distúrbio de ressonância. Uma possível justificativa para essa alteração funcional, poderia ser uma incompetência velofaríngea.

Diante deste fato, houve a necessidade de se fazer uma análise acústica do sinal vocálico de cada pessoa. Como citado na teoria, esta análise foi feita através do programa Analise de Voz, o que proporcionou uma avaliação não invasiva das condições da laringe. O resultado desta análise está descrito na tabela 5.1:

Tabela 5.1: Parâmetros Acústicos

Parâmetros-Valores de normalidade	F1	F2	F3	F4	M1	M2	M3	M4
F₀ :Homem-160 Mulher-250	215.74	184.71	204.91	204.43	121.26	139.95	93.6	114.56
Jitter: 2 %	0.63	0.23	3.97	0.57	0.15	2.62	0.17	0.29
Shimmer: 8 %	8.2	7.08	10.7	8.09	4.34	10.31	4.33	6.58
SFR: -5.0	-12.31	-10.13	-8.86	-11.76	-10.46	-7.4	-6.99	-9.31
SFF: 0.3	-15.43	-8.1	-16.73	-17.95	-15.3	-17.66	-18.11	-20.88
PA: 0.3	0.53	0.67	0.47	0.59	0.62	0.38	0.39	0.43
Ex: 2.0	22.18	6.51	16.28	17.91	21.6	11.42	21.42	44.17

Analisando os resultados da tabela acima, pode-se observar uma expressiva diferença nos valores dos parâmetros SFF e Ex. Apesar de estarem dentro do padrão de normalidade estipulado pelo programa Análise de Voz, há uma diferença significativa em relação aos demais pacientes. Sendo o parâmetro SFF relacionado com a quantidade de ruído do espectro do filtro e o parâmetro Ex uma medida correlacionada com a relação sinal/ruído, esses valores possivelmente explicam um escape de ar devido a interação trato vocal e glote, justificando desta forma o pulso derivativo ruidoso observado na análise.

● ETAPA 2

Como a etapa anterior baseou-se na análise qualitativa, nesta etapa, foram discriminados os parâmetros quantitativamente. Para isso, foram selecionados novos 15 indivíduos com vozes consideradas normais diagnosticadas pela videolaringostroboscopia realizada no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto – USP, não discriminando em relação ao sexo nem em relação à idade dos pacientes, pois esses detalhes não são relevantes para a análise proposta, já que não interferem na forma glotal.

Como foi citado, para o ajuste do modelo LF, levou-se em consideração os parâmetros retirados do sinal derivativo glotal real. Para uma possível corroboração dos

resultados encontrados, já que são explorados na literatura, a análise iniciou-se com vozes normais, separadas entre homens e mulheres.

Os parâmetros foram comparados através das médias aritméticas e como foi citado na fundamentação teórica, os parâmetros foram divididos entre parâmetros T, R, Q e parâmetros relacionados a forma da derivativa da onda glotal. Os resultados para as vozes normais podem ser visualizados através da tabela 5.2 e das figuras 5.17 (a-e).

Tabela 5.2: médias dos parâmetros para vozes normais

Vozes Normais			
1	médias	homens	mulheres
2	Fo	132,11	211
3	Ee	0,98	0,94
4	Te	54,89	36,33
5	U0	11,98	10,37
6	Rd	1,44	2,09
7	Ra	0,06	0,09
8	Rk	0,39	0,47
9	Rg	0,96	0,94
10	Td	12,19	11,06
11	T0	76,33	47,17
12	Ta	4,49	4,26
13	Tp	39,93	25,07
14	Te	54,89	36,33
15	Tc	76,33	47,17
16	Pp-período pitch	77,22	47,83
17	td	12,19	11,06
18	tc	76,33	47,17
19	tp	45,05	24,57
20	te	58,78	33,00
21	ta	4,49	4,26
22	e0	0,53	0,65
23	α	0,02	0,02
24	ω_0	0,07	0,13
25	β	0,22	0,23
26	SQ	1,43	1,10
27	OQ	0,99	0,99
28	AQ	12,19	11,06
29	NAQ	0,16	0,23
30	CQ	0,41	0,47

Na tabela 5.2, os parâmetros T, estão diferenciados pelas letras T e t, o que correspondem aos resultados dos dados reais e dos ajustados pelo modelo LF, respectivamente.

As figuras 5.6 (a - e), trazem a visualização dos resultados obtidos:

Figuras 5.6 (a – e) : Parâmetros relacionados a vozes femininas e masculinas normais

Em relação aos parâmetros T, observa-se que para o ajuste do modelo LF, foram necessárias pequenas alterações, que foram feitas manualmente.

Os valores de R_d encontrados, mostram-se coerentes segundo [Fant/1997], que afirma que R_d para homens se encontra numa faixa de 0.5-1.5 e para mulheres numa faixa de 0.8-2.5. Um valor padrão proposto pelo autor, está em torno de 0.7. Confirmando também que um aumento no valor R_d implica em um aumento nos valores de R_k , R_a e um decréscimo no valor de R_g .

Em relação aos parâmetros Q, os resultados são compatíveis aos apresentados por [ALKU/1996].

