Campus de São Carlos

ANÁLISE RESIDUAL DO SINAL DE VOZ

Autor: Arlindo Neto Montagnoli

Orientador: Prof. Dr. José Carlos Pereira

UNIVERSIDADE DE SÃO PAULO

**ESCOLA DE ENGENHARIA
DE SÃO CARLOS**

ANÁLISE RESIDUAL DO SINAL DE VOZ

Arlindo Neto Montagnoli

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica.

ORIENTADOR: Prof. Dr. José Carlos Pereira

São Carlos
1998

DEDALUS - Acervo - EESC

31100017205

Class.	TESE-EEEC
Cutt.	3720 ✓
Tombo	0186/98

06/05/98.

st 0984326

**Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca - EESC-USP**

Montagnoli, Arlindo Neto
M758r Análise residual do sinal de voz / Arlindo Neto
Montagnoli. -- São Carlos, 1998.

Dissertação (Mestrado) -- Escola de Engenharia
de São Carlos-Universidade de São Paulo, 1998.
Área: Engenharia Elétrica.
Orientador: Prof. Dr. José Carlos Pereira.

1. Sinal de voz. 2. Filtro residual.
3. Patologias na laringe. I. Título.

FOLHA DE APROVAÇÃO

Candidato: Engenheiro **ARLINDO NETO MONTAGNOLI**

Dissertação defendida e aprovada em 06-05-1998
pela Comissão Julgadora:

Prof. Associado **JOSÉ CARLOS PEREIRA (Orientador)**
(Escola de Engenharia de São Carlos - Universidade de São Paulo)

Prof. Associado **EDSON DOS SANTOS MOREIRA**
(Instituto de Ciências Matemáticas de São Carlos - Universidade de São Paulo)

Prof. Doutor **EDUARDO TAVARES COSTA**
(Universidade Estadual de Campinas - UNICAMP)

Prof. Doutor **DENIS VINICIUS COURY**
Coordenador da Área de Engenharia Elétrica

JOSÉ CARLOS A. CINTRA
Presidente da Comissão de Pós-Graduação da EESC

Aos meus pais Arlindo e Pierina
e a minha esposa Marley
que tanto me apoiaram e
incentivaram na realização
deste trabalho.

AGRADECIMENTOS

Ao Orientador Prof. Dr. José Carlos Pereira que orientou o desenvolvimento deste trabalho com dedicação e incentivo.

As minhas irmãs que sempre acreditaram em meu trabalho e sempre me apoiaram.

Ao Prof. Dr. Marcos Grellet do Hospital Central em Ribeirão Preto - USP, pela coordenação na realização das gravações das vozes com disfonias.

Aos meus amigos que sempre me incentivaram na realização desta dissertação.

Ao CNPq pela bolsa de estudo concedida.

SUMÁRIO

LISTA DE FIGURAS	i
RESUMO	iii
ABSTRACT	iv
1 INTRODUÇÃO	01
1.1 Processo de Produção da Voz	02
1.1.1 Estruturas do Aparelho Fonador	02
1.1.2 Funcionamento do Aparelho Fonador	05
1.1.3 Características Acústicas da Voz	06
2 PROCESSAMENTO DIGITAL DA VOZ	08
2.2 Modelos Digitais para Sinais de Voz	10
2.2.1 Modelo da Excitação	10
2.2.2 Modelo da Glote	11
2.2.3 Modelo do Trato Vocal	12
2.2.4 Modelo da Radiação	13
2.2.5 Filtro Inverso da Voz: Solução das Equações de Modelamento	13
3 TÉCNICAS PARA AVALIAÇÃO DAS DISFONIAS	17
3.1 Avaliação do Comportamento Vocal	17
3.2 Técnicas Acústicas	18
3.2.1 Sinal Direto	18
3.2.2 Sinal Indireto	18
3.2.3 Parâmetros Acústicos no Domínio do Tempo	20
3.2.4 Parâmetros Acústicos no Domínio da Frequência	24
4 DESENVOLVIMENTO	27
4.1 Considerações Iniciais do Programa	27
4.2 Efeitos de Bordas (janelamento)	30
4.3 Programa para Análise Acústica da Voz	32
4.2 Levantamento de um Padrão Normal de Voz	36
5 RESULTADOS	47
5.1 Análises de Vozes Normais	47
5.2 Análises de Vozes Patológicas	55

6 CONCLUSÕES	62
7 TRABALHOS FUTUROS	64
8 REFERÊNCIAS BIBLIOGRÁFICAS	65

ANEXOS

- A1. Listagem das Rotinas do Software desenvolvido
- A2. Resultados das análises das vozes normais
- A3. Resultados das análises das vozes patológicas

Lista de Figuras

Figura 1.1-	Aparelho Fonador.....	3
Figura 1.2-	Imagem da laringe.....	4
Figura 1.3-	Diagrama esquemático do sistema vocal.....	6
Figura 2.1-	Modelo linear para produção da fala.....	10
Figura 4.1a	Gráficos do espectro do filtro conforme a ordem.....	28
Figura 4.1b	Gráficos do espectro do filtro conforme a ordem.....	29
Figura 4.2-	Transformada de Fourier da amostra de voz.....	30
Figura 4.3-	FFT da janela retangular.....	31
Figura 4.4a	Logaritmo da FFT da senoide (janela retangular).....	31
Figura 4.4b	FFT do sinal senoidal (janela Hamming).....	33
Figura 4.5-	Diagrama Esquemático do Programa	32
Figura 4.6-	Classe Matemática: Matem.cpp.....	33
Figura 4.7-	Classe p/ processamento de sinais: Tdsp.cpp.....	34
Figura 4.8-	Fluxograma para o cálculo do SFR, SRR e espectros.....	34
Figura 4.9-	Classe para processamento de sinaisII: TDspII.cpp.....	35
Figura 4.10-	Fluxograma para cálculo de frequências e bandas.....	35
Figura 4.11-	Classe p/ processamento de sinais III: TDspIII.cpp.....	36
Figura 4.12-	Fluxograma para cálculo do PPQ, APQ e Ex.....	36
Figura 4.13-	Histograma da frequência fundamental.....	37
Figura 4.14-	Histograma do coeficiente de excesso.....	38
Figura 4.15-	Histograma da amplitude de pitch.....	39
Figura 4.16-	Histograma do quociente de perturbação do pitch.....	40
Figura 4.17-	Histograma do quociente de perturbação do pitch eliminando valores acima de 1%.....	42
Figura 4.18-	Histograma do quociente de perturbação da amplitude.....	43
Figura 4.19-	Histograma do nivelamento espectral do resíduo.....	44
Figura 4.20-	Histograma do nivelamento espectral do filtro.....	45
Figura 5.1-	Amostra selecionada do sinal de voz.....	48
Figura 5.2-	Transformada de Fourier do sinal amostrado.....	48
Figura 5.3-	FFT da amostra selecionada com escala logarítmica.....	49
Figura 5.4-	Histograma da FFT da amostra selecionada com 20 barras.....	50
Figura 5.5-	Histograma com quatro barras da FFT da amostra selecionada.....	51
Figura 5.6-	Todos os resultados gráficos.....	52
Figura 5.7-	Espectro do filtro.....	52
Figura 5.8-	Sinal residual da amostra selecionada.....	53
Figura 5.9-	Auto correlação do resíduo.....	53
Figura 5.10-	Resultados numéricos.....	54
Figura 5.11-	Resultados numéricos em forma de barra.....	55
Figura 5.12-	Análise de uma voz patológica.....	56
Figura 5.13-	Resultados numéricos para voz patológica.....	57
Figura 5.14-	Resultados numéricos em forma de barra para voz patológica	57
Figura 5.15-	Resultados para uma voz patológica (leocoplasi).....	58

Figura 5.16- Resultados para um mesmo paciente (papilomatose).....	59
Figura 5.17- Resultados numéricos antes da cirurgia.....	59
Figura 5.18- Resultados numéricos um mês após cirurgia.....	60
Figura 5.19- Resultados numéricos um ano após cirurgia.....	60
Figura 5.20- Resultados gráficos para um mesmo paciente.....	61

Resumo

Este trabalho consiste no desenvolvimento de instrumentação destinada à pré-diagnósticos de disfunções na laringe, bem como no fornecimento de auxílio no acompanhamento dos tratamentos das patologias ou uma avaliação pós-operatória e ainda pode facilitar a tarefa de triagem de pacientes com problemas na laringe em hospitais.

O programa desenvolvido consiste em um método não invasivo para pré-diagnósticos da laringe. Através da análise do resíduo do sinal de voz obtém-se seis parâmetros que permitem indicar os casos de patologias.

Os sinais de vozes foram adquiridos diretamente com um microfone ligado a uma placa de aquisição no Departamento de Otorrinolaringologia da Faculdade de Medicina de Ribeirão Preto, USP. Este sistema oferece procedimentos clínicos mais confortáveis e permite um rápido diagnóstico sem restrições a qualquer paciente.

Abstract

This work describes the development of a device destined for pré-diagnoses of larynx dysfunctions, as well as, to supply aid in the accompaniment of the treatments of the pathologies or a postoperative evaluation and it can still facilitate the task of the patients' screen with problems in the larynx in hospitals.

A softaware package has been developed in order to set a non-invasive method for pré-diagnosis of the larynx. Through the residue voice signal analysis, six parameters are extracted from the signal in order to assess the pathological cases.

The voice signal was acquired directly from a microphone and an acquisition board at the Department of Otorrinolaringology of the Faculty of Medicine of Ribeirão Preto, USP. That system makes the clinical procedures more comfortable and allows for a faster diagnosis with no restriction to any patient.

1 INTRODUÇÃO

Atualmente, existe várias técnicas para se diagnosticar as patologias da laringe, sendo que as mais utilizadas são: análise perceptual da voz ou visualização da laringe através de um espelho ou laringoscópio.

O primeiro método é subjetivo, pois depende da avaliação do especialista e sua percepção em identificar os parâmetros característicos da voz patológica. Não raro, diferentes diagnósticos são emitidos por diferentes especialistas.

O segundo método é mais eficiente pois visualiza as estruturas do aparelho fonador, bem como os movimentos das cordas vocais durante a fonação, embora de uma forma limitada. As desvantagens deste método são: natureza invasiva, pequeno campo de visão face ao tamanho do aparelho vocal e incapacidade de verificar movimentos vibratórios de alta frequência e patologias em estágio inicial. Ainda, segundo BEHLAU & PONTES (1995), este método é ineficiente no diagnóstico de disfonias funcionais.

O trabalho proposto aqui consiste no desenvolvimento de um método de diagnóstico não invasivo usando parâmetros acústicos para a análise de sons vocálicos. Esta técnica permite, além do diagnóstico, a avaliação do comportamento vocal e o acompanhamento da evolução do tratamento do paciente.

Com base no modelo linear da fala, um filtro inverso glotal (filtro IIR – tudo-pólo) é utilizado para compensar as contribuições espectrais do trato vocal e dos lábios, fornecendo, desta maneira, uma estimativa da velocidade volumétrica glotal (pulso glotal). A saída do filtro é denominada sinal residual da voz, e é uma estimativa da função periódica gerada pela glote durante a fonação, DAVIS (1976).

A utilização do sinal residual é vantajosa em relação ao sinal direto da fala por realçar as componentes de alta frequência contidas na velocidade volumétrica glotal, possibilitando mais subsídios sobre o tempo de vibração das pregas vocais, ESKENAZI et. Al. (1990).

A classificação das disfonias é uma tarefa difícil requerendo profissionais especialmente treinados com habilidades perceptual para este tipo de avaliação. Várias desordens vocais são provenientes de disfunções na laringe. Alguns métodos podem auxiliar o treinamento profissional na classificação e modelamento das funções da laringe, por exemplo: o sinal da eletroglotografia (EGG), o sinal acústico, filmes de ultra alta-velocidade da laringe e síntese da fala, CHILDERS (1990).

Os parâmetros utilizados para a análise acústica do sinal de voz desenvolvidos neste trabalho são:

a) Domínio do Tempo

No domínio do tempo pode-se obter os seguintes parâmetros: a frequência fundamental (F_0), o quociente de perturbação da amplitude (APQ), o quociente de perturbação da “pitch” (PPQ), o coeficiente de excesso (EX) e a amplitude do “pitch” (PA).

b) Domínio da Frequência

No domínio da frequência tem-se o nivelamento espectral do filtro inversor (SFF) e nivelamento espectral do resíduo (SFR), frequências e bandas formantes.

Através do levantamento dos parâmetros acima foram realizados estudos estatísticos e elaboração de gráficos que auxiliam no diagnóstico das disfunções do aparelho fonador, assim como proporcionam facilidades para o acompanhamento de um tratamento fonoaudiológico.

1.1 Processos de Produção da Voz

1.1.1 Estruturas do aparelho fonador

A produção da voz humana requer uma unidade funcional obtida através da ação integrada e harmoniosa de uma série de estruturas que não participam de uma mesma unidade orgânica, por não existir um único órgão especialmente desenvolvido para a produção da fala, como é o caso da audição.

Os sons de nossa fala resultam da ação do sistema de órgãos que constituem o aparelho fonador e que criam as três condições necessárias para a produção da voz,

como a geração da corrente de ar pelos pulmões, um obstáculo a esta corrente de ar feita pelas cordas vocais, e uma caixa de ressonância feita pela cavidade nasal e bucal. O aparelho fonador envolve estruturas anatômicas que se iniciam nos pulmões e terminam nos lábios e nariz, por onde o som é irradiado conforme a figura 1.1:

FIGURA 1.1 - Aparelho Fonador

Fazem parte do aparelho fonador: os pulmões, os brônquios e a traquéia, que constituem os órgãos respiratórios e funcionam como dois foles gerando o fluxo de ar; a laringe, que possui nove cartilagens, situada na parte ântero-superior do pescoço, sob a base da língua, adiante da faringe e acima da traquéia. Nela se localiza a glote, pequena abertura em forma triangular na altura do “pomo-de-adão” situada entre as duas pregas musculares distendidas horizontalmente diante da glote (cordas vocais, figura 1.2) ; a epiglote, válvula que obstrui a glote durante a deglutição; a faringe que consiste em uma cavidade ligeiramente afinilada, entre a

boca e a parte superior do esôfago, que conduz o ar para a boca e fossas nasais; o palato mole (véu palatino) que quando levantado cola-se à parte posterior da laringe deixando livre apenas o conduto bucal e quando abaixado conecta as fossas nasais ao sistema; por último, encontra-se a boca que irradia os sons emitidos.

Cordas Vocais

FIGURA 1.2- Imagem da laringe

Pode-se dizer que os fonemas nascem na laringe e se completam na boca e isso acontece graças ao concurso das arcadas dentárias, dos alvéolos do palato duro (céu da boca) e do palato mole e, sobretudo, à atividade da língua, dos lábios e das bochechas, os quais se movimentam para modificar a corrente de ar e modular os fonemas. As fossas nasais, cavidades situadas no maxilar superior, funcionam como caixa de ressonância dos fonemas nasais. O centro do cérebro que intervém na faculdade de falar encontra-se no córtex cerebral. Qualquer lesão tanto nas extremidades das células do córtex ou das fibras que as unem interferem nos processos da fala.