Analisando a correlação entre os valores relacionados aos parâmetros CQ e NAQ, observa-se que tanto para homens, como para mulheres, CQ apresenta valores mais altos do que NAQ, confirmando os resultados propostos por [Alku/2002], que afirma que NAQ é aproximadamente 50% menor que CQ.

• ETAPA 3

Diante da robustez encontrada no resultados dos algoritmos propostos para a análise de vozes normais, neste tópico serão apresentados os resultados do ajuste do modelo LF tanto para a representação de um sinal normal, de um sinal contendo patologia, bem como a discriminação destas patologias. O ajuste foi realizado sobre parâmetros retirados dos sinais derivativos reais e analisado estatisticamente para garantir seu nível de significância.

Esta etapa do trabalho discorre sobre a análise aplicada a todo o banco de vozes disponível. Foram selecionadas 82 pessoas diagnosticadas através da videoestrobolaringoscopia realizada no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto – USP, não discriminando em relação ao sexo nem em relação à idade dos pacientes, pois esses detalhes não são relevantes para a análise proposta, já que não interferem na forma glotal. A tabela 5.3 apresenta os tipos de vozes que foram coletadas e a quantidade destas.

Tabela 5.3: relação dos tipos de vozes e quantidade de casos

Tipo de Vozes	Número de Casos
Normal	48
Edema de Reinke	9
Nódulo	25

Estas patologias foram seleccionadas, (Edema de Reinke e nódulo) por serem as de maior incidência cotidiana.

Os valores médios dos parâmetros relacionados às vozes normais, com Nódulo e edema de Reinke podem ser conferidos na tabela 5.4.

Tabela 5.4: parâmetros relacionados as vozes normais, nódulo e Reinke.

	normal	nódulo	Reinke
Fo	196,83	237,88	154,56
Ee	31271,28	30749,16	32067,22
U0	340524,7	306490,7	380274,6
Rd	1,89	2,07	1,62
Ra	0,08	0,09	0,07
Rk	0,45	0,47	0,42
Rg	0,94	0,94	0,95
To	53,35	43,52	68,67
Ta	4,19	3,81	4,52
Tp	28,27	23,09	35,85
Te	40,31	33,32	49,89
Tc	53,35	43,52	68,67
Td	10,87	9,92	11,86
pp	53,87	44,52	69,55
tc	53,35	43,52	68,67
tp	29,13	24,08	44,03
te	37,23	32,39	56,5
ta	4,19	3,81	4,52
eo	2,07E+04	2,08E+04	1,84E+04
α	0,02	0,019	0,017
wg	0,12	0,14	0,079
β	0,24	0,25	0,21
ece	0,03	0,099	0,16
B1	0,006	0,0059	0,0053
B2	6,59	5,98	7,09
B1/B2	0,001	0,0011	0,00089
B2/B1	1197,08	1430,02	1919,44
SQ	1,18	1,22	1,78
OQ	0,77	0,82	0,88
AQ	10,87	9,92	11,86
NAQ	0,21	0,23	0,17
CQ	0,45	0,44	0,36
TRp	1,12	1,12	1,08
TRs	0,31	0,22	0,14

Para validar os resultados encontrados na etapa anterior, houve um aumento no número de indivíduos com vozes normais e novamente as análises para esse tipo de voz confirmaram os encontrados anteriormente, que se baseavam em resultados encontrados na literatura.

Buscando uma originalidade no modelamento dos sinais glotais derivativos patológicos, o interesse da análise, aqui está situado na determinação do estado físico da laringe, já que a quantificação paramétrica existente na literatura associa-se às características acústicas. Em cima deste fato, a análise, baseou-se na quantificação dos seguintes parâmetros:

- de forma (E_0 , α , β , ω_0), por serem independentes do período e por estarem intrinsecamente associados ao movimento glotal;
- relacionados a extensão de banda, ($B1/B2$ e $B2/B1$);
- ece ; e
- relacionados aos efeitos não-lineares do sub-sistema glotal, TR (TRp e TRs)

A análise para o ajuste do modelo LF baseou-se principalmente na hipótese de que há uma diferença significativa em pelo menos 1 parâmetro dentro do grupo em análise que se pudesse discriminar o tipo de voz. Para demonstrar essa hipótese utilizou-se a análise variância (ANOVA), com nível de significância de 0,05. Para os parâmetros onde houve um resultado estatisticamente significativo utilizou-se o teste-t, pelo método Welch-Aspin [DRETZKE/1998], com $\alpha=0,05$, comparando os grupos dois a dois na intenção de investigar a causa da rejeição da hipótese de igualdade.

Os resultados foram avaliados seguindo:

- comparação visual das médias dos parâmetros, através de gráficos;
- tabela ANOVA;
- teste-t; e
- interpretação dos resultados obtidos.

- AVALIAÇÃO DO MODELO LF PARA VOZES NORMAIS E PATOLÓGICAS (EDEMA DE REINKE E NÓDULO)

PARÂMETROS DE FORMA (e_0 , α , β , ω_0)

A análise desses parâmetros já foi comentada numa primeira investigação, nesta seção, serão mostradas as análises estatísticas.

Gráficos de médias – figuras 5.7 (a) – (d)

Figura 5.7 (a) – (d): Comparação das médias dos parâmetros de forma para vozes normais, com nódulo e edema de Reinke.