Em uma voz alterada é importante avaliar as estruturas de fonação, para verificar se há alterações quanto a forma ou tonicidade que possam interferir na fala provocando ajustes compensatórios inadequados, levando a uma fonação sob esforço. Segundo BEHLAU & PONTES (1995) devem ser obtidos dados em repouso e durante a função de fonação e fala das seguintes estruturas: lábios, língua, dentes e oclusão dentária, palato duro, véu palatino, mandíbula e laringe.

1.1.2 Funcionamento do aparelho fonador

O ar expelido dos pulmões, por via dos brônquios, penetra na traquéia e chega à laringe, onde, ao atravessar a glote, encontra o primeiro obstáculo à sua passagem. O fluxo de ar pode encontrá-las fechadas ou abertas, em virtude das bordas das cordas vocais estarem aproximadas ou afastadas. No primeiro caso, o ar força a passagem através das cordas vocais esticadas, fazendo-as vibrar e produzir o som musical característico das articulações sonoras. No segundo caso, relaxando as cordas vocais, o ar escapa sem vibrações laríngeas, produzindo-se assim as articulações denominadas de surdas.

Ao sair da laringe, a corrente expiratória entra na cavidade faríngea, que lhe oferece duas vias de acesso ao exterior, o canal bucal e o nasal. Suspenso no entrecruzar desses dois canais fica o véu palatino, órgão dotado de mobilidade capaz de obstruir ou não a passagem de ar na cavidade nasal e, conseqüentemente, de determinar a natureza oral ou nasal de um som. É porém, na cavidade bucal que se produzem os movimentos fonadores variados, graças a maior ou menor separação dos maxilares, das bochechas e, sobretudo, à mobilidade da língua e dos lábios.

1.1.3 Características acústicas da voz

Na produção dos sons da fala o trato vocal age como uma série de filtros acústicos que atenuam ou reforçam o tom original e fundamental. Estes filtros ou ressonadores são capazes de assumir uma infinidade de formas diferentes com orifício de tamanho variável, conjugado com os movimentos dos lábios, língua,

palato e maxilares que produzem os sons característicos que distinguem uma vogal da outra.

No estudo do processo de produção da fala se torna útil conhecer importantes características físicas do sistema, de maneira a se obter um modelo matemático real e tratável. O modelo de produção da fala mais usual é o sistema linear variante no tempo e é baseado na premissa que o sinal de fala contém informações redundantes. O modelo inclui uma fonte de excitação que modela a velocidade volumétrica do ar que passa pela glote, um filtro linear variante no tempo que representa os ressonadores do trato vocal e uma impedância de radiação dos lábios, que é caracterizado como um circuito de diferenciação, CHILDERS (1990). Na figura 1.3 tem-se um diagrama esquemático de um sistema vocal.

Figura 1.3 Diagrama esquemático do sistema vocal

Para os sons sonoros existe um período de repetição chamado de período de tom ou “pitch”, sendo que seu inverso é chamado de frequência fundamental. Os sons não sonoros produzem um sinal semelhante a um ruído sem periodicidade definida. O espectro de frequências é definido de acordo com a fala e varia de instante a instante, porém, possui uma composição de frequências espaçadas e picos de uma forma bem definida, JOAQUIM (1990)

Os sons da fala podem ser classificados dentro de três classes distintas de acordo com o modo de excitação. Sons vocálicos são produzidos forçando o ar através da glote com as cordas vocais, com uma tensão ajustada para que vibrem em uma oscilação relaxada, produzindo pulsos quase periódicos do ar que excitam o

trato vocal. Sons fricativos ou não vocálicos são gerados fazendo uma contração em algum ponto do trato vocal, geralmente em direção ao fim da boca e forçando o ar através desta contração com uma velocidade suficiente para gerar uma turbulência. Os sons plosivos são resultados do fechamento completo do trato vocal, geralmente em direção frontal, aumentando-se a pressão no ponto de oclusão e liberando o ar subitamente, como na produção dos sons [p], [b], [t] e [d]. Para a classificação automática dos sons vocálicos e não vocálicos, existem várias técnicas desenvolvidas, como a que apresenta a determinação do período de pitch para decidir se um intervalo é ou não vocálico, KNORR (1979), KASUYA & WAKITA (1979), que auxiliam a determinação de algoritmos automáticos que selecionam os trechos vocálicos do sinal de voz para uma posterior análise.

Para os fonemas vocálicos, pode-se observar que apenas os três primeiros formantes possuem características semelhantes, e a partir do terceiro formante, frequências acima de 3200 Hz, não são mais relevantes para a produção dos mesmos, apenas caracterizam fatores pessoais, ROCHA (1988). Este dado é relevante na determinação do sistema utilizado, assim como auxilia a determinação das bandas, quantidade de formantes, frequências de corte e ordem dos filtros a serem utilizados em sistemas para análise de voz.

2 PROCESSAMENTO DIGITAL DA VOZ

O processamento de sinais consiste em obter a representação do sinal baseado em um dado modelo e a aplicação de alguma transformada para colocar o sinal em uma forma de visualização mais conveniente. A transformação envolvida no processamento de voz geralmente possui uma entrada $x(n)$ e uma saída $y(n) = T[x(n)]$ sendo utilizada uma simplificação, onde a classe estudada é linear e invariante ao deslocamento, que por sua vez, é completamente caracterizada pela sua resposta a uma entrada de pulso unitário. A saída pode ser calculada pelo sinal de entrada, $x(n)$ e a resposta ao impulso unitário, $h(n)$, usando a expressão da soma de convolução: $y(n) = \sum_k x(k) h(n-k) = x(n) * h(n)$.

A representação de uma transformada-z de uma sequência é definida pelo par de equações:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n} \quad (1)$$

$$x(n) = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz \quad (2)$$

A transformada-z de $x(n)$ é definida pela equação (1) e pode ser vista como uma série de potências infinitas na variável z^{-1} , onde a sequência de valores, $x(n)$, têm a função dos coeficientes na série de potências. Em geral cada série de potências converge para um único valor finito para certos valores de z . O conjunto de valores para que uma série convirja define a região no plano complexo- z conhecida como região de convergência.

Outra transformada muito utilizada em processamento de sinais de voz é a Transformada de Fourier na qual a representação de um sinal discretizado no tempo é dado pelas seguintes equações:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega.n} \quad (3)$$

$$x(n) = \frac{1}{2\pi} \cdot \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega.n} d\omega \quad (4)$$

Estas equações poder ser vistas como um caso particular da transformada-z, especificamente, a representação de Fourier é obtida pela restrição da transformada-z para o círculo unitário do plano-z.

Um filtro digital consiste em um sistema linear discreto no tempo e invariante ao deslocamento que têm a propriedade da entrada e saída satisfazerem a equação linear da diferença dada por:

$$y(n) - \sum_{k=1}^N a_k y(n-k) = \sum_{r=0}^M b_r x(n-r) \quad (5)$$

Aplicando-se a transformada-z em ambos os lados da equação pode-se mostrar que:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{r=0}^M b_r z^{-r}}{1 - \sum_{k=1}^N a_k z^{-k}} \quad (6)$$

A função sistema, $H(z)$, é em geral uma função racional de z^{-1} . A qual é caracterizada pela localização dos pólos e zeros no plano-z, RABINER & SCHAFER (1978).

Nos estudos de patologias da voz geralmente são utilizados os sons vocálicos devido a produção de vibrações das cordas vocais que por sua vez torna mais fácil a obtenção de característica da laringe, sendo que a vogal /a / é uma vogal oral, central, aberta, que por sua natureza articulatória evidencia as mínimas articulações no equilíbrio mioelástico da laringe, BEHLAU & PONTES. Um modelo linear para a produção da fala não anasalada foi proposto por FANT(1959) e subseqüentemente este modelo foi analisado em detalhes por FLANAGAN (1972).

figura 2.1 Modelo linear para produção da fala

Conforme a figura 2.1, este modelo é composto por três filtros acústicos. A principal justificativa para se utilizar este modelo é baseado na teoria de tubos acústicos, medidas da velocidade volumétrica, pressão sonora e resultados de circuitos elétricos para síntese de voz. A característica mais importante deste modelo é a separabilidade dos segmentos de filtros. A banda limite do sinal de voz $s_a(t)$ é dada em termos de um sinal $s(n)$ amostrado a cada T segundos:

$$s(n) = s_a(nT), -\infty < n < \infty \quad (7)$$

e com a transformada z $[S(z)]$ dada por:

$$S(z) = \sum_{n=-\infty}^{\infty} s(n)z^{-n} \quad (8)$$

2.2 Modelos Digitais para Sinais de Voz

2.2.1 Modelo da Excitação

A excitação $e(n)$ é modelada como uma série de impulsos unitários espaçados pelo período de pitch para sons vocálicos e é expressa como:

$$e(n) = E_0 \sum_{j=0}^{\infty} \delta(n - jP), n = 0, 1, \dots \quad (9)$$

onde E_0 é o fator de escala, $\delta(n)$ é a função delta. A transformada-z $[E(z)]$ e de $e(n)$ é dada por:

$$E(z) = \frac{E_0}{1 - z^{-P}} \quad (10)$$

onde P é o período de “pitch” e a frequência fundamental (F_0) é igual à $1/PT$.

2.2.2 Modelo da Glote

O filtro glotal pode ser representado por vários modelos tudo-pólo, FLANAGAN (1972), que são baseados em observações fisiológicas. Um dos modelos mais simples é o filtro em que a resposta ao impulso unitário é dado por:

$$g(n) = G_0 (n + 1) e^{-cnT}, \quad n=0, 1, 2, \dots \quad (11)$$

e a transformada-z $G(z)$ dada por:

$$G(z) = \frac{G_0}{(1 - e^{cT} z^{-1})^2} \quad (12)$$

onde G_0 é o fator escala e e^{-cT} é um pólo no círculo unitário do plano-z. Usando um trem de pulso $e(n)$ como entrada do filtro glotal $G(z)$, a velocidade volumétrica $u(n)$ é modelada como a saída do filtro e é representada por:

$$u(n) = E_0 \cdot G_0 \cdot \sum_{j=0}^{\infty} (n + 1 - jP) e^{-c(n-jP)}, n = 0, 1, \dots \quad (13)$$

2.2.3 Modelo do Trato Vocal

O modelo do trato vocal é usado para modelar características acústicas de ressonância do espaço contido entre a glote e os lábios. Este modelo é baseado em aproximações acústicas do trato vocal. Esta aproximação do tubo acústico é composta de um número específico de seções, onde cada seção tem uma área da seção transversal variável, SONDHI (1979). A estimação da área do trato vocal baseia-se na predição linear da voz onde as limitações que envolvem o método estão relacionadas com o limites das bandas e frequências do sinal, características da fonte, condições limites, energia perdida com o trato vocal e o método de avaliação empregado, WAKITA (1979). A transformada-z $V(z)$ do filtro do trato vocal pode ser modelada por um número finito K de pólos complexos da banda limite na qual a frequência central estende de zero à $\frac{1}{2} T$ Hz. A transformada-z é dada por:

$$V(z) = \prod_{i=1}^k \frac{1}{(1 - z^{-1}z_i)(1 - z^{-1}z_i^*)} \quad (14)$$

ou

$$V(z) = \prod_{i=1}^k \frac{1}{1 - e^{-\pi \cdot b_i T} \cos(2\pi \cdot f_i T) z^{-1} + e^{2\pi \cdot b_i T} z^{-2}} \quad (15)$$

Uma ressonância espectral da frequência central f_i e os dois lados da banda b_i são definidos por cada par de polo complexo (z_i, z_i^*) , onde $z_i = \exp[-\pi b_i + j2\pi f_i T]$, z_i^* é o conjugado complexo de z_i e T é o período de amostragem. Estas ressonâncias também são chamadas de formantes. Os formantes no modelo têm uma frequência central geralmente menor que 5000 Hz e a banda menor que 150 Hz.

2.2.4 Modelo da Radiação

O filtro da radiação dos lábios representa a transformação do sinal da velocidade volumétrica nos lábios para o sinal da pressão sonora. A função de transferência é bastante simplificada em relação a de FLANAGAN (1972), sendo que a transformada- z $L(z)$ é aproximada por um filtro transversal dado por:

$$L(z) = L_0(1 - \mu.z^{-1}) \quad (16)$$

onde L_0 é o fator de escala e μ é um pólo dentro de .95 e .98 no plano- z .

O modelo linear de produção da fala para uma pressão sonora acústica $S(z)$ é produto da transformada- z do filtro da radiação, do filtro do trato vocal, do filtro glotal e a fonte de excitação e é dado por

$$S(z) = L(z).V(z).G(z).E(z) \quad (17)$$

ou seja:

$$S(z) = \frac{L_0.G_0.E_0.(1 - \mu.z^{-1}).(1 - e^{cT}z^{-1})^{-2}(1 - z^{-P})^{-1}}{\prod_{i=1}^k 1 - e^{\pi.b_i T} \cos(2\pi.f_i T)z^{-1} + e^{-2.\pi.b_i T} z^{-2}} \quad (18)$$

2.2.5 Filtro Inverso da Voz: Soluções das Equações de Modelamento

O modelo da produção da fala dado pela equação (18) pode ser simplificado pelos seguintes processos: O fator do numerador $[1 - \mu z^{-1}]$ é utilizado para cancelar um dos fatores do numerador $[1 - e^{cT}z^{-1}]^2$, sendo que cT é geralmente muito menor que a unidade. O restante do numerador $[1 - e^{cT}z^{-1}]^{-1}$ é incluído como fator de produto do denominador. Na prática, alguns pólos podem ser adicionados arbitrariamente para permitir uma conformação do espectro do modelo da fala com o espectro real e para corrigir alguns erros de modelamento, ou seja, a voz não pode ser representada simplesmente como um modelo tudo-pólo baseada na fonte periódica impulsiva e três filtros independentes. MARKEL(1971) ilustra o uso de pólos adicionais para uma melhoria no espectro. Finalmente, a potência de z são arranjadas em ordem crescente e as constantes são combinadas para produzir a seguinte equação:

$$S(z) = \frac{\sigma}{\sum_{i=0}^M a_i \cdot z^{-i}} \quad (19)$$

onde a_i é real, $a_0 = 1$, $M = 2K + m$ sendo que $1 \leq m \leq 6$ e σ o ganho do filtro A transformada-z da excitação é obtida da equação (19) é dada como:

$$E(z) = \frac{1}{\sigma} \sum_{i=0}^M a_i z^{-i} S(z) \quad (20)$$

e a correspondente seqüência no tempo é dada por:

$$e(n) = \frac{1}{\sigma} \sum_{i=0}^M a_i s(n-i), \infty < n < \infty \quad (21)$$

fazendo-se

$$A(z) = \sum_{i=0}^M a_i z^{-i} \quad (22)$$

então

$$E(z) = \frac{1}{\sigma} A(z) \cdot S(z) \quad (23)$$

e

$$LM(E_T) = LM(S_T) - LM(1/A) - 10 \log \sigma^2 \quad (24)$$

onde o correspondente espectro logarítmico é obtido pela janela de Hamming da excitação e calculado por $LM(E_T) = 10 \log |E_T(z)|^2$ no círculo unitário $z = e^{j\theta}$, onde $E_T(z)$ é a transformada-z da sequência amostrada $e_T(n)$, sendo que, a frequência normalizada é dada por $\theta = 2\pi f/F_s$, a frequência de amostragem dada por $F_s = 1/T$ e f sendo a frequência contínua.