Pelos gráficos das médias (figuras 5.7 (a) – (d)), pode-se observar que existe uma diferença significativa entre os parâmetros de diferentes tipos de vozes, com exceção do $e_{0\text{normal}}$ e $e_{0\text{nódulo}}$ que aparentam ter a mesma média. Portanto esse teste poderia sugerir uma possível discriminação entre vozes normais e vozes com edema de Reinke e entre as patologias.

A seguir, continuando os testes para uma avaliação das diferenças discriminativas entre os parâmetros, o teste estatístico ANOVA, (tabelas 5.5 à 5.8) foi aplicado.

• ANOVA – e_0

Tabela 5.5: ANOVA para o parâmetro e_0

RESUMO

Grupo	Contagem	Soma	Média	Variância
e_0 normal	48	995216,6	20733,68	24439164
e_0 nodulo	25	520909	20836,36	14341957
e_0 reinke	9	165482,6	18386,96	18162290

ANOVA

Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	45629637	2	22814819	1,10025	0,337832	3,112262
Dentro dos grupos	1,64E+09	79	20736025			
Total	1,68E+09	81				

A ANOVA para o parâmetro e_0 mostra que para a análise da variância, o parâmetro MQG (quadrado médio entre os grupos) não é muito superior ao MQR (quadrado médio dentro do grupo), o que indica que não há uma variabilidade entre os grupos. Confirmando esse resultado, observa-se que $F \leq F_{\text{crítico}}$, o que indica que num nível de significância de 95%, não há uma diferença significativa entre as médias dos parâmetros.

- ANOVA – α

Tabela 5.6: ANOVA para o parâmetro α

RESUMO

Grupo	Contagem	Soma	Média	Variância
alpha normal	48	0,9647	0,020098	3,14E-05
alpha nodulo	25	0,4651	0,018604	4,14E-05
alpha reinke	9	0,1504	0,016711	6,92E-05

ANOVA

Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,000103	2	5,15E-05	1,345976	0,266188	3,112262
Dentro dos grupos	0,00302	79	3,82E-05			
Total	0,003123	81				

- ANOVA – β

Tabela 5.7: ANOVA para o parâmetro β

RESUMO

Grupo	Contagem	Soma	Média	Variância
beta normal	48	11,5431	0,240481	0,00269
beta nodulo	25	6,2013	0,248052	0,004175
beta reinke	9	1,8718	0,207978	0,012003

ANOVA

Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,010811	2	0,005406	1,323404	0,272064	3,112262
Dentro dos grupos	0,322683	79	0,004085			
Total	0,333494	81				

O resultado da ANOVA para os parâmetros α e β traz a mesma justificativa para o resultado da ANOVA do parâmetro e_0 , ou seja, para a análise da variância, o parâmetro MQG (quadrado médio entre os grupos) não é muito superior ao MQR (quadrado médio dentro do grupo), o que indica não variabilidade entre os grupos. E em relação ao teste F,

observa-se que $F \leq F_{\text{crítico}}$, o que indica que para um nível de significância de 95%, não há uma diferença significativa entre as médias dos parâmetros.

- **ANOVA – wg**

Tabela 5.8: ANOVA para o parâmetro wg.

RESUMO

Grupo	Contagem	Soma	Média	Variância
wg normal	48	5,6948	0,118642	0,001141
wg nodulo	25	3,5557	0,142228	0,001239
wg reinke	9	0,7095	0,078833	0,000779

ANOVA

Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,027517	2	0,013759	12,13259	2,54E-05	3,112262
Dentro dos grupos	0,089588	79	0,001134			
Total	0,117105	81				

Para o parâmetro wg, observa-se que em relação à variância, há uma diferença significativa entre os grupos, já que MQG é bem maior que MQR. Em relação à média, o teste F também indica diferença significativa, já que $F \geq F_{\text{crítico}}$ e a probabilidade de encontrar um $F \geq 12,13$ é muito pequena em relação ao nível de significância de 95%. Portanto, para os parâmetros $e0$, α , β e wg, o resultado da ANOVA não comprova os resultados sugeridos visualmente, concluindo que há uma diferença entre os valores desse parâmetros para os três tipos de vozes.

Para identificar quais as médias que diferem entre si para parâmetro wg, o próximo passo será análise das médias pelo Teste-t.

• **Teste-t - wg**

Tabela 5.9 (a) – (c): Teste-t para o parâmetro wg

(a)	wg reinke	wg normal
Média	0,078833333	0,118641667
Variância	0,000778513	0,001140804
Observações	9	48
Hipótese da diferença de média	0	
gl	13	
Stat t	-3,790957934	
P(T<=t) bilateral	0,002245581	
t crítico bilateral	2,16036824	

(b)	wg reinke	wg nodulo
Média	0,078833333	0,142228
Variância	0,000778513	0,001239252
Observações	9	25
Hipótese da diferença de média	0	
gl	18	
Stat t	-5,434619963	
P(T<=t) bilateral	3,66021E-05	
t crítico bilateral	2,100923666	

(c)	wg normal	wg nodulo
Média	0,118641667	0,142228
Variância	0,001140804	0,001239252
Observações	48	25
Variância agrupada	0,001174082	
Hipótese da diferença de média	0	
gl	71	
Stat t	-2,790874738	
P(T<=t) bilateral	0,006746639	
t crítico bilateral	1,993944352	

O teste-t (tabelas 5.9 (a) – (c)), foi realizado com a finalidade de se testar as médias referentes ao parâmetro wg para os três tipos de vozes. Com um nível de significância $\alpha = 0,05$ o resultado pode ser considerado estatisticamente significativo, pois observa-se nas três comparações, que as probabilidades de aceitar a hipótese nula tem valores menores que α , semelhantemente os valores críticos nos três casos, são menores que os

valores absolutos dos t 's obtidos estatisticamente. Com isso rejeita-se a hipótese que as médias são iguais, concluindo que há uma diferença entre os valores desse parâmetro para os três tipos de vozes.

RESULTADOS DOS PARÂMETROS DE FORMA

Um resumo dos testes estatísticos usado na discriminação entre os parâmetros de forma, é mostrado na tabela 5.10.

Tabela 5.10: comparação dos resultados dos parâmetros $e0$, α , β , wg .

Diferenças Significativas

grupo\testes	Gráfico das médias	ANOVA	Teste-t
$e0$	Normal/reinke	Não há	
	Nodulo/reinke		
α	Normal/reinke	Não há	
	Nodulo/reinke		
	Normal/nodulo		
β	Normal/reinke	Não há	
	Nodulo/reinke		
	Normal/nodulo		
wg	Normal/reinke	Normal/reinke	Normal/reinke
	Nodulo/reinke	Nodulo/reinke	Nodulo/reinke
	Normal/nodulo	Normal/nodulo	Normal/nodulo

PARÂMETROS RELACIONADOS À EXTENSÃO DE BANDA ($B1$, $B2$, $B1/B2$ e $B2/B1$)

Gráficos de médias – figuras 5.8 (a) – (d)

(a)

(b)

(c)

(d)

Figuras 5.8 (a) – (d): Comparação das médias dos parâmetros de banda para vozes normais, com nódulo e edema de Reinke.

Através dos gráficos das médias, percebe-se que em relação aos parâmetros $B1$ e $B1/B2$, não há uma diferença muito grande entre as médias para diferentes tipos de vozes. O parâmetro $B2$, pode sugerir uma possível discriminação entre as médias das vozes

B2normal em relação à B2nódulo, B2normal em relação à B2reinke e B2nódulo em relação à B2reinke, e em relação ao parâmetro B2/B1, a discriminação poderia ser sugerida através dos parâmetros B2B1normal em relação à B2/B1reinke e B2/B1nódulo em relação à B2/B1reinke.

Para buscar uma referência mais sólida, o próximo passo, foi a análise estatística, com os testes ANOVA e t.

• ANOVA B1

Tabela 5.11: ANOVA para o parâmetro B1

RESUMO

Grupo	Contagem	Soma	Média	Variância
B1 normal	48	0,3081	0,006419	3,21E-06
B1 nodulo	25	0,1482	0,005928	4,21E-06
B1 reinke	9	0,048	0,005333	7,12E-06

ANOVA

Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	1,07E-05	2	5,35E-06	1,367884	0,26061	3,112262
Dentro dos grupos	0,000309	79	3,91E-06			
Total	0,00032	81				

• ANOVA B2

Tabela 5.12: ANOVA para o parâmetro B2

RESUMO

Grupo	Contagem	Soma	Média	Variância
B2 normal	48	316,1295	6,586031	1,604042
B2 nodulo	25	149,5124	5,980496	1,990053
B2 reinke	9	63,8472	7,094133	3,650943

ANOVA

Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	10,129	2	5,064502	2,626009	0,078677	3,112262
Dentro dos grupos	152,3588	79	1,928592			
Total	162,4878	81				

• **ANOVA B1/B2**

Tabela 5.13: ANOVA para o parâmetro B1/B2

RESUMO						
Grupo	Contagem	Soma	Média	Variância		
B1/B2 normal	48	0,049354	0,001028	1,49E-07		
B1/B2 nodulo	25	0,027198	0,001088	2,4E-07		
B1/B2 reinke	9	0,00805	0,000894	4,51E-07		
ANOVA						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	2,49E-07	2	1,24E-07	0,601155	0,550666	3,112262
Dentro dos grupos	1,64E-05	79	2,07E-07			
Total	1,66E-05	81				

A ANOVA para os parâmetros B1, B2 e B1/B2, mostra que para a análise da variância, o parâmetro MQG (quadrado médio entre os grupos) é inferior ao MQR (quadrado médio dentro do grupo), o que indica que não há uma variabilidade entre os grupos. Confirmando esse resultado, observa-se que $F \leq F_{\text{crítico}}$, o que indica que num nível de significância de 95%, não há uma diferença significativa entre as médias dos parâmetros.

- **ANOVA B2/B1**

Tabela 5.14: ANOVA para o parâmetro B2B1

RESUMO

<i>Grupo</i>	<i>Contagem</i>	<i>Soma</i>	<i>Média</i>	<i>Variância</i>
B2B1 normal	48	57459,94	1197,082	526657,7
B2B1 nodulo	25	35750,43	1430,017	1919828
B2B1 reinke	9	17274,99	1919,443	1883702

ANOVA

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Entre grupos	4200320	2	2100160	1,931498	0,151714	3,112262
Dentro dos grupos	85898409	79	1087322			
Total	90098729	81				

A ANOVA para o parâmetro B2/B1, em relação à variância, o parâmetro MQG é superior à MQR. Por outro lado, observa-se que $F \leq F_{\text{crítico}}$, o que indica que num nível de significância de 95%, não há uma diferença significativa entre as médias dos parâmetros.