$A(z)$ é chamado de filtro inverso do espectro de voz porque $A(z)S(z)$ aproxima a excitação desejada $E(z)$ dentro dos limites do ganho constante σ ANANTHAPADMANABHA & YEGNANARAYANA (1979).

O filtro inverso $A(z)$ é encontrado determinando -se σ e a_i que minimizem a energia ξ em $LM(E_T)$ sobre toda frequência sujeitas à constante $a_0 = 1$:

$$\xi = \int_{-\pi}^{\pi} |E_T(e^{j\theta})|^2 \frac{d\theta}{2\pi} \quad (25)$$

Substituindo a equação (20) pela (25) com $z = e^{j\theta}$, expandindo e rearranjando os fatores temos:

$$\xi = \frac{1}{\sigma^2} \sum_{i=0}^M \sum_{k=0}^M a_i a_k \int_{-\pi}^{\pi} |S_T(e^{j\theta})|^2 e^{j\theta(i-k)} \frac{d\theta}{2\pi} \quad (26)$$

A integral da equação (26) é a transformada inversa do espectro de potência do sinal de voz e conseqüentemente é um termo da autocorrelação no tempo do sinal,

$$r(m) = \int_{-\pi}^{\pi} |S_T(e^{j\theta})|^2 e^{j\theta m} \frac{d\theta}{2\pi} = \sum s_T(n) s_T(n-m) \quad (27)$$

onde $s_T(n)$, $n = 0, 1, \dots, N-1$ é a amostragem truncada do sinal de voz, e os termos da autocorrelação são calculados assumindo $s_T(n)$ como um zero fora do intervalo $[0, N-1]$. Portanto,

$$\xi = \frac{1}{\sigma^2} \sum_{i=0}^M \sum_{k=0}^M a_i a_k r(i-k) \quad (28)$$

A solução é obtida pelo conjunto $\delta \xi / \delta a_i = 0, i = 1, 2, \dots, M$. Pode se encontrar os a_i pela solução de

$$\sum_{i=0}^M a_i r(i-k) = -r(k), k = 1, 2, \dots, M \quad (29)$$

e a solução é obtida iterativamente usando o método de Levinson, O ganho do filtro σ é muitas vezes é escolhido para igualar a energia total no espectro do sinal de voz ao da energia total do espectro do filtro inverso, MARKEL & GRAY (1976), ou σ pode ser igualado à unidade se a energia absoluta do sinal não for considerada, DAVIS (1976).

3 TÉCNICAS PARA AVALIAÇÃO DAS DISFONIAS

3.1 Avaliação do comportamento vocal

A faculdade de emitir sons articulados para expressar o pensamento nos torna capazes de transmitir nossas idéias e sentimentos de pessoa para pessoa, possibilitando a interação social. Quando uma voz não tem uma boa qualidade e não é capaz de cumprir este papel sem causar desconforto e dificuldade para o falante ou para o ouvinte, temos o que é considerado uma disfonia onde muitas vezes a causa é de origem orgânica e deve ser tratada sempre no início. Também as disfonias podem ser de origem funcionais ou orgânico-funcionais, porém, as disfonias se tornam evidentes e preocupantes para o paciente geralmente em estágio avançado, muitas vezes dificultando o tratamento.

Constantemente a disfonia é sintoma de diferentes distúrbios no aparelho fonador, as vezes como sintoma secundário inserido num quadro muito maior, como na doença de Parkinson, ou se caracteriza como sintoma principal sendo por isto encarada como a própria doença, como nas disfonias funcionais.

BELHAU & PONTES (1995) descrevem alguns procedimentos para se determinar vários parâmetros para a avaliação dos sons da fala, por exemplo, mede-se o tempo máximo de fonação, pede-se ao paciente que emita uma vogal /a / pelo maior tempo possível, cronometra-se este tempo que deve estar ao redor de 20s para homens e 14s para mulheres, posteriormente mede-se a relação S/Z, ou seja, mede-se os tempos de emissão das fricativas médias surdas e sonoras /S / e /Z / respectivamente e calcula-se a sua razão, a maneira como o som se inicia, que pode ser isocrônico, brusco e aspirado, onde a forma correta é a isocrônico, havendo um sincronismo entre a fase expiratória da respiração e o início da vibração da mucosa das pregas vocais, seguindo-se assim uma variada seqüência de procedimentos que o especialista pode verificar o comportamento da voz com o mínimo de instrumentação

necessária. Contudo, atualmente, a prova definitiva da existência de uma patologia de origem orgânica está em se fazer uma laringoscopia indireta ou direta, que mostra ao médico a laringe, onde o reconhecimento da patologia confirma o diagnóstico.

3.2 Técnicas Acústicas

3.2.1 Sinal Direto

Supondo que um sinal de voz seja gravado sobre condições ideais, todas as informações sobre as variações acústicas que o fluxo de ar sofre durante o seu percurso, desde o pulmão até os lábios são armazenadas neste sinal. A análise acústica da forma de onda direta do sinal de voz é a forma mais rápida, apenas depende de comparações visuais. Seu formato pode variar com as características de cada locutor, dificultando a verificação da existência de uma patologia. Outro fator importante a ser observado é que a maioria das disfunções ocorrem na laringe e a estrutura supra-glotal por sua vez se encontra sem anomalias, atuando de forma a amenizar as disfonias, portanto, pode esconder as características sonoras comuns de uma patologia na laringe, tornando mais difícil a análise direta. Um microfone conectado à garganta é às vezes utilizado para evitar efeitos da estrutura supra-glotal CHILDERS (1990), mas as informações são limitadas devido à ação dos tecidos como um filtro passa-baixas. Portanto, para a análise das disfunções da laringe, a análise direta serve como uma prévia visual, que com a experiência do analista torna-se possível fazer comparações de uma forma subjetiva, muitas vezes eficiente, devido a grande capacidade cerebral em se fazer comparações visuais.

3.2.2 Sinal Indireto

O trato vocal atua como uma série de filtros acústicos que atenuam ou reforçam o tom original e fundamental, para um som vocálico com a sustentação de uma vogal pode simplesmente ser modelado como uma forma de onda de pressão acústica, resultado de uma fonte quase periódica, que corresponde à glote, e por um

tubo acústico, que corresponde ao trato vocal e lábios. Como geralmente as patologias se encontram na laringe, a técnica de filtragem inversa aplicada a sinais de pressão sonora pode remover efeitos do tubo acústico e resultar um sinal aproximado da fonte periódica contendo informações suficientes para análise de efeitos acústicos de uma patologia na laringe, DAVIS (1976) . Medidas baseadas no sinal de voz filtradas são potencialmente mais informativas que as medidas baseadas em sinais sem filtragem.

Existem dois métodos para a filtragem inversa usada para se obter informações acústicas sobre a fonte glotal. A primeira consiste no filtro inverso glotal. Neste processo o inverso da radiação dos lábios e a contribuição do espectro do trato vocal são utilizados para estimar a velocidade volumétrica glotal como uma função do tempo, WONG et al. (1979)

Basicamente o estudo do filtro inverso glotal usa técnicas analógicas e usualmente elimina somente o primeiro e segundo harmônico. Posteriormente estuda a envoltória através de técnicas computacionais. As frequências e bandas formantes são estimadas e usadas para o ajuste do filtro inverso glotal até a obtenção da forma de onda esperada surgir sob um aspecto triangular seguido por uma porção em zero.

Contudo toda aparência do resultado da onda não satisfazem conceitos intuitivos das cordas vocais fechadas. Outra técnica de predição linear do sincronismo do pitch para o filtro inverso glotal produz aceitáveis formas de onda, sem estimar os formantes, mas o ponto de fechamento glotal para cada período de pitch deve ser localizado pela inspeção visual WONG et al.(1979).

O segundo método indireto para a obtenção das atividades da laringe consiste na filtragem inversa residual. Esta técnica é baseada na modelagem linear da fala. O filtro inverso residual consiste no inverso da radiação labial estimada, do trato vocal e das contribuições espectrais do modelamento glotal para o sinal de voz. O resultado da filtragem do sinal de voz por um filtro inverso residual consiste no sinal residual. Este sinal é uma estimação da função da fonte periódica para um modelo tudo-pólo da produção da fala e exibe picos reforçados para cada início do período de pitch e um ruído quase aleatório entre cada período de pico. KOIKE & MARKEL (1975) utilizaram o sinal residual em várias análises de vozes normais e patológicas, indicando que para alguns casos entre intermediário e avançado casos de patologias

na laringe o sinal residual não trazia informações convenientes sobre a disfunção vocal que estava presente. Contudo para alguns casos precoces, KOIKE & MARKEL (1975) observaram que o sinal residual traziam qualitativamente evidências da patologia o que não era obtido com o sinal de voz não filtrado. Subseqüentemente DAVIS (1976) verificou a hipótese de maiores informações acústicas serem transmitidas pelo sinal residual do que através do sinal de voz direto.

3.2.3 Parâmetros Acústicos no Domínio do Tempo

Medidas baseadas na perturbação da amplitude e período de pitch trazem informações capazes de diferenciar sinais de vozes patológicas das normais. Uma das primeiras destas medidas, o fator de perturbação do período de pitch LIEBERMAN¹ apud DAVIS (1976), é definido como a frequência relativa da perturbação do período de pitch maior que 0,5 milisegundos ocorrendo em sons vocálicos constantes. A perturbação do período de pitch é definida como a diferença entre a duração de sucessivos períodos de pitch do sinal de voz. Lieberman mostrou que a voz patológica geralmente tem um fator de perturbação maior que a voz normal quando comparadas com a frequência fundamental. A segunda medida de perturbação é relativa a média de perturbação KOIKE² apud DAVIS (1976). Este parâmetro é diferente do fator de perturbação do pitch de Lieberman em vários aspectos. Observou-se que sons constantes de vogais deve normalmente exibir trocas relativamente suaves no período de pitch. A velocidade das trocas ocorridas no período de pitch foram medidas e, adicionalmente, Koike sugeriu normalizar a medida da perturbação do pitch. Finalmente ele sugeriu que o sinal de contato com a garganta é melhor indicador da periodicidade da laringe que o sinal de voz, porque os efeitos da estrutura supra-glotal no sons da voz são reduzidos. Em consideração a estas observações, Koike definiu a média de perturbação relativa (RAP) como:

¹LIEBERMAN, P. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. Journal of the Acoustical Society of America, 1963 apud DAVIS, S.B. Acoustic Characteristics of Normal and Pathological Voices, Academic Press Inc, Santa Barbara, CA, 1976

$$\text{RAP} = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| \frac{P(i-1) + P(i) + P(i+1)}{3} - P(i) \right|}{\frac{1}{N} \sum_{i=1}^N P(i)} \quad (30)$$

onde $P(i)$, $i = 1, 2, \dots, N$ denota os sucessivos períodos de pitch. O numerador é a média da perturbação para cada período de pitch suavizado pela média de uma janela de três pontos e o denominador é a média de cada período, Koike concluiu que RAP para vozes normais e patológicas possuem diferentes valores e que varia significativamente entre pacientes com neoplasia e pacientes com paralisia unilateral.

A base para uma terceira medida da perturbação é a observação que a modulação da amplitude é quase periódica em sons vocálicos constantes para vozes patológicas. Koike computou os coeficientes de correlação serial para vários valores de amplitude da série no tempo e cada período de pitch descobrindo que estes coeficientes para vozes normais e patológicas são geralmente distinguíveis uns dos outros. Estes coeficientes para vozes com tumores na laringe usualmente mostraram picos de correlação significantes para atrasos entre três e doze períodos fundamentais e para vozes com paralisia na laringe, não mostrou a existência de picos. Koike concluiu que pode ser possível desenvolver métodos para diferenciar evoluções patológicas na laringe baseada em informações contidas em sons vocálicos constantes.

Uma quarta aproximação, KITAJIMA, et al.³ apud DAVIS (1976) definiu a medida de perturbação da frequência fundamental (F_0) como a média de 100 sucessivas perturbações de F_0 de uma frase toda vocálica. Uma escala de semi-tons (relativa a 16,35 Hz ou quatro oitavas entre o Dó médio) foi utilizada desde que a percepção auditiva de F_0 fosse aproximadamente proporcional ao logaritmo da frequência. Eles descobriram que as vozes femininas exibem uma medida de perturbação maior que as vozes masculinas, que pode ser explicada devido a

³KOIKE, Y. Application of some acousti measures for the evaluation of laryngeal dysfunction. *Studia Phonologica*, 1973. apud DAVIS, S.B. *Acoustic Characteristics of Normal and Patological Voices*, Academic Press Inc, Santa Barbara, CA, 1976

frequência fundamental feminina ser geralmente maior que a masculina. Também, perceberam que as vozes masculinas afetadas pelo câncer de laringe são distinguíveis das vozes femininas e masculinas normais através da medida da perturbação de F_0 .

Outros avanços feitos nas medidas de perturbação por TAKAHASHI & KOIKE⁴ apud DAVIS (1976) que combinou os resultados iniciais de Koike obtendo duas medidas no domínio do tempo de vozes patológicas baseadas em sinais captados por microfones em contato com a garganta: o quociente de perturbação de frequência (FPQ) e o quociente de perturbação em amplitude (APQ). O FPQ, por definição, é análogo ao RAP, sendo que os períodos de “pitch” foram substituídos por F_0 (definidos como uma reciprocidade do período de “pitch”). O APQ é, também por definição, análogo ao RAP, sendo que, os valores de amplitude dos picos para cada período de pitch e uma janela de doze pontos, que é preferível no lugar da de três pontos, para se fazer a média. Koike e Takahashi descobriram que o APQ contribui significativamente na análise da componente principal da qualidade vocal. Apesar da FPQ estar fortemente correlacionada com APQ, o FPQ não traz contribuições significantes na análise da componente principal.

DAVIS (1976) definiu o quociente do período de perturbação do pitch (PPQ) e o quociente de perturbação da amplitude (APQ) que são semelhantes ao RAP de Koike e ao APQ de Koike e Takahashi. Contudo havia várias diferenças nas definições acústicas, mais apropriadamente que utilizar uma janela de três ou doze pontos fixa, Davis, sistematicamente investigou os benefícios das alterações no tamanho das janelas, encontrando que a janela com cinco pontos para o cálculo da média produz resultados melhores nas medidas de perturbação APQ e PPQ na discriminação de vozes patológicas ou normais. Ele também verificou que as medidas de perturbação baseadas no sinal residual são melhores que as baseadas em sinal direto de voz, contudo ele não fez nenhuma comparação com sinais glotais ou com o microfone em contato com a garganta usados em outras pesquisas.