Portanto, graficamente, poderia até sugerir uma possível diferença entre os parâmetros, o que não está comprovado pela ANOVA, pois para todos os parâmetros entre os grupos, não foi encontrado um valor estatisticamente significativo, o que pode ser comprovado pelas tabelas 5.11 à 5.14. Diante deste fato, não houve a necessidade de se realizar o teste t. Enfim, dentro deste grupo, não houveram valores discriminativos para diferentes vozes.

- Resultados dos parâmetros B1, B2, B1/B2 e B2/B1 – tabela 5.15

Tabela 5.15: comparação dos resultados dos parâmetros B1, B2, B1/B2 e B2/B1.

Diferenças Significativas		
grupos\testes	visual	ANOVA
B1	Não há	Não há
B2	normal/nódulo nódulo/reinke	Não há
B1/B2	normal/reinke nodulo/reinke	Não há
B2/B1	normal/reinke nodulo/reinke	Não há

PARÂMETRO ECE

Gráfico de médias – figura 5.9

Figura 5.9: Comparação das médias do parâmetro ece vozes normais, com nódulo edema de Reinke

Graficamente, percebe-se que há uma diferença entre os parâmetros ece para diferentes vozes. Para confirmar essa hipótese, foram realizados os testes estatísticos, ANOVA e t.

- **ANOVA – ece**

Tabela 5.16: ANOVA para o parâmetro ece

RESUMO						
Grupo	Contagem	Soma	Média	Variância		
ece normal	48	1,4976	0,0312	0,000773		
ece nodulo	25	2,471	0,09884	0,014933		
ece reinke	9	1,4591	0,162122	0,048373		
ANOVA						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,168244	2	0,084122	8,501631	0,000453	3,112262
Dentro dos grupos	0,781689	79	0,009895			
Total	0,949933	81				

Para o parâmetro ece, observa-se que em relação à variância, há uma diferença significativa entre os grupos, já que MQG é bem maior que MQR. Em relação à média, o teste F também indica diferença significativa, já que $F \geq F_{\text{crítico}}$ e a probabilidade de encontrar um $F \geq 8,50$ é muito pequena em relação ao nível de significância de 95%. Portanto, para o parâmetro ece, o resultado da ANOVA comprova os resultados sugeridos visualmente, concluindo que há uma diferença entre os valores desse parâmetros para os três tipos de vozes.

Para identificar quais as médias que diferem entre si para parâmetro ece, o próximo passo analisou as médias pelo Teste-t.

• **Teste-t – tabela 5.17 (a) – (c)**

Tabela 5.17 (a) – (c): Teste-t para o parâmetro wg

(a)	<i>ece reinke</i>	<i>ece normal</i>
Média	0,162122222	0,0312
Variância	0,048373004	0,000772771
Observações	9	48
Hipótese da diferença de média	0	
gl	8	
Stat t	1,78313256	
P(T<=t) bilateral	0,112411858	
t crítico bilateral	2,306005626	

(b)	<i>ece reinke</i>	<i>ece nodulo</i>
Média	0,162122222	0,09884
Variância	0,048373004	0,014932707
Observações	9	25
Hipótese da diferença de média	0	
gl	10	
Stat t	0,818877004	
P(T<=t) bilateral	0,431938372	
t crítico bilateral	2,228139238	

(c)	<i>ece normal</i>	<i>ece nodulo</i>
Média	0,0312	0,09884
Variância	0,000772771	0,014932707
Observações	48	25
Hipótese da diferença de média	0	
gl	25	
Stat t	-2,731045925	
P(T<=t) bilateral	0,011406445	
t crítico bilateral	2,05953711	

Através do teste-t (tabelas 5.17 (a) – (c)), foi feito o teste para as médias referentes ao parâmetro ece para os três tipos de vozes. Com um nível de significância $\alpha = 0,05$ o resultado pode ser considerado estatisticamente significativo, pois observa-se nas três comparações, que as probabilidades de aceitar a hipótese nula tem valores menores que α .

Por outro lado, a análise em relação à comparação $t_{\text{crítico}}$ e t_{obtido} , o único valor discriminativo é em relação aos parâmetros e_{cenormal} e e_{cenodulo} .

• **RESULTADO DO PARÂMETROS ECE – TABELA 5.18**

Tabela 5.18: comparação dos resultados do parâmetro *ece*.

grupo\testes	Diferenças Significativas		
	visual	ANOVA	Teste-t
ece	normal/reinke	normal/reinke	normal/nodulo
	nodulo/reinke	nodulo/reinke	
	normal/nodulo	normal/nodulo	

Apesar de graficamente e o teste ANOVA, discriminarem estatisticamente todos os parâmetros dentro do grupo, o teste-t, só aceita a hipótese de diferença entre os parâmetros *ece* para voz normal e *ece* para nodulo.