³KITAJIMA,K.TANABE,M.&ISSHIKI,N. Pitch perturbation in normal and patologic voice. *Studia Phonologica*,1975 apud DAVIS, S.B. *Acoustic Characteristics of Normal and Patological Voices*, Academic Press Inc, Santa Barbara, CA,1976

⁴TAKAHASHI,H.KOIKE,Y. Some perceptual dimensions and acoustical correlates of patologic Voices. *Acta Oto-Laryngologica, Supplementum*,1975,No.338,1-24 apud DAVIS, S.B. *Acoustic Characteristics of Normal and Patological Voices*, Academic Press Inc, Santa Barbara, CA,1976

Para o cálculo do APQ é escolhida dentre as amplitudes positivas e negativas o menor valor de APQ, e então, é calculada o PPQ da sequência do período de pitch correspondente ao menor APQ.

Davis desenvolveu, também, duas medidas acústicas no domínio do tempo para a verificação de patologias na laringe. Uma medida foi baseada nas observações que a relação sinal-ruído do sinal residual é uma boa alternativa para a discriminação de patologias. O sinal neste caso é a sequência de picos espaçados pelos intervalos períodos de pitch, e a energia do ruído quase randômico entre os picos do sinal. KOIKE & MARKEL (1975) sugeriram que uma medida da relação entre sinal-ruído pode ser feita através da energia de pico para cada período dividida pela energia do ruído na última metade do período para sucessivos períodos de “pitch” no intervalo especificado, contudo não tentaram nenhuma medida quantitativa.

Para o sinal residual de uma voz normal, a separação do sinal-ruído para cada período de pitch é bastante clara, contudo para um sinal residual de uma voz patológica os picos de pitch não são sempre distintos do ruído e podem ser dificilmente encontrados desta maneira. O aparecimento de um ruído maior em casos de patologias sugeriu a Davis que a distribuição de amplitudes de um sinal residual pode ser usada para medidas estatísticas da relação entre sinal e ruído, PROSEK et al (1987).

A forma de onda desta distribuição pode ser quantificada pela medida estatística chamada de coeficiente de excesso (EX), este coeficiente é definido como a razão do quarto momento da distribuição para o quadrado do segundo momento da distribuição. O EX é zero para uma distribuição Gaussiana, e é dado por :

$$EX = \frac{E\{(x - \bar{x})^4\}}{E\{(x - \bar{x})^2\}^2} - 3 \quad (31)$$

onde

$$E\{(x - \bar{x})^k\} = \frac{1}{N} \sum_{i=0}^{N-1} [x(i) - \bar{x}]^k \quad (32)$$

e

$$\bar{x} = \frac{1}{N} \sum_{i=0}^{N-1} x(i) \quad (33)$$

para o sinal $x(i)$, $i = 0, 1, \dots, N-1$. O coeficiente de excesso é positivo se a distribuição for mais estreita e mais alta que a Gaussiana e negativa se for mais baixa e mais larga que a Gaussiana.

Davis (1976a) definiu um parâmetro acústico no domínio do tempo chamado amplitude de pitch (PA), como o maior valor de pico de cada período de pitch na função de autocorrelação do sinal residual. A amplitude de pitch têm um alto valor para vogais e um valor pequeno para fricativas vocálicas e zero para fricativas não vocálicas.

3.2.4 Parâmetros acústicos no domínio da frequência

Outra técnica utilizada para análise acústica da voz é a análise no domínio da frequência, a qual fornecem um conjunto de informações, inclusive numéricas, o que a torna mais eficiente para uma análise computacional.

A espectrografia acústica é um dos instrumentos mais comuns para a análise em frequência de vozes. Filtrando-se o sinal de voz com um passa banda sintonizado temos a distribuição da energia espectral em um gráfico da frequência versus tempo, podendo observar com facilidade a frequência fundamental e seus formantes. Vários estudos foram feitos verificando-se as diferenças apresentadas no espectrograma entre vozes normais e patológicas. Harmônicos atenuados, componentes de ruído podem ser originados devido a turbulência do ar ao passar pela glote não completamente fechada ou por vibrações irregulares do trato vocal. Alguns laringologistas utilizam-se da espectrografia para ter acesso a informações da quantidade de ruído que se encontra na voz, pois geralmente a rouquidão é um dos primeiros sintomas das disfonias BEHLAU & PONTES (1995). Estes resultados apenas ilustram as vantagens e possibilidades do uso da espectrografia, porém não foram quantificados, tornando-se assim, subjetiva a análise no domínio da frequência, dependendo da experiência do analista.

Medidas da inclinação espectral de um sinal glotal foram as primeiras tentativas de se quantificar a espectrografia como forma de análise acústica, HIKI et al.⁵ apud DAVIS (1976), porém eles não apresentaram resultados quantitativos. Contudo, outras pesquisas usando tubos reflexivos indicavam que a constante de inclinação não era uma boa forma de se avaliar disfunções na laringe.

Usando um banco de filtros, FROKJAER-JESEN e PRYTZ⁶ apud DAVIS (1976) pesquisaram trocas na média espectral de um termo longo (LTAS) em pacientes em terapia para correção de desordem vocal devido a paralisia do nervo recorrente, o LTAS é determinado pela média do espectro obtida de um segmento de voz de 80 milisegundos de uma amostra de voz de 45 segundos. Eles sugeriram que a razão da energia entre a banda de 1 - 5 KHz e de 0 - 1 KHz é um bom parâmetro espectral para quantificar a qualidade vocal. Outros estudos foram feitos posteriormente aos de Frokjaer-Jensen e Prytz, mudando-se as bandas utilizadas, comprovando os bons resultados para este tipo de análise quantificando a qualidade vocal.

DAVIS (1976) utilizando conceitos de nivelamento espectral do filtro inverso residual (SFF) e o nivelamento espectral do sinal residual (SFR) GRAY & MARKEL, (1975). O nivelamento espectral é definido como a razão em decibéis da média geométrica pela média aritmética do espectro, sendo:

$$s1 = \sum_{i=1}^N \log(x[i]) \quad p/x[i] \neq 0 \quad (34)$$

e

$$s2 = \sum_{i=1}^N x[i] \quad (35)$$

para um sinal $x(i)$, $i = 1, 2, \dots, N$ onde N é o número de pontos da amostra temos como o nivelamento espectral do sinal:

⁵HIKI, SIMAIZUMI, S. HIRANO, M. MATSUSHITA, H. & KAKITA, Y. Acoustical analysis for voice disorders. Conference Record, IEEE International Conference on Acoustics, Speech and Signal Processing, 1976, pp.613-616 apud DAVIS, S.B. Acoustic Characteristics of Normal and Patological Voices, Academic Press Inc, Santa Barbara, CA, 1976

⁶ FROKJAER-JENSEN, B. PRYTZ, S. Registratio of voice quality. Bruel and Kjaer Technical Review, 1976, No 3, 3-17 apud DAVIS, S.B. Acoustic Characteristics of Normal and Patological Voices, Academic Press Inc, Santa Barbara, CA, 1976

$$\text{SFS} = 10 \log \frac{\exp(s1 / N)}{s2 / N} \quad (36)$$

e o nivelamento espectral do filtro como:

$$\text{SFF} = 10 \log \frac{\in [(M + 1)]}{\in [1]} \quad (37)$$

onde $\in [.]$ é igual aos termos do vetor energia para cada iteração, e M a ordem do filtro inverso. A partir destes valores pode ser calculado o nivelamento espectral do resíduo SFR como a diferença entre SFS e SFF.

GRAY e MARKEL (1975) observaram que quanto mais ruidoso é um espectro, maior é o valor do nivelamento espectral, sendo que, para um espectro constante o valor é de 0 dB. Pode-se observar também que para sons vocálicos o valor de SFF é menor para sons não vocálicos que são produzidos com a glote aberta e uma significativa constrição do trato vocal, este resultado é consequência da natureza harmônica do sinal vocálico e consequentemente para sinais patológicos onde há um acréscimo do ruído e uma diminuição dos harmônicos temos um acréscimo do nivelamento espectral em relação a vozes normais.

4 - DESENVOLVIMENTO

4.1 – Considerações Iniciais do Programa

Devido as facilidades oferecidas para programação estruturada e um excelente desempenho para programas que utilizam cálculos matemáticos foi escolhida a linguagem C++, NORTON & YAO (1992).

As variáveis do programa que podemos ressaltar como relevantes para a análise dos resultados são as seguintes:

a.) Número de pontos da amostra (nNpts)

Esta variável estabelece o número de pontos a serem considerados na amostra do sinal. Para uma análise acurada é necessário que a amostra contenha pelo menos 4 períodos do sinal. Por exemplo, utilizando-se uma frequência amostral de 22 KHz (fixada pela placa de aquisição – SB16) para um sinal cuja frequência é 90 Hz (voz masculina grave) teremos 244 pontos amostrados por período, portanto para se amostrar 4 períodos necessita-se de 977 pontos.

Para o uso eficiente do algoritmo FFT (Fast Fourier Transform), o número de pontos da amostra deve ser uma potência de 2. Portanto, no nosso exemplo, deve-se escolher uma janela com nNpts igual à 1024.

b) Frequência de amostragem (fa)

A referência de tempo no software é baseada na frequência de amostragem do sinal. Por ser utilizada a placa de aquisição SB-16 que fornece arquivos padrão com extensão .WAV, este parâmetro é extraído automaticamente do cabeçalho. Caso outro padrão seja utilizado, é necessário fornecer a frequência de amostragem para a correta definição da base de tempo.

c) Ordem do Filtro (nOrdem)

O filtro inverso utilizado na análise residual do sinal é um filtro casado cuja finalidade é anular as contribuições espectrais do trato vocal e da radiação dos

lábios. Para se estabelecer a ordem do filtro a ser utilizada, foram simulados diversos filtros com diferentes ordens. Os resultados dos filtros inversos simulados estão mostrados na figura 4.1.a e 4.1.b.

Figura 4.1-a Gráficos do espectro do filtro conforme a ordem

Figura 4.1-b Gráficos do espectro do filtro conforme a ordem

Buscando encontrar a ordem ótima do filtro inverso comparamos sua resposta em frequência com o espectro de um sinal de voz não patológica (figura 4.2)

Figura 4.2 Transformada de Fourier de um sinal de voz não patológica

É importante observar que a FFT do sinal real apresenta duas componentes, uma relativa à frequência fundamental (glotal) e outra relativa à estrutura supraglotal. Retirando-se a componente de baixa frequência superposta, suavizamos o espectro do sinal, o que corresponde a eliminarmos a contribuição da vibração glotal e realçarmos a contribuição da estrutura supraglotal. Por comparação, vemos que o filtro que mais se assemelha ao espectro da estrutura supraglotal apresenta ordem oito (8). Para ordens inferiores a oito, a influência da estrutura supra glotal deixa de ser eliminada e por outro lado, acima de dez, os formantes básicos da vogal são removidos.

4. 2 – Efeito de Bordas (janelamento)

Um efeito indesejável no processamento de sinal é aquele produzido pela segmentação do sinal. Por exemplo, uma janela retangular com 512 pontos possui um espectro característico mostrado na figura 4.3.

Figura 4.3 - FFT da janela retangular

Quando segmentamos o sinal no domínio do tempo por uma janela retangular, teremos como resultados a convolução do espectro da janela (figura 4.3) com o espectro do sinal. Portanto, o modo como se segmenta os dados é um fator importante na precisão dos resultados. Conforme estudos de MARKEL (1971), a utilização da janela de Hamming proporciona as melhores características para aplicações em sinais de voz. Para efeito de ilustração, apresentamos um sinal senoidal segmentado por uma janela retangular e uma janela de Hamming, ambas com 512 pontos (figura 4.4.a e 4.4.b, respectivamente).

Figura 4.4.a Logaritmo FFT da senoide (janela retangular)

Figura 4.4.b - FFT do sinal senoidal (Janela de Hamming)

4.3 - Programa para Análise Acústica da Voz

O programa principal, análise.cpp, gerencia as rotinas que compõem o software, realizando cálculos e apresentando resultados numéricos e gráficos. As rotinas tem como finalidade organizar todas as classes utilizadas nos cálculos e também iniciar as variáveis junto ao menu ferramentas. As listagens das rotinas encontram-se no apêndice A1. O diagrama em blocos do programa está mostrado na figura 4.5.

Figura 4.5 Diagrama esquemático do programa

A parte do programa que realiza os cálculos foi dividida em quatro classes hereditárias, ou seja, a classe posterior herda as funções das classes anteriores.

A primeira classe desenvolvida é a mais simples, onde se realiza todos os cálculos vetoriais envolvidos no programa, como rotacionar, multiplicar por constante, zerar vetores, cálculos das raízes de um polinômio, localizar picos, converter inteiro para real e interpolação parabólica.

A segunda classe, derivada da primeira, inicia o processamento digital de sinais, calculando a transformada rápida de Fourier, recursão de Levinson, e cálculo do resíduo através da execução de um filtro inverso. Posteriormente tem-se uma terceira classe, derivada da anterior, onde são calculados o período de tom "pitch" e os três primeiros formantes da vogal analisada, e, por fim, uma última classe onde se calcula o quociente de perturbação.

A primeira classe denominada matem.cpp esta estruturada por funções que executam os cálculos matemáticos mais simples conforme mostrado na figura 4.6.

Figura 4.6 Classe Matemática: Matem.cpp

A segunda classe é derivada da anterior e executa as funções básicas do processamento de sinais conforme mostra o diagrama de blocos da figura 4.7:

Figura 4.7 Classe para processamento de sinais: Tdsp.cpp

Com esta classe implementada já se torna possível o cálculo dos seguintes parâmetros: SFR, SFF, o espectro do sinal, o resíduo e seu espectro. O procedimento de cálculo destes parâmetros pode ser visualizado na figura 4.8:

Figura 4.8 Fluxograma para o cálculo do SFR , SFF e espectros

A terceira classe, derivada da anterior, acrescenta novos cálculos de processamento de sinais e seu diagrama de blocos é indicado a seguir, na figura 4.9:

Figura 4.9 Classe p/ processamento de sinais II: TDspII.cpp

A partir desta classe, pode-se calcular as frequências formantes e suas respectivas bandas, a frequência fundamental (pitch), e amplitude de pitch, conforme fluxograma da figura 4.10:

Figura 4.10 fluxograma par cálculo de frequências e bandas

Na última classe serão executados os cálculos de perturbação conforme mostrados na figura 4.11:

Figura 4.11 Classe p/ processamento de sinais III: TDspIII.cpp

Os últimos cálculos executados após a implementação desta classe são os parâmetros PPQ, APQ e EX, conforme a ilustração da figura 4.12:

Figura 4.12 Fluxograma para cálculo do PPQ, APQ e Ex

4.2 Levantamento dos Padrões Normais da Voz

Para o levantamento de um padrão normal de vozes foram selecionadas 25 pessoas que não tinham problemas aparentes com a voz e a laringe. Com dificuldade de se obter uma confirmação laringoscópica, por ser um exame demorado e incomodo, fez-se um estudo dos resultados e eliminou-se os casos discrepantes.