PARÂMETROS RELACIONADOS AOS TEMPOS - TR (TR_p E TR_s)

Estes parâmetros estão relacionados aos efeitos não-lineares do sub-sistema glotal

Graficamente – figuras 5.10 (a) – (b)

Figuras 5.10 (a) – (b): Comparação das médias dos parâmetros TR_p e TR_s, para vozes normais, com nódulo e edema de Reinke.

Graficamente, pode-se perceber que em relação ao parâmetro TR_p, não há uma diferença capaz de discriminar entre vozes normais e com nódulo. Já em relação ao parâmetro TR_s, pode-se sugerir que há uma diferença entre todos os grupos.

A seguir a análise estatística para esses parâmetros com os testes ANOVA e t.

• ANOVA – TRp (tabela 5.19)

Tabela 5.19: ANOVA para o parâmetro TRp

RESUMO

<i>Grupo</i>	<i>Contagem</i>	<i>Soma</i>	<i>Média</i>	<i>Variância</i>
trpnormal	48	53,82124	1,121276	0,001261
trpnodulo	25	28,08084	1,123233	0,000707
trpreinke	9	9,765187	1,085021	0,000928

ANOVA

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Entre grupos	0,010987	2	0,005494	5,187645	0,007642	3,112262
Dentro dos grupos	0,083661	79	0,001059			
Total	0,094649	81				

• ANOVA – TRs (tabela 5.20)

Tabela 5.20: ANOVA para o parâmetro TRs

RESUMO

<i>Grupo</i>	<i>Contagem</i>	<i>Soma</i>	<i>Média</i>	<i>Variância</i>
trsnormal	48	15,07557	0,314074	0,014185
trsnodulo	25	5,432265	0,217291	0,007479
trsreinke	9	1,307952	0,145328	0,008777

ANOVA

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Entre grupos	0,301306	2	0,150653	12,98742	3,33E-05	3,112262
Dentro dos grupos	0,916393	79	0,0116			
Total	1,217699	81				

Para os parâmetro TRp e TRs, em relação à variância, há uma diferença significativa entre os grupos, já que MQG é maior que MQR. Em relação à média, o teste F também indica diferença significativa, já que $F \geq F_{\text{crítico}}$ e as probabilidades de se encontrar um $F \geq 5,19$ para o parâmetro TRp e um $F \geq 12,99$ para o parâmetro TRs são muito pequenas em relação ao nível de significância de 95% . Portanto, para os parâmetros TRp e TRs, o resultado da ANOVA mostra que há uma diferença entre os parâmetros, podendo, deste modo, serem aptos para a discriminação entre as vozes.

O próximo passo é analisar através do teste t, a diferença significativa entre as médias discriminadas no teste ANOVA.

- **Teste-t**

TRp – tabelas 5.21 (a) – (c)

Tabela 5.21 (a) – (c): Teste-t para o parâmetro TRp

(a)	<i>trpnormal</i>	<i>trpnodulo</i>
Média	1,121276	1,123233
Variância	0,001261	0,000707
Observações	48	25
Hipótese da diferença de média	0	
gl	62	
Stat t	-0,26509	
P(T<=t)	0,791818	
bilateral		
t crítico	1,998969	
bilateral		
(b)	<i>trpnormal</i>	<i>trpreinke</i>
Média	1,121276	1,085021
Variância	0,001261	0,000928
Observações	48	9
Hipótese da diferença de média	0	
gl	12	
Stat t	3,186874	
P(T<=t)	0,007821	
bilateral		
t crítico	2,178813	
bilateral		
(c)	<i>trpnodulo</i>	<i>Trpreinke</i>
Média	1,123233	1,085021
Variância	0,000707	0,000928
Observações	25	9
Hipótese da diferença de média	0	
gl	13	
Stat t	3,333499	
P(T<=t)	0,005389	
bilateral		
t crítico	2,160368	
bilateral		

Pelo teste t, o parâmetro de primeira ordem (TRp) é efetivo na discriminação do Edema de Reinke, porém mostrou-se ineficaz na detecção do nódulo.

TRs – tabelas 5.22 (a) – (c)

Tabela 5.22 (a) – (c): Teste-t para o parâmetro TRs

(a)	<i>trsnormal</i>	<i>trsnodulo</i>
Média	0,314074	0,217291
Variância	0,014185	0,007479
Observações	48	25
Hipótese da diferença de média	0	
gl	63	
Stat t	3,968812	
P(T<=t) bilateral	0,000188	
t crítico bilateral	1,998342	
(b)	<i>trsnormal</i>	<i>trsreinke</i>
Média	0,314074	0,145328
Variância	0,014185	0,008777
Observações	48	9
Hipótese da diferença de média	0	
gl	13	
Stat t	4,733775	
P(T<=t) bilateral	0,000391	
t crítico bilateral	2,160368	
(c)	<i>trsnodulo</i>	<i>trsreinke</i>
Média	0,217291	0,145328
Variância	0,007479	0,008777
Observações	25	9
Hipótese da diferença de média	0	
gl	13	
Stat t	2,015838	
P(T<=t)	0,0	
bilateral	64969	
t crítico	2,1	
bilateral	60368	

Em relação ao parâmetro de ordem superior, o teste t , mostra que TRs é um excelente detetor de vozes patológicas (nódulo e Edema de Reinke), porém não é capaz de discriminar as patologias. A avaliação do TRs, computa as não linearidades reais do sistema glotal.