Através das 25 amostras de vozes gravadas, determina-se os valores máximos e mínimos dos parâmetros para estipular se uma voz é normal.

Através da função do menu *Resultados Numéricos*, do programa desenvolvido, fez-se a análise de cinco quadros para cada sinal de voz gravado resultando em 125 amostras e os resultados foram apresentados na forma de histograma, pois, muitos são os dados, impossibilitando uma apresentação de todos, porém, estes foram analisados individualmente e armazenados em arquivos textos, os quais são listados integralmente no apêndice A2.

Alguns cálculos estatísticos foram executados como cálculo da média, mediana, desvio padrão, valores máximos e mínimos, assim como, a curva gaussiana para cada histograma, exceto para a frequência fundamental, onde as amostras masculinas e femininas estão no mesmo histograma.

O primeiro parâmetro a ser analisado é a frequência fundamental e seu histograma é indicado na figura 4.13:

Figura 4.13 Histograma da frequência fundamental.

Pela análise do histograma da figura 4.13, pode-se verificar uma boa separação entre as vozes masculinas, frequências entre 90 a 150 Hz, com as vozes

femininas, frequências entre 70 a 210 Hz, comprovando os resultados de BEHLAU & PONTES (1995) que indicam que as vozes masculinas variam de 80 a 150 Hz e as femininas de 150 a 250 Hz. Devido aos estudos realizados com as diversas amostras, percebeu-se que este parâmetro têm pouco peso na indicação da existência da patologia e que uma variação de 30% dos limites estabelecidos acima a discriminação da patologia começa a ser evidenciada.

O segundo parâmetro a ser analisado é o coeficiente de excesso (EX) indicado na figura 4.14:

Figura 4.14 Histograma do coeficiente de excesso.

Estatística do coeficiente de excesso:

Curva Gaussiana:

Média: 7.7853 SD: 4.4831 Área: 231.246

Distribuição Normal:

Media: 10.0655 SD: 6.6957

Mediana: 8.047 Máx.: 38.6 Mín: 1.4218

Com estes resultados temos que 90% das vozes normais estão entre 0 e 16, porém, o coeficiente de excesso é um coeficiente inversamente proporcional a patologia, ou seja, quanto maior for este parâmetro, menor a probabilidade da ocorrência da patologia, portanto, com uma margem de segurança de 90% determina-se, através do gráfico da figura 4.14, este parâmetro com o mínimo valor de 2.0, sendo que, acima deste valor considera-se as vozes como normais e abaixo, como vozes patológicas.

O terceiro parâmetro analisado é a amplitude de pitch (PA), conforme mostrado na figura 4.15:

Figura 4.15 Histograma da amplitude de pitch.

Estatística da amplitude de pitch:

Curva Gaussiana

Média: 0.5195 SD: 0.1465 Área: 6.521

Distribuição Normal:

Média: 0.5210 SD: 0.1329

Mediana 0.503 Máx.: 0.7651 Mín: 0.227

Este resultado indica que 90% das vozes normais estão entre 0.3 e 0.72. Este parâmetro é inversamente proporcional a patologia, ou seja, quanto maior este, menor a probabilidade da ocorrência da patologia, portanto, com uma margem de segurança de 90%, determina-se este parâmetro, através do gráfico 4.15, o mínimo valor de 0.32, sendo que, acima deste valor considera-se como vozes normais e abaixo como vozes patológicas.

O parâmetro a seguir é o coeficiente de perturbação do pitch (PPQ), seu histograma e cálculos estatísticos estão na figura 4.16.

Figura 4.16 Histograma do quociente de perturbação do pitch

Estatística para perturbação do período de pitch:

Curva Gaussiana:

Média: -74.6001 SD: 98.25114 Área: 246.75887

Distribuição Normal:

Média: 0,891 SD: 1.8957

Mediana: 0.24 Máx: 12.30 Mín: 0.0

Este parâmetro está relacionado diretamente com a patologia, sendo que, baseado em simulações anteriores, percebeu-se-se que as vozes normais, na maioria dos casos, apresentam um limite superior de 1% de perturbação de “pitch”. Nota-se que a média está em 0.89, a mediana em 0.24 e o máximo valor é de 12.3. Se o máximo valor fosse utilizado como limite, provavelmente, muitos casos de patologias seriam desconsiderados. Através do gráfico da figura 4.19, utilizando uma margem segurança de 90%, teria que se considerar o limite de 2.0, contudo, o valor limite utilizado para o programa é de 1%, aconselhando-se realizar novas amostras se o valor for acima deste e abaixo de 2.0 e verificar os outros parâmetros para confirmar a existência de patologia ou não.

Se considerarmos os valores obtidos para vozes normais acima de 1% como valores discrepantes e eliminando-se estes valores temos os resultados mostrados na figura 4.17, que fornece 90% dos casos entre 0 e 0.5%, corroborando a escolha de 1% como limite.

Figura 4.17 Histograma do quociente de perturbação do pitch eliminando valores acima de 1%

Estatística do PPQ eliminando-se valores acima de 1%:

Curva Gaussiana:

Média: 0.19283 SD: 0.1885 Área: 9.8948

Distribuição Normal:

Média: 0.24041 SD: 0.18664 Máx: 0.87 Mín: 0.01

O quarto parâmetro é o coeficiente de perturbação da amplitude (APQ) e o histograma com as estatísticas estão na figura 4.18.

Figura 4.18 Histograma do quociente de perturbação da amplitude

Estatística para o quociente de perturbação de amplitude

Curva Gaussiana:

Média: 7.332 SD: 4.104 Área: 254,08

Distribuição Normal:

Média: 7.799 SD: 3.943

Mediana: 7.06 Máx: 22.74 Mín: 0.60

Este parâmetro, assim como o quociente de perturbação de pitch, está diretamente relacionado com a patologia e utilizou-se 90% das contagens, obtendo-se um valor máximo de 12.5 para se considerar a voz como normal, acima deste valor considera-se a possibilidade da voz ser patológica.

O quinto parâmetro é o nivelamento espectral do resíduo (SFR), indicado na figura 4.19.

Figura 4.19 Histograma do nivelamento espectral do resíduo.

Estatística do nivelamento espectral do resíduo:

Curva Gaussiana:

Média: -5.228 SD: 1.4029 Área: 62.394

Distribuição Normal:

Média: -5.2113 SD: 1.38

Mediana -5.088 Mín: -8.079 Máx: -1.871

Este parâmetro, também está diretamente relacionado com a patologia, portanto, para 90% das contagens, gráfico da figura 4.19, temos um valor de -3.5, que será estipulado como limite máximo para vozes normais.

O último parâmetro é o nivelamento espectral do filtro, indicado na figura 4.20

Figura 4.20 Histograma do nivelamento espectral do filtro.

Estatística para o nivelamento espectral do filtro:

Curva Gaussiana:

Média: -12.96076 SD: 2.50137 Área: 123.1502

Distribuição Normal:

Média: -13.10955 SD: 2.75139

Mediana: -12.954 Máx: -20.51 Mín: -8.45

Quanto maior for o parâmetro SFF, maiores são as possibilidades da existência de patologias, com uma margem de segurança de 90% e através do gráfico da figura 4.20, temos o valor máximo de -10.0, sendo que, abaixo deste valor o programa considerará a voz como normal.

5 Resultados

5.1 Análises de vozes normais

Os resultados das disfonias neste trabalho estão limitados às amostras obtidas no Hospital Central em Ribeirão Preto sob a coordenação do Prof. Dr. Marcos Grellet, e algumas amostras obtidas no Laboratório de Instrumentação e Microeletrônica do departamento de Engenharia Elétrica da USP - São Carlos. Também foram utilizadas as vozes da fita cassete incluída no livro de BEHLAU & PONTES(1995) apenas para obter alguns dados complementares, através de testes e comparações, pois os autores não garantem a qualidade de gravação, portanto os resultados obtidos não serão considerados como conclusivos.

Para fazer a análise dos resultados utiliza-se o programa desenvolvido Analise.exe, que segue os padrões "Windows". Inicialmente deve-se abrir um arquivo de voz utilizando-se o menu Arquivos no item Abrir, ou utilizar o primeiro botão de atalho: , posteriormente seleciona-se uma amostra de voz consistente, ou seja, que não esteja no início ou término da produção da vogal gravada, utilizando do sétimo ao décimo botão de atalho: , que consiste respectivamente em retrocesso rápido, retrocesso, avanço e avanço rápido, podendo utilizar o menu Ferramentas no item Ampliar ou último botão de atalho: , que executa uma ampliação no quadro amostrado, para obter maiores detalhes da amostra, tendo-se assim, resultados semelhantes a figura 5.1.

Posteriormente pode-se fazer as análises gráficas ou numéricas. Utilizando o menu Executa, selecionou-se o item FFT, que executa a transformada rápida de Fourier da amostra escolhida utilizando uma resolução de 1024 pontos, obtendo-se resultados semelhantes a figura 5.2.

Figura 5.1 Amostra selecionada do sinal de voz

Figura 5.2 Transformada de Fourier do sinal amostrado

Utilizando o menu Ferramentas no item Ampliar ou o último botão de atalho:

obtemos a transformada de Fourier com escala logarítmica, conforme a figura 5.3.

Figura 5.3- FFT da amostra selecionada com escala logarítmica

Outro recurso utilizado é o histograma da transformada de Fourier ou seja as frequências podem ser agrupadas escolhendo o número de barras através do menu Ferramentas no item Num de Barras e para obter o histograma seleciona-se o menu Ferramentas no item Histograma ou pode-se utilizar o botão de atalho: resultando um histograma semelhante ao da figura 5.4.

Figura 5.4 Histograma da FFT da amostra selecionada com 20 barras

Este recurso permite a análise das frequências em grupos, por exemplo, pode se observar os resultados dos exercícios que um paciente executou para melhorar o timbre de voz, ou seja reforçando as altas frequências. Neste caso pode se escolher o número de barras igual a quatro, obtendo o resultado conforme a figura 5.5.

Figura 5.5- Histograma com quatro barras da FFT da amostra selecionada

Outro recurso que o programa oferece está no item Automático do Menu Executa, ou utilizando a tecla de atalho: obtendo todos os resultados gráficos, figura 5.6, que o programa oferece e organiza todas as janelas obtidas no formato lado a lado horizontal.

Figura 5.6 Todos os resultados gráficos

Analisando a janela do espectro do filtro podemos observar a atuação do filtro eliminando as componentes que não pertençam aos três primeiros formantes, figura 5.7, da vogal analisada de uma voz normal, com isto, elimina-se as influências do trato vocal, obtendo-se assim, um sinal residual, figura 5.8, com picos bem definidos no início de cada período do sinal amostrado, semelhante a excitação do modelo estudado.

Figura 5.7 - Espectro do filtro.

Figura 5.8- Sinal Residual da amostra selecionada.

Outra característica que pode ser observada a uma voz normal é que a autocorrelação do resíduo também possui picos bem definidos que decrescem exponencialmente conforme pode ser observado na figura 5.9.

Figura 5.9 Auto Correlação do Resíduo

Para a obtenção dos resultados numéricos, pode-se utilizar o menu Executa no item Res. Numéricos ou através do botão de atalho: , que além de fornecer os seis parâmetros estudados fornece as frequências e bandas formantes do filtro, conforme a figura 5.10.

Figura 5.10 Resultados Numéricos

Pode-se observar que os resultados estão contidos nos limites estabelecidos para a voz normal, como pode ser visualizado na figura 5.11, onde estes resultados aparecem em forma de barra, as quais foram normalizadas com o limite estabelecido. Portanto, quanto menor a amplitude das barras, menor a possibilidade das vozes serem patológicas. Estas barras são de cor branca e podem ser comparadas com as barras cinza, que indicam os limites, se acaso os limites forem ultrapassados as barras mudam de cor para o cinza escuro.

Figura 5.11 Resultados numéricos em forma de barras

5.2 Análises de vozes patológicas

Seguindo os mesmos procedimentos executados anteriormente foi analisada uma voz patológica, figura 5.12, e todas as amostras analisadas estão listadas no apêndice A3.

Figura 5.12 Análise de uma voz patológica

Os resultados mais expressivos a esta análise consistem no fato do resíduo não apresentar os picos de sinal bem definidos (parte inferior esquerda da figura 5.12) no início de cada período, consequentemente os picos da autocorrelação são bem menores que para uma voz normal (parte superior esquerda da figura 5.12). Estes são os dois resultados conclusivos a respeito da análise gráfica, enquanto os três outros quadros, o espectro do filtro, a transformada de Fourier (FFT) e o sinal gravado, (lado direito, de cima para baixo) não são conclusivos para esta amostra.

Para a mesma amostra, calculamos os resultados numéricos que estão mostrados na figura 5.13, que sem os valores limites estipulados não permite qualquer inferência a respeito de patologias, porém, a análise numérica em forma de barras mostrado na figura 5.13, evidencia a existência da patologia, indicando que os parâmetros quociente de perturbação de pitch ($PPQ = 1.44$) e o nivelador de espectral do filtro ($SFF = -18.34$) ultrapassaram os limites estabelecidos máximos de 1.00 para o PPQ e de -10.0 para o SFF, os quais podem ser vistos comparativamente no gráfico de barras da figura 5.14.

Para um caso confirmado clinicamente (leocoplasia), os resultados da análise efetuada estão mostradas na figura 5.15. Para esta amostra verifica-se que quatro parâmetros acústicos excederam seus respectivos limites, nominalmente o coeficiente de excesso (Ex), amplitude de “pitch” (PA), quociente de perturbação do “pitch” (PPQ) e o nivelador espectral do filtro (SFF).

Figura 5.13- Resultados numéricos para voz patológica.

Figura 5.14- Resultados numéricos em forma de barras para voz patológica.

Figura 5.15 Resultados para uma voz patológica (leocoplasi)

Na figura 5.16, foi realizado um teste de acompanhamento clínico de um paciente do sexo feminino com 56 anos de idade portador de papilomatose da corda vocal direita.

A parte inferior do gráfico mostra o sinal de voz para diferentes épocas de aquisição. As análises estão mostradas na parte superior do gráfico. A primeira análise deu-se em agosto de 1994 antes da paciente se submeter a cirurgia, parte superior esquerda da figura 5.15. A Segunda aquisição e análise, parte central da figura 5.15, ocorreu em novembro de 1994 (60 dias após a cirurgia). Finalmente a parte direita da figura 5.15 mostra o sinal e sua análise realizada em novembro de 1995, após o tratamento fonoterápico.

Figura 5.16 Resultados para um mesmo paciente(papilomatose)

Figura 5.17 Resultados numéricos antes da cirurgia

Figura 5.18 Resultados numéricos um mês após cirurgia

Figura 5.19- Resultados numéricos um ano após cirurgia

Figura 5.20- Resultados gráficos para um mesmo paciente

Na figura 5.20 acima pode-se observar como a voz do paciente progrediu, analisando-se o gráficos da autocorrelação, onde os picos esperados para uma voz normal começam a aparecer conforme a regressão da disfonia, observa-se também que a frequência fundamental é muito alta para esta paciente, mesmo sendo uma voz feminina. Existe grande quantidade de energia no segundo, terceiro e quarto formantes antes da cirurgia, conforme pode ser visto nos gráficos da transformada de Fourier (FFT) que se encontram em forma de histogramas na figura 5.20 acima, isto se deve a uma oscilação irregular das cordas vocais que dificultam a fonação, observa-se, que a cirurgia e o tratamento fonoaudiológico surtem efeitos positivos onde a paciente começa a produzir vozes cada vez mais semelhante a uma voz normal, comprovando a eficiência do tratamento.

6. Conclusões

O trabalho desenvolvido utiliza a técnica de filtragem inversa do sinal de voz, removendo os efeitos da estrutura supra-glotal através da aplicação de um filtro inverso do trato vocal no sinal de voz real. A saída deste filtro (denominada resíduo) representa a excitação glotal e contém informações fisiológicas importantes das características da laringe.

Através do estabelecimento dos parâmetros amplitude do “pitch” (PA), quociente de perturbação do “pitch” (PPQ), quociente de perturbação da amplitude (APQ), nivelamento espectral do resíduo (SFR), nivelamento espectral do filtro (SFF) e coeficiente de excesso (EX), pode-se delinear as regiões características das vozes normais e patológicas. Baseado nas simulações de 30 pacientes com patologias na laringe e 25 pessoas normais é proposto o estabelecimento de limites para a determinação da existência de patologias.

Analisando os desvios padrões para as vozes normais, verifica-se que os parâmetros amplitude do “pitch” (PA) e quociente de perturbação do “pitch” (PPQ) são os menores e através dos estudos em vozes patológicas conclui-se que estes são os melhores indicadores para a verificação da existência da patologia. Os parâmetros nivelamento espectral do resíduo (SFR) e o nivelamento espectral do filtro (SFF), que estão relacionados com a quantidade de ruído do espectro do resíduo e filtro respectivamente são bons indicadores da existência de patologia ou não. O coeficiente de excesso (EX) é uma medida correlacionada com a relação sinal/ruído e juntamente com o quociente de perturbação da amplitude (APQ), que está relacionado com a variação da amplitude, demonstraram ser parâmetros pouco consistentes para verificação da patologia das amostras analisadas neste trabalho.

Adicionalmente, o programa desenvolvido representa uma ferramenta útil para a avaliação do quadro clínico pós operatório de uma forma quantitativa. Também, permite a visualização do sinal de voz, bem como a sua representação no domínio da frequência que podem auxiliar na terapia dos deficientes da fala, constituindo-se em uma ferramenta adequada ao fonoaudiólogo.

O programa elaborado tem como uma das suas características a interface compatível com o usuário das áreas de laringologia e fonoaudiologia agilizando os resultados bem como apresentando as telas concisas e completas utilizando os padrões “windows”.

7. Trabalhos Futuros

O estudo do sinal de voz possui diversas ramificações, onde recentemente a utilização de redes neurais vem sendo aplicada. Portanto a utilização destas, pode trazer resultados complementares a este trabalho.

Torna-se importante a obtenção de um número maior de amostras de vozes patológicas, obtendo-se grupos consistentes de cada patologia, o que facilitaria em trabalhos futuros a discriminação de cada patologia utilizando os parâmetros utilizados neste trabalho.

Pelo fato da perturbação do período de “pitch” é um dos melhores indicadores de patologia estudado neste trabalho, seria interessante estudar outros parâmetros que tivessem uma boa correlação com este, assim como a descoberta de outros parâmetros que viessem a complementar este estudo.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- ANANTHAPADMANABHA, T. V., YEGNANARAYANA, B. Epoch extraction from linear prediction residual for identification of closed glottis interval. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. ASSP-27, n.4, p.309-18, ago. 1979
- BEHLAU, M., PONTES, P. *Avaliação e tratamento das disfonias*. São Paulo, Lovise, 1995.
- CHILDERS, D. G. Speech processing and synthesis for assessing vocal disorders. *IEEE Engineering in Medicine and Biology Magazine*, p.69-71, mar. 1990.
- DAVIS, B.S. *Acoustic characteristics of normal and pathological voices*. Santa Barbara, Academic Press, 1976.
- ESKENAZI, L.; CHILDERS, D. G. Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research*, v.33, p.298-306, jun.1990.
- FLANAGAN, J. L. *Speech analysis synthesis and perception*. Expanded Edition Head Acoustic Research Department - Bell Laboratories, Murray Hill, New Jersey, 1972.
- HOLZNER, S. *Borland C++ programação for windows*. Trad. de Katia A. Roque. São Paulo, Makron Books, 1994.
- JOAQUIM, M. B. *Representação senoidal da voz*. São Paulo, 1990. Tese (Doutorado) - Escola Politécnica - Universidade de São Paulo.
- KASUYA, H., WAKITA, H. An approach to segmenting speech into vowel- and nonvowel-like intervals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. ASSP-27, n.4, p.319-27, ago. 1979
- KNORR, S. G. Reliable voiced/ unvoiced decision. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. ASSP-27, n.3, p.263-7, jun. 1979.
- KOIKE, Y., MARKEL, J. Application of inverse filtering for detecting laryngeal pathology. *Ann. Otol.*, v.84, p. 117-124, 1975.
- MARKEL, J. D., GRAY, A. H. *Linear prediction of speech*. New York, Springer Verlag, 1976.
- NORTON, P., YAO, P. *Programando em borland c ++ para windows*. Trad. de Katia de Almeida Guimarães. Rio de Janeiro, Berkeley Brasil, 1992.

- PROSEK, A.R., MONTGOMERY, B.E., & HAWKINS, D. B., (1987). **An evaluation of residue features as correlates of voice disorders.** *Journal of Communication Disorders*, 20, 105-117.
- RABINER, L.R., SCHAFER, R.W. **Digital processing of speech signals.** New Jersey, Prentice-Hall, 1978.
- ROCHA, L. **Procesamiento de voz.** Buenos Aires, Kapelusz, 1988.
- SONDHI, M. M. Estimation of vocal-tract areas: the need for acoustical measurements. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. ASSP-27, n.3, p.268-73, jun. 1979.
- WAKITA, H. Estimation of vocal-tract shapes from acoustical analysis of the speech wave: the state of the art. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. ASSP-27, n.3, p.281-5, jun. 1979.
- WONG, D. Y., MARKEL, J. D., GRAY, A. H. Least squares glottal inverse filtering from the acoustic speech waveform. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. ASSP-27, n.4, aug. 1979 .

A1. Listagem das Rotinas do Software Desenvolvido


```

#include "mat.h"
#include <conio.h>
#include <stdio.h>
TMat::TMat(int nN,int nO)
{
    nNpts=nN;
    nOrdem=nO;
    afAux=new float[nOrdem+5];
    afCoefR= new float[nOrdem+5];
}
TMat::~TMat()
{
    delete afAux;
}
void TMat::Inicia(int*X,int N,int CTE)
{
    int I;
    for(I=1;I<=N;I++)
        X[I]=CTE;
}
void TMat::Inicia(float*X,int N,float CTE)
{
    int I;
    for(I=1;I<=N;I++)
        X[I]=CTE;
}
void TMat::Reverte(int*X,int*Y,int N,int SCL)
{
    int I,NP1,nAux;

    NP1=N+1;
    for(I=1;I<=NP1/2;I++)
    {
        nAux=X[I];
        Y[I]=X[NP1-I]*SCL;
        Y[NP1-I]=nAux*SCL;
    }
}
void TMat::Reverte(float*X,float*Y,int N,float SCL)
{
    int I,NP1,nAux;

    NP1=N+1;
    for(I=1;I<=NP1/2;I++)
    {
        nAux=X[I];
        Y[I]=X[NP1-I]*SCL;
        Y[NP1-I]=nAux*SCL;
    }
}
void TMat::IntReal(int*X,float*Y,int N,float SCL)
{
    int I;

```

```

for(I=1;I<=N;I++)
Y[I]=(float)X[I]*SCL;
}

```

```

void TMat::MultCte(int*X,int*Y,int N,int SCL)

```

```

{
int I;
for(I=1;I<=N;I++)
Y[I]=X[I]*SCL;
}

```

```

void TMat::MultCte(float*X,float*Y,int N,float SCL)

```

```

{
int I;
for(I=1;I<=N;I++)
Y[I]=X[I]*SCL;
}

```

```

float TMat::InterParab(float YM,float Y0,float YP,float *XMAX)

```

```

{
float A,B,YMAX;

YMAX=0.0;
*XMAX=0.0;
A=(YP+YM)/2.0-Y0;
B=(YP-YM)/2.0;
if (A!= 0)
{
*XMAX=-B/(2.0*A);
YMAX=Y0+(B/2.0)*(*XMAX);
}
return(YMAX);
}

```

```

void TMat::RaizPol(float* afCoef,float* afRaizesReais,float* afRaizesImag)

```

```

{
int ifit,n,nx,nxx,n2,kj1,l,mt,in,ict,fi,ityp,i;//IER:

```

```

float xo,yo,temp,x,y,ypr,xpr,sumsq,dx,dy,
alpha,ux,yt,xt,xt2,yt2,uy,v,u;

```

```

ifit=0;
n=nOrdem;
// IER=0;
/* if(nOrdem<1){IER=1;exit_e(1);}
if(nOrdem>36){IER=2;exit_e(2);}
if(afCoefR[n+1]==0){IER=4;exit_e(4);}*/
Reverte(afCoef,afCoefR,n+1,1,0);
nx=n;
nxx=n+1;
n2=1;
kj1=n+1;

```

```

for(l=1;l<=kjl;l++)
{
    mt=kjl-l+1;
    afAux[mt]=afCoefR[l];
}
L45:
xo=-0.09;
yo=-0.03;
in=0;
L50:
x=xo;
xo=-10.0*yo;
yo=-10.0*x;
x=xo;
y=yo;
in=in+1;
goto L59;
L55:
ifit = 1;
xpr=x;
ypr=y;
L59:ict=0;
L60:ux=uy=v=yt=0.0;
    xt=1;
    u=afAux[n+1];
    if(u==0) goto L130;
    for (i=1;i<=n;i++)
    {
        l=n-i+1;
        temp=afAux[l];
        xt2=x*xt-y*yt;
        yt2=x*yt+y*xt;
        u=u+temp*xt2;
        v=v+temp*yt2;
        fi=i;
        ux=ux+fi*xt*temp;
        uy=uy-fi*yt*temp;
        xt=xt2;
        yt=yt2;
    }
    sumsq=ux*ux+uy*uy;
    if(sumsq==0) goto L110;
    dx=(v*uy-u*ux)/sumsq;
    x=x+dx;
    dy=-(u*uy+v*ux)/sumsq;
    y=y+dy;
    if(fabs(dy)+fabs(dx)-.00001 <0) goto L100;
    ict=ict+1;
    if(ict-500<0) goto L60;
    if(ifit !=0) goto L100;
    if (in-5 <0) goto L50;
L95://IER=3;
    //exit_c(3);
L100:for(l=1;l<=nxx;l++)

```



```

        {
            mt=kj1-l+1;
            temp=afCoeR[mt];
            afCoeR[mt]=afAux[l];
            afAux[l]=temp;
        }
        itemp=n;
        n=nx;
        nx=itemp;
        if(ifit==0) goto L55;
        else goto L120;
L110:
        if (ifit==0) goto L50;
        x=xpr;
        y=ypr;
L120:
        ifit=0;
        if(fabs(y/x)-0.0001<0) goto L135;
        alpha=x+x;
        sumsq=pow(x,2)+pow(y,2);
        n=n-2;
        goto L140;
L130:x=0.0;
        nx=nx-1;
        nxx=nxx-1;
L135:y=0.0;
        sumsq=0.0;
        alpha=x;
        n=n-1;
L140:afAux[2]=afAux[2]+alpha*afAux[1];
        for(l=2;l<=n;l++)
            afAux[l+1]=afAux[l+1]+alpha*afAux[l]-sumsq*afAux[l-1];
L155:afRaizesImag[n2]=y;
        afRaizesReais[n2]=x;
        n2=n2+1;
        if(sumsq==0.0) goto L165;
        y=-y;
        sumsq=0.0;
        goto L155;
L165:if (n>0) goto L45;

    }

void TMat::Momento(float*X,int nNpts,int M,float*XMOM)
{
    float RPTS,S,T;
    int I,J;

    RPTS=1.0/nNpts;
    MultCte(XMOM,XMOM,M,0.0);
    for(I=1;I<=nNpts;I++)
    {
        S=1.0;
        T=X[I];

```

```

        for(J=1;J<=M;J++)
        {
            S=S*T;
            XMOM[J]=XMOM[J]+S;
        }
    }
    MultCte(XMOM.XMOM.M.RPTS);
}

int TMat::MaxAmpPos(float*X,int JMI,int JMA)
{
    int J,JLOC;
    float XMAXP;

    XMAXP=-1*pow10(35);
    JLOC=JMI;
    for(J=JMI;J<=JMA;J++)
    {
        if(X[J]>XMAXP)
        {
            XMAXP=X[J];
            JLOC=J;
        }
    }
    return(JLOC);
}