A seguir a tabela 5.23, traz um resumo da análise em relação aos parâmetros TRp e TRs.

- **RESULTADOS DOS PARÂMETROS TRP E TRS**

Tabela 5.23: comparação dos resultados dos parâmetros TRp e TRs.

grupos\testes	Diferenças Significativas		
	Gráfico	ANOVA	Teste-t
TRp	Normal/reinke	Normal/nodulo	Normal/reinke
	Reinke/nodulo	Normal/reinke	Reinke/nodulo
		Reinke/nodulo	
TRs	Normal/nodulo	Normal/nodulo	Normal/nodulo
	Normal/reinke	Normal/reinke	Normal/reinke
	Reinke/nodulo	Reinke/nodulo	

Os possíveis valores médios para a discriminação entre as vozes aqui estudadas, são dados na tabela 5.24:

Tabela 5.24: Possíveis resultados para a discriminação entre vozes normais, com nódulo e com edema de Reinke.

Parâmetros	Resultados Discriminativos	Valores médios de referência
w_g	Normal/reinke	0,12/0,08
	Nodulo/reinke	0,14/0,079
	Normal/nodulo	0,12/0,14
ece	Normal/nodulo	0,03/0,1
TRp	Normal/reinke	1,12/1,08
	Nodulo/reinke	1,12/1,08
TRs	Normal/reinke	0,31/0,14
	Nodulo/reinke	0,22/0,14

6 CONCLUSÕES

Na busca de uma seleção de parâmetros que pudesse discriminar vozes normais, vozes com nódulo e vozes com Edema de Reinke, a análise, baseou-se na quantificação dos parâmetros de forma (E_0 , α , β , ω_g), onde wg foi o único discriminativo na diferença de vozes; nos parâmetros relacionados a extensão de banda ($B1$, $B2$, $B1/B2$ e $B2/B1$), que a análise estatística mostrou que não são confiáveis na discriminação entre as vozes; no parâmetro *ece*, que através da análise estatística o parâmetro *ece* pode ser uma opção na discriminação entre as vozes normais, nódulo e Reinke e por fim, nos parâmetros relacionados a fase de retorno do sub-sistema glotal, TR (TR_p e TR_s), onde o parâmetro TR_p , mostrou-se eficiente na discriminação da patologia Reinke, e o parâmetro TR_s , na discriminação entre voz normal e patológica avaliando desta forma a alteração da onda mucosa.

Portanto, baseando-se nos resultados resumidos acima, foram selecionados os parâmetros wg , *ece*, além de TR_p e TR_s .

Embora o banco de dados de voz utilizado contenha poucas amostras, os resultados estatísticos são significantes demonstrando o potencial de detecção/discriminação dos parâmetros propostos (TR_p e TR_s).

Um diagnóstico preciso dentro da área de saúde se torna complicado diante da imensa responsabilidade do profissional pelo indivíduo a ser tratado. Os resultados deste trabalho complementam o diagnóstico permitindo uma avaliação clínica mais fidedigna.

Um diagnóstico médico para determinadas patologias é muito difícil, com por exemplo, nódulo e cisto. Diante deste fato, sugere-se para trabalhos futuros encontrar os parâmetros referentes a diversas patologias e assim determinar uma possível discriminação entre elas.

Uma outra sugestão é a extração dos parâmetros em crianças entre sete e dez anos, buscando uma padronização antes da muda vocal.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ALKU, P. and Vilkman, E., "Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering", *Speech Communication* 18, (1996), 131-138.

ALKU, P., and Backstrom, T. "Normalized amplitude quotient for parametrization of the glottal flow", *J. Acoust. Soc. Am.* 112 (2), August, 2002.

ANANTHPADMANABHA, T. V. and Fant, G., "Calculation of true glottal flow and its components", *Speech Commun.*, pp. 167 –184, 1982.

BOX, G. E. P., Hunter, W. G., Hunter, J. S., "Statistics for Experiments", John & Sons Inc., 1978, p.165 – 193.

CHILDERS, D.G. and C. Ahn. "Modeling the glottal volume-velocity waveform for three voice types". *Journal of the Acoustical Society of America* 97(1), 505- 519, 1995.

CHILDERS, Donald G. "Speech processing and synthesis toolboxes", 1999.

CUMMINGS. K. E., "Analysis Synthesis and Recognition of Stressed Speech". *Ph. D. thesis*, Georgia Institute of Technology, Atlanta, GA, September, 1992.

CUMMINGS. K. E. and Clements, M. A., "Glottal Models for Digital Speech Processing: A Historical Survey and New Results", *Digital Processing* 5, (1995), 21-42.

DELLER Jr., P. R.; Proakis, J. G. and Hansen, J. H. L. "Discrete-Time Processing of Speech Signals". Macmillan Publishing Company, New York, NY, USA. 1993.

DRETZKE, B. J. and Heilman, K. A. "Statistics with Microsoft Excel", Prentice Hall, Upper Saddle River, NJ 07458, 1998.