```

```
#include "dsp.h"
```

```
TDsp::TDsp(int nN,int nOrd,BOOL bWin): TMat (nN,nOrd)
```

```
{
    nNpts=nN;
    nOrdem=nOrd;
    bJanela=bWin;
    afImag=new float[1024+nOrdem+15];
    afReal=new float[1024+nOrdem+15];
    // afModulo=new float[nNpts+nOrdem+5];
    // afResiduo=new float[nNpts+nOrdem+5];
    afAlpha=new float[nOrdem+5];
    afRc=new float[nOrdem+5];
    afCoef=new float[nOrdem+5];
    afSAux=new float[1024+nOrdem+15];
```

```

}
TDsp::~~TDsp()
```

```
{
    delete afImag;
    delete afReal;
    // delete afModulo;
    // delete afResiduo;
    delete afAlpha;
    delete afRc;
    delete afCoef;
    delete afSAux;
}
```

```
void TDsp::Fft(float* afSinal,float* afModulo)
```

```
{
    //afModulo=afMod;
    int aux.cont,AUX5,AUX2.IT,AUX3.cont2,AUX6,AUX7.NI;
    int bit_rev.nNPA;
    float AUX4,cont1,
            AUX1,WR,WI,TR,TI;
    Inicia(afReal,1024+nOrdem,0.0);
    MultCte(afSinal,afReal,nNpts,1.0);
    Inicia(afImag,1024+nOrdem,0.0);
    nNPA=nNpts;
    nNpts=1024;
    NI=log10(nNpts)/log10(2.0);
    AUX1=2*PI/nNpts;
    AUX2=nNpts;
    for(IT=1;IT<=NI;IT++)
    {
        AUX3=AUX2;
        AUX2=AUX2/2;
        AUX4=0;
        for (cont1=1;cont1<=AUX2;cont1++)
        {
```



```

WR=cos(AUX4);
WI=sint(AUX4);
AUX4=AUX4+AUX1;
AUX5=1;

for(cont2=AUX3;cont2<=nNpts;cont2=AUX5*AUX3)
{
    AUX5++;

    AUX6=cont1+cont2-AUX3;
    AUX7=AUX6+AUX2;
    TR=(afReal+AUX6)-(afReal+AUX7);
    TI=(afImag+AUX6)-(afImag+AUX7);
    *(afReal+AUX6)=*(afReal+AUX6)+*(afReal+AUX7);
    *(afImag+AUX6)=*(afImag+AUX6)+*(afImag+AUX7);
    *(afReal+AUX7)=TR*WR-TI*WI;
    *(afImag+AUX7)=TI*WR+TR*WI;
}
}

AUX1=2*AUX1;
}

for(cont=0;cont<nNpts;cont++)
{
    bit_rev=0;
    aux=cont;
    for(cont2=0;cont2<NI;cont2++)
    {
        bit_rev=(bit_rev<<1) + (aux&1);
        aux>>=1;
    }

    if (bit_rev >cont )
    {
        TR=afReal[cont+1];
        TI=afImag[cont+1];
        afReal[cont+1]=afReal[bit_rev+1];
        afImag[cont+1]=afImag[bit_rev+1];
        afReal[bit_rev+1]=TR;
        afImag[bit_rev+1]=TI;
    }
}

for(cont=1;cont<=nNPA;cont++)
afModulo[cont]=sqrt(afReal[cont]*afReal[cont]+afImag[cont]*afImag[cont])/nNpts;
nNpts=nNPA;
}

void TDsp::Residuo(float* afSinal,float *afResiduo)
{
    // float afModulo[1200];
    // afSinal=afSin;
    // afResiduo=afWork;

```

```

MultCte(afSinal.afSAux.nNpts+nOrdem+10,1.0);
for(int cont=1;cont<=nNpts;cont++)
    *(afSAux+cont)=*(afSAux+cont+nOrdem+2);
/*int k=1;
for(int l=0;l<nNpts*2-5;l=l+5)
    for (int j=1;j<=5;j=j+2)
        afSAux[k++]=afSAux[l+j];
//for(;k<nNpts*2+nOrdem+5;k++) afSAux[k]=0.0;
*/

if(bJanela) Janela();
Levinson();
for(int i=2;i<=nOrdem+1;i++)
    R[nOrdem+3-i]=afSAux[i];

Filtro(afResiduo);
// delete afSAux:

}
void TDsp::Flatness(float* afSinal,float* afModulo)
{
    MultCte(afSinal.afSAux.nNpts+nOrdem+10,1.0);
    for(int cont=1;cont<=nNpts;cont++)
        *(afSAux+cont)=*(afSAux+cont+nOrdem+2);
    Fft(afSAux,afModulo);
    float SUM1=0;
    float SUM2=0;
    float T;
    for (int J=1;J<=nNpts;J++)
    {
        T=pow(afModulo[J],2);
        if (T > 0) SUM1+=log(T);
        SUM2+=T;
    }
    float fSfd=10*log10(exp(SUM1/nNpts)/(SUM2/nNpts));
    fSff=10*log10(afAlpha[nOrdem]/afAlpha[1]);
    fSfr=fSfd-fSff;
    //delete afSAux:
}

void TDsp::Janela()
{
    for(int cont=1;cont<=nNpts;cont++)

        *(afSAux+cont)=*(afSAux+cont)*(0.54 - 0.46*cos((cont-1)*2*PI/nNpts));
}

void TDsp::Levinson()
{
    float S.AT;
    int nCont,nCont2,nMin,nMH,nInd;

```

```

R=new float[4*nOrdem+10];
for (nCont=1;nCont<=4*nOrdem+1;nCont++)
{
    R[nCont]=0.0;
    for(nCont2=1;nCont2<=nNpts-nCont+1;nCont2++)
        R[nCont]=R[nCont]+afSAux[nCont2]*afSAux[nCont2+nCont-1];
}

afCoef[1]=1.0;
afAlpha[1]=R[1];
if (nOrdem == 0) return;
afRc[1]=-R[2]/R[1];
afCoef[2]=afRc[1];
afAlpha[2]=R[1]+R[2]*afRc[1];
if (nOrdem == 1) return;
for(nMin=2;nMin<=nOrdem;nMin++)
{
    S=0.0;
    for (nCont=1;nCont<=nMin;nCont++)
        S=S+R[nMin-nCont+2]*afCoef[nCont];
    afRc[nMin]=-S/afAlpha[nMin];
    nMH=nMin/2 + 1;
    for(nCont=2;nCont<=nMH;nCont++)
    {
        nInd=nMin-nCont+2;
        AT=afCoef[nCont]+afRc[nMin]*afCoef[nInd];
        afCoef[nInd]=afCoef[nInd]+afRc[nMin]*afCoef[nCont];
        afCoef[nCont]=AT;
    }
    afCoef[nMin+1]=afRc[nMin];
    afAlpha[nMin+1]=afAlpha[nMin]-afAlpha[nMin]*afRc[nMin]*afRc[nMin];
    if(afAlpha[nMin]<=0) break;
}
//    nOrdem=nMin-1;
}

```

```

void TDsp::Filtro(float *afResiduo)
{
    int M=0;
    int nMax,nCont,nCont2;
    float fAux,*afAux,YY;
    afAux=new float[nOrdem+2];
    afAux[1]=1.0;
    nMax=nOrdem;
    // if(M>nOrdem) nMax=M;
    for(nCont2=1;nCont2<=nNpts;nCont2++)
    {
        fAux=afSAux[nCont2];
        if(M!=0)
        {
            for (nCont=2;nCont<=M+1;nCont++)
                fAux=fAux-R[nCont]*afAux[nCont];
        }
    }
}

```



```
R[1]=fAux:
YY=fAux*afCof[1]:
if(nOrdem!=0)
{
    for(nCont=2:nCont<=nOrdem+1:nCont++)
        YY=YY+R[nCont]*afCof[nCont]:
}
afResiduo[nCont2]=YY:
for(nCont=nMax+1:nCont>=2:nCont--)
    R[nCont]=R[nCont-1]:
```

```
}
}
```

```

#include "dsp2.h"

#include <conio.h>
#include <stdio.h>
TDsp2::TDsp2(int F,int N,int M,BOOL J): TDsp(N,M,J)
{
    fSf=F;
    //X=afResiduo;
    // nOrdem=M;
    // nNpts=N;
    // afCoef=Coef;
    // afAutoCorrelacao = new float[nNpts+nOrdem+5];
    afFt=new float[nOrdem+5];
    afBt=new float[nOrdem+5];
    afRaizr=new float[nOrdem+5];
    afRaizi=new float[nOrdem+5];
    afFf=new float[4];
    afBf=new float[4];

}
TDsp2::~~TDsp2()
{
    // delete afAutoCorrelacao;
    delete afFt;
    delete afBt;
    delete afRaizr;
    delete afRaizi;
    delete afFf;
    delete afBf;
}

void TDsp2::Reson(float*afResiduo,float*afAutoCorrelacao)
{
    //afAutoCorrelacao=afWork;
    //X=sinal;
    nNpeak=3;
    fBand=0.0;
    fQ=1.0;
    Formantes();

    if(nNpeak<=2) afFt[3]=afBt[3]=0.0;
    if(nNpeak<=1) afFt[2]=afBt[2]=0.0;
    if(nNpeak==0) afFt[1]=afBt[1]=0.0;
    MultCte(afFt,afFf,3,1.0);
    MultCte(afBt,afBf,3,1.0);
    Pitch(afResiduo,afAutoCorrelacao);

}

void TDsp2::Formantes()
{
    int I,L,K,JJ;
    float RTR,RTI,RAD,ANGLE,QT;//w,z:

```

```

NP=0;
if(nOrdem<=0) return;
//afCoef[1]=1.0;afCoef[2]=3.0;afCoef[3]=5.0;afCoef[4]=7.0;
// for(int i=5;i<10;i++) afCoef[i]=0;
RaizPol(afCoef.afRaizr.afRaizi);
/* if(IER!=0)
{
NP=-1*(IER);
return;
}*/

```

```

for(I=1;I<=nOrdem;I++)
{
RTR=afRaizr[I];
RTI=afRaizi[I];
afFt[I]=0.0;
afBt[I]=0.0;
if((RTI==0)&&(RTR==0)) break;
RAD=sqrt(RTR*RTR+RTI*RTI);
if(RTI<0) RTI=-RTI;
ANGLE=PI/2.0;
if(RTR!=0) ANGLE=atan(RTI/RTR);
if(RTR<0) ANGLE=PI+ANGLE;
afFt[I]=fSf*ANGLE*0.5/PI;
afBt[I]=-fSf*log(RAD)/PI;
}

```

Ordena(nOrdem.afBt.afFt.afRaizr.afRaizi);

```

if(NP==0) return;
if(NP<nNpeak) nNpeak=NP;
if(nNpeak>0) goto L100;
nNpeak=0;
for(JJ=1;JJ<=NP;JJ++)
if(afBt[JJ]>fBand) goto L80;
nNpeak=NP;
goto L90;
L80:
nNpeak = JJ-1;
L90:
if(nNpeak>0) goto L100;
NP=0;
return;
L100:
Ordena(nNpeak.afFt.afBt.afRaizr.afRaizi);
if((NP<=0)||fQ<=0) return;
K=1;
L110:
if(afBt[K]<=0) goto L120;
QT=afFt[K]/afBt[K];
if(QT<=fQ) goto L120;
if(K==NP) return;

```



```

K=K+1;
goto L110;
L120:
NP=NP-1;
if(NP<=0) return;
for(L=K;L<=NP;L++)
{
    afFt[L]=afFt[L+1];
    afBt[L]=afBt[L+1];
}
goto L110;
}

```

```

void TDsp2::Ordena(int M,float*X,float*Y,float*AR,float*BR)
{
    float SMIN,TMIN,YP,YPP;
    int JX,JY,JLOC;

```

```

    MultCte(AR,AR,M,0.0);
    MultCte(BR,BR,M,0.0);
    NP=M;

```

```

    SMIN=1*pow10(-37);
    YP=0.0;
    for(JX=1;JX<=(NP);JX++)
    {
        TMIN=1*pow10(37);
        JLOC=0;
        YPP=1*pow10(37);
        for(JY=1;JY<=(NP);JY++)
        {
            if(Y[JY]<=0) goto L60;
            if((X[JY]-SMIN)<0) goto L60;
            if((X[JY]-SMIN)==0) goto L30;
            if((X[JY]-TMIN)<0) goto L50;
            if((X[JY]-TMIN)==0) goto L40;
            if((X[JY]-TMIN)>0) goto L60;
L30:
            if((Y[JY]-YP)<=0) goto L60;
L40:
            if((Y[JY]-YPP)>=0) goto L60;
L50:
            YPP=Y[JY];
            TMIN=X[JY];
            JLOC=JY;
L60:continue;
        }
        if(JLOC>0) goto L70;
        NP=JX-1;
        goto L90;

```

```

L70:
AR[JX]=X[JLOC];
YP=Y[JLOC];
BR[JX]=YP;
SMIN=TMIN;
}
L90:
if(NP<=0) return;
MultCte(AR,X,NP,1.0);
MultCte(BR,Y,NP,1.0);
return;
}

void TDsp2::Pitch(float *X,float* afAutoCorrelacao)
{

float IPX,IPPT[20],PAT[20],SFL,SFH,PINCR;
int IP,I,LPER,LOC,ILOW,IHGH;

LPER=4;
SFL=20.0;
SFH=400.0;

ILOW=fSf/SFH;
IHGH=fSf/SFL;
if(IHGH>nNpts) IHGH=nNpts;

//DWINDF(X,nNpts,IHAM,0.0,0.0);
//RACOR(IHGH,X,afAutoCorrelacao);
AutoCorrelacao(nNpts,X,afAutoCorrelacao);
for(I=1;I<=LPER;I++)
{
LOC=MaxAmpPos(afAutoCorrelacao,ILOW,IHGH);
IPPT[I]=LOC;
PAT[I]=afAutoCorrelacao[LOC];
afAutoCorrelacao[LOC]=0.0;
}
IP=IPPT[1];
for(I=2;I<=LPER;I++)
{
IPX=IPPT[I];
if(IPX<(IP-3)) IP=IPX;
}
for(I=1;I<LPER;I++)
{
LOC=IPPT[I];
afAutoCorrelacao[LOC]=PAT[I];
}
PVA=InterParab(afAutoCorrelacao[IP-1],afAutoCorrelacao[IP],afAutoCorrelacao[IP+1].&PINCR);
PPER=IP+(PINCR);
FO=fSf/PPER;

```

```
}
```

```
void TDsp2::AutoCorrelacao(int M,float*Y,float*R)
```

```
{
```

```
int I,J;
```

```
float SS,RSCL,SCL;
```

```
for( I=1;I<=M;I++)
```

```
{
```

```
    SS=0.0;
```

```
    for(J=1;J<=nNpts-I+1;J++)
```

```
        SS=SS+Y[J]*Y[J+I-1];
```

```
    R[I]=SS;
```

```
}
```

```
for(I=M+1;I<=nNpts;I++)
```

```
R[I]=0.0;
```

```
SCL=R[1];
```

```
RSCL=1.0/SCL;
```

```
MultCte(R,R.M.RSCL);
```

```
}
```



```

#include "dsp3.h"
#include <como.h>

TDsp3::TDsp3(int F,int N,int M,BOOL J,float PR,int P,int A)
:TDsp2(F,N,M,J){
    nNpts=N;
    PRNGE=PR;
    LP=P;
    LA=A;
    IPL=new int[nNpts+nOrdem+15];
    PTAMP=new float[nNpts+nOrdem+15];
    APQQ=new float[5];
    PPQQ=new float[5];
    PTLOC=new float[nNpts+nOrdem+15];
    XMOM=new float[100];
}
TDsp3::~~TDsp3()
{
    delete IPL;
    delete PTAMP;
    delete APQQ;
    delete PPQQ;
    delete PTLOC;
    delete XMOM;
}

void TDsp3::Perturbacao(float *sinal)
{
    X=sinal;
    int NNPP[5],K,J,L;
    float PTOLD,PTNEW,PTINC,VAR,VAR2;
    for(K=1;K<=2;K++)
    {
        NNPP[K]=LocPicos(X);
        PTOLD=0.0;
        for(J=1;J<=NNPP[K];J++)
        {
            L=IPL[J];

            if((L>1)&&(L<nNpts))
                PTAMP[J]=InterParab(X[L-1],X[L],X[L+1].&PTINC);
            else
            {
                PTINC=0.0;
                PTAMP[J]=X[L];
            }

            PTNEW=L+PTINC;
            PTLOC[J]=PTNEW-PTOLD;
            PTOLD=PTNEW;
        }
        // PPQQ[K]=QuocPert(&PTLOC[2],NNPP[K]-1,LP);
        PPQQ[K]=QuocPert(PTLOC,NNPP[K]-1,LP);
    }
}