EPSTEIN, M., B. Gabelman, N. Antoñanzas-Barroso, B. Gerratt, and J. Kreiman.
 "Source model adequacy for pathological voice synthesis". *Paper presented at the XIVth International Congress of Phonetic Sciences, San Francisco, 1999b.*

FANT, G., Liljencrants, J., and Lin, Q. 1985. A four-parameter model of glottal flow. *Paper presented at the French-Swedish Symposium, Grenoble, France.*

FANT, G. Liljencrants, J. and Lin, Q. "A Four-Parameter Model of Glottal Flow". *In Speech Transmiss. Lab. Q. Prog. Stat. Rep.*, pages 1–13, 1985.

FANT, G. "Glottal flow: models and interaction", *J. Phonet.*, 14, (1986), 393 – 399.

FUJISAKI, H. and Ljungqvist, M. 1986. Proposal and evaluation of models for the glottal source waveform. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 4*, 1605-1608.

FUJISAKI, H., and Ljungqvist, M. "Estimation of voice source and vocal tract parametrs basead on ARMA analysis and model for the glottal source waveform", *Proc. 1987. Int. Conf. Acoust. Speech Signal Process.* 1 (April 1987). 637-640.

GABELMAN, B. and Alwan, A. "Analysis by Synthesis of FM Modulation and Aspiration Noise Components in Pathological Voices." *In ICASSP Conference Proceedings*, pp. 449-452, Orlando, FL, May, 2002.

HEDELIN, P. "A glottal LPC vocoder". *Proc. 1984 Int Conf. Acoust. Speech Signal Process*, 1 (April 1984), 1.6.1-1.6.4.

- HOLMES, J.** "The influence of glottal waveform on the naturaness of speech from a parallel formant synthesizer". *IEEE Trans. Aud. Electroacoust.* AU21 3, (1973), 298-305.
- HOLMES, J.** "Formant Synthesizers: Cascade or Parallel?", *Speech Commun*, 2 (Dec, 1983), 251 – 273.
- ISHIZAKA, K. and Flanagan, J. ,** "Systhesis of voiced sounds from a two-mass model of the vocal cords", *Bell System Tech Journal*, 51, 1233-1268, 1972.
- ISSHIKI, N., Kitagima, K., Kojima, H. and Harita, Y.** "Turbulent noise in Dysphonia". *Folia Phoniatica*, 30, 214 – 224, 1978.
- ISSHIKI, N., Yanagihara, N. and Morimoto, M.** "Approach to the Objective Diagnosis of hoarseness." *Folia Phoniatica*, 18, 393 – 400, 1966.
- ISSHIKI, N., Okamura, H., Tanave, M, and Morimoto, M.** "Differential Diagnosis of Hoaseness". *Folia Phoniatica*, 21, 9 – 19, 1969.
- KENT R. D. and Read C.** "The Acoustics Analysis of Speech", *Singular Publishing Group, Inc.*, San Diego, CA.1992.
- KOIKE, Y. and Markel, J.,** "Application of inverse filtering for detecting laryngeal pathology." *Annals of Otolaryngology*, 84, 117 – 124, 1975.
- KLATT, D. H. and L. C. Klatt.** "Analysis, synthesis, and pereception of voice quality variations among female and male talkers". *Journal of the Acoustical Society of America* 87(2), 820-857, 1990.
- MATSUI, K., Pearson, S., Hata. K., and Kamai, T.** "Improving naturralness in text-to-speech synthesis using natural glottal source". *Proc., 1991 Int. Conf.*

Acoust. Speech Signal Process. 2 (May 1991), 769-772.

MONTAGNOLI, Arlindo Neto, “Análise residual do sinal de voz”. *Dissertação de mestrado – Escola de Engenharia de São Carlos - Universidade de São Paulo*, 1998.

NÍ CHASAIDE, A. and C. Gobl. Voice Source Variation”. In W. J. Hardcastle and J. Laver, eds., *The Handbook of Phonetic Sciences*. Blackwell Publishers Ltd.: Oxford. 427-461, 1997.”

PEREIRA, J. C. e Montagnoli, A. N., “Análise Acústica do sinal de Voz”, *Apostila*, Dezembro, 1999.

PLUMPE, M. D., Quatieri, T. F. And Reynolds, D. A., “Modeling of the glottal flow derivative waveform with application to speaker identification”, *IEEE transactions on Speech and Audio Processing*, 7 n° 5, (1999), 569-586.

QI, Y. and Bi, N., “A Simplified Approximation of the Four – Parameter LF Model of Voice Source”, *J. Acoust. Soc. Am.* 96 (2) Pt. 1, August, 1994.

RABINER, L. R. and Schafer, R. W., “Digital Processing of Speech Signals”. *Englewood Cliffs*, NJ: Prentice – Hall, 1978.

ROSA, M. O., Pereira, J. C. and Grellet, M., “Adaptative Estimation of Residue Signal for Voice Pathology Diagnosis”, *IEEE Transaction on Biomedical Engineering*, vol. 47, No 1, pp 96 – 104, January, 2000.

ROSENBERG, A.E. 1971. “Effect of glottal pulse shape on the quality of natural Vowels”. *Journal of the Acoustical Society of America* 49, 583-598.

TITZE, I., “Principles of voice production”. *Englewood Cliffs*: Prentice-Hall. (1994)