```

```

APQQ[K]=QuocPert(PTAMP.NNPP[K].LA);
MultCtet(X.X.nNpts.-1.0);
}

```

```

K=1;

```

```

if(APQQ[2]<=APQQ[1]) K=2; fQAP=APQQ[K];
if(PPQQ[2]<=PPQQ[1]) fQPP=PPQQ[2]; else fQPP=PPQQ[1]; //dependia antes do teste do APQQ
//fQPP=PPQQ[K];
fNPP=NNPP[K];

```

```

K=1;

```

```

if(APQQ[2]<=APQQ[1]) K=2;
fQAP=APQQ[K];
fQPP=PPQQ[K];
fNPP=NNPP[K];
Momento(X.nNpts.4.XMOM);
VAR=XMOM[2]-XMOM[1]*XMOM[1];
VAR2=VAR*VAR;
fEx=0.0;
if(VAR>0)
fEx=(XMOM[1]*(XMOM[1]*(XMOM[1]*(-3*XMOM[1])+6*XMOM[2])-4*XMOM[3])+XMOM[4])/
    VAR2-3.0;
}

```

```

float TDsp3::QuocPert(float*D,int N,int K){

```

```

float XNUM.YNUM.XDEN;
int I,J,M,NK;
XNUM=0.0;
XDEN=0.0;
M=K/2;

```

```

if(K-M-M-1 != 0) return(0.0);

```

```

NK=N-K+1;
if(NK<=0) return(0.0);
for(I=3;I<=NK;I++) // modificado: despreza o primeiro elemento
{
    YNUM=0.0;
    for (J=1;J<=K;J++)
        YNUM=YNUM+D[I+J-1];
    XNUM=XNUM+fabs(YNUM/K-D[I+M]);
}

```

```

XNUM=XNUM/NK;
for(I=2;I<=N;I++)
    XDEN=XDEN+D[I];
XDEN=XDEN/N;

```

```
return(XNUM/XDEN);  
}
```

```
int TDsp3::LocPicos(float *X){  
int INDX,INDX1,INDX2,MAX;  
float MLOW,MHGH;  
MAX=MaxAmpPos(X,1,nNpts);  
IPL[1]=MAX;  
MLOW=(1.0-PRNGE)*PPER;  
MHGH=(1.0+PRNGE)*PPER;  
int NP=1;  
INDX=MAX;  
INDX1=INDX-MHGH;  
INDX2=INDX-MLOW;  
while(INDX1>1){  
INDX=MaxAmpPos(X,INDX1,INDX2);  
NP++;  
IPL[NP]=INDX;  
INDX1=INDX-MHGH;  
INDX2=INDX-MLOW;  
}  
Reverte(IPL,IPL,NP,1);  
INDX=MAX;  
INDX1=INDX+MLOW;  
INDX2=INDX+MHGH;  
while(INDX2<nNpts){  
INDX=MaxAmpPos(X,INDX1,INDX2);  
NP++;  
IPL[NP]=INDX;  
INDX1=INDX+MLOW;  
INDX2=INDX+MHGH;  
}  
return(NP);  
}
```


A2. Resultados das análises das vozes normais

Fo	Ex	PA	PPQ	APQ	SFR	SFF
102	27.6527	0.5583	0.03	5.3	-3.9261	-11.797
102.1	38.632	0.5203	0.02	4.68	-3.6704	-11.2516
102	25.036	0.6557	0.08	4.25	-4.9153	-12.562
101.8	20.8134	0.3963	0.1	4.91	-3.7067	-13.2008
101.2	24.4604	0.445	0.01	2.75	-3.887	-13.2498
178	8.2894	0.4207	0.43	9.41	-4.8077	-15.0567
174.6	10.7934	0.5338	5.06	9.7	-4.8448	-14.7241
174.7	6.351	0.4369	0.34	10.45	-4.1132	-15.3606
178.1	15.6817	0.4237	5.03	12.81	-3.7215	-14.5188
174.6	100.3413	0.2445	6.79	10.56	-4.2852	-14.4884
200.6	3.3521	0.6392	1.85	22.74	-5.5921	-10.7375
201	5.7415	0.4978	2	19.76	-5.4013	-10.5692
200.8	6.4671	0.6381	12.3	10.59	-5.8637	-10.8208
200.8	5.9936	0.5988	0.34	12.75	-5.959	-10.763
200.8	5.9655	0.6213	8.71	11.64	-4.9769	-10.2059
174.6	4.7773	0.5933	0.36	10.01	-5.6829	-10.6211
175	7.0817	0.5003	4	-112.96	-5.3885	-10.3795
173.8	8.0471	0.5019	0.18	5.02	-6.1206	-10.229
174.2	7.435	0.5107	7.08	8.81	-4.9228	-9.8997
173	5.7045	0.5933	0.17	7.02	-5.9008	-10.1599
175	6.254	0.5009	0.28	3.71	-5.1054	-9.8682
174.4	6.9552	0.5836	0.24	10.19	-6.3373	-10.0148
212.1	7.0782	0.694	0.06	5.88	-8.0278	-14.4306
211.9	16.2964	0.6012	0.57	7.06	-7.2227	-13.8274
211.7	4.183	0.6706	0.29	6.89	-6.9356	-13.186
210.7	5.0389	0.6303	7.32	10.16	-6.6764	-13.2202
212.1	5.3581	0.5209	0.37	11.62	-7.1156	-13.0555
211.5	6.6146	0.6461	0.23	10.08	-6.5479	-13.0759
209.7	14.9574	0.4582	0.49	12.5	-5.9925	-13.3182
167.1	6.7151	0.7126	0.68	7.73	-5.7527	-14.7543
167.3	7.1059	0.6487	0.37	10.97	-6.2044	-14.6337
167.5	11.5094	0.6	1.18	11.55	-5.8984	-14.3599
167.5	11.5094	0.6	1.18	11.55	-5.8984	-14.3599
100	9.4398	0.469	0.05	2.69	-1.871	-17.0212
100.2	29.4917	0.3458	6.83	12.09	-3.3363	-20.5147
100.4	16.7897	0.3233	0.07	17.86	-2.5316	-19.0388
99.4	10.6742	0.4291	3.21	2.52	-2.8574	-18.9571
98.7	21.2713	0.3128	0.02	14.31	-3.0201	-20.1612
99.4	12.9787	0.3353	0.06	9.14	-2.8066	-19.9974
98.8	25.5066	0.275	4.19	2.98	-2.9751	-20.2647
99.3	16.2642	0.3561	1.45	8.56	-3.0376	-20.4532
116.9	7.806	0.3073	0.55	4.79	-4.6327	-14.443
119.7	5.369	0.2936	0.49	6.07	-5.0854	-14.8131
115.8	4.4154	0.379	0.14	6.51	-6.1418	-15.4058
116.7	4.9596	0.4104	0.28	11.64	-5.2374	-15.1234
116.7	7.9075	0.4404	0.16	6.45	-4.5929	-13.7293
196.3	2.4863	0.461	0.43	11.6	-4.5939	-16.207
193.5	8.9133	0.7021	0.34	14.19	-7.5323	-8.4569
193.3	11.3718	0.7438	1.45	17.88	-7.8148	-8.8014

193.3	10.4785	0.7027	0.53	10.49	-7.0901	-8.8404
193.4	12.4474	0.7639	0.1	7.29	-7.7087	-8.5114
193.4	12.4474	0.7639	0.1	7.29	-7.7087	-8.5114
193.4	10.0001	0.7031	0.24	13.01	-7.7921	-8.6885
193.4	8.9477	0.6512	1.95	14.22	-7.2762	-9.2213
193.4	9.1603	0.7409	0.11	6.09	-6.7853	-9.3014
193.4	8.3005	0.7112	1.03	10.57	-7.0526	-9.4569
193.4	7.6356	0.7502	0.08	9.88	-7.5146	-9.2697
94.8	34.5824	0.2892	0.1	3.03	-2.9077	-14.3724
93.5	4.4105	0.3862	0	4.63	-3.3681	-17.1733
96.7	9.5525	0.282	0.04	14.85	-3.0626	-16.1047
94.6	16.3653	0.3917	0.05	3.93	-3.0293	-14.2536
94.2	17.4682	0.4751	0.21	1.1	-3.4824	-12.6142
145.3	8.6449	0.7149	0.13	7.15	-5.9098	-13.6003
145.4	8.1081	0.6387	0.11	3.95	-5.8918	-14.0343
125.5	10.3525	0.5192	0.12	2.91	-5.5795	-12.339
125.3	8.2842	0.5228	0.17	5.08	-5.3761	-12.9542
124.1	10.0102	0.6754	0.02	5.45	-5.6351	-12.4717
124.1	9.5658	0.6549	0.22	3.05	-5.4725	-12.8441
123.8	8.4958	0.6538	0.2	5	-6.6539	-13.6577
123.9	14.2062	0.559	0.22	4.35	-5.1218	-12.006
102.5	6.5687	0.3852	0.11	1.95	-4.1958	-15.9746
101.4	8.0167	0.42	0.17	4.2	-4.4185	-15.5911
100.8	8.761	0.4518	0	3.97	-4.0619	-15.1002
100	9.5386	0.4866	0.03	4.44	-4.5268	-15.0006
99.7	14.3044	0.4106	0.13	4.55	-4.1024	-14.2305
99.5	8.6073	0.3795	0.04	4.72	-5.2298	-15.146
137.8	8.6209	0.4926	0.11	9.44	-6.0741	-12.3615
136.7	14.2241	0.4745	0.27	8.42	-4.5551	-11.6505
137.5	12.1765	0.5058	0.17	4.43	-6.212	-12.6849
136.7	14.2241	0.4745	0.27	8.42	-4.5551	-11.6505
137.5	12.1765	0.5058	0.17	4.43	-6.212	-12.6849
139.1	15.3285	0.2748	0.11	9.05	-4.3003	-13.2624
136.8	7.1487	0.4162	0.27	8.82	-4.8479	-13.0221
208.2	4.7434	0.5297	0.32	7.4	-5.2356	-13.7496
208.9	1.5899	0.5664	0.48	6.97	-4.9482	-12.8571
211.5	1.4218	0.5667	0.34	8.72	-5.8253	-14.3851
211.3	3.5498	0.5238	0.41	12.02	-5.4641	-14.1847
174.8	2.1923	0.4802	0.56	7.98	-4.5968	-13.0136
174.8	1.4277	0.567	0.62	8.6	-5.6024	-12.3972
172.5	5.9676	0.5422	0.29	8.51	-5.0884	-13.213
186.6	1.8862	0.5264	0.46	9.82	-5.2921	-10.5262
187.7	1.8618	0.5095	1.14	11.17	-5.7552	-10.0549
186.6	7.1539	0.4222	2.76	10.18	-5.981	-10.5678
186.8	2.1809	0.5542	0.59	9.11	-5.6144	-9.5931
193.3	12.7555	0.7651	0.07	4.78	-7.9501	-12.2106
192.6	9.7111	0.6211	0.25	4.38	-6.2063	-12.2199
190.2	6.8944	0.6249	0.2	7.97	-5.4373	-12.3066
193.4	7.7454	0.614	0.23	4.88	-6.7181	-12.5034
192.1	7.9973	0.6016	0.16	6.33	-6.4917	-12.0842
97.8	14.254	0.3603	0.06	2.81	-3.6556	-15.1217
99	15.7909	0.4527	0.07	8.91	-4.4019	-15.2735
99.2	17.1011	0.6039	0	0.6	-5.2113	-14.2676
99.3	14.8832	0.558	0	0.66	-5.5923	-14.3595
99.5	15.2958	0.6213	0.03	2.89	-4.5616	-14.0699

99.5	15.2958	0.6213	0.03	2.89	-4.5616	-14.0699
99.5	17.4429	0.6238	0.07	2.07	-5.0212	-14.1855
99.3	20.2469	0.4518	0.06	3.6	-5.2336	-14.7505
99	15.1007	0.5572	0.02	3.99	-5.5275	-15.1074
99.4	17.9984	0.5146	0.08	2.85	-5.6558	-14.6882
193.5	2.3822	0.4014	0.35	8.98	-4.458	-14.3896
192.9	2.2797	0.4391	0.3	12.53	-4.7578	-14.5822
193.1	1.6043	0.4673	0.49	7.85	-4.7381	-14.5399
192	1.8538	0.4568	0.83	11.46	-4.7302	-14.4333
191.6	2.3825	0.4145	0.87	7.34	-4.3716	-14.8629
200.8	4.4846	0.6705	0.16	7.45	-6.6118	-12.0679
203.6	3.079	0.6561	0.23	7.07	-6.9764	-12.2285
203.8	3.4166	0.6646	0.33	9.68	-6.8753	-11.788
200.5	7.3909	0.5475	5.64	10.94	-5.9493	-10.2412
200.7	7.0521	0.7444	0.28	10.73	-8.0709	-9.6779
143	12.5276	0.4051	0.39	8.78	-3.8308	-10.3296
143.2	7.8256	0.2267	1.33	7.03	-2.7515	-9.551
143.2	8.2562	0.2402	1.35	6.87	-3.1842	-9.1087
134.8	8.972	0.3888	0.23	8.62	-3.524	-9.0614
134.2	5.3097	0.3966	0.43	5.62	-4.0272	-9.796
135.6	3.7676	0.3545	0.2	7.05	-3.7777	-10.7582

A3. Resultados das análises das vozes patológicas

Fo	EX	PA	PPQ	APQ	SFR	SFF
230.4	0.3371	0.1256	14.62	13.59	-3.343	-3.9454
208.2	8.5936	0.262	0.34	7.26	-2.8076	-8.9787
117.3	3.8694	0.3065	0.13	6.92	-3.0523	-11.7354
210.6	1.64	0.3585	4.79	10.18	-4.1197	-9.4919
133.3	7.0971	0.2991	1.19	14.99	-4.2977	-24.5008
282.4	0.7372	0.2615	10.55	12.6	-2.9733	-4.7651
191.9	1.4336	0.3241	2.69	9.57	-5.0675	-11.6703
295.5	-0.0027	0.1742	11.56	14.08	-3.1881	-1.9701
226.8	1.2095	0.345	5.05	8.76	-5.3181	-11.5911
206.3	8.8239	0.5012	7.49	11.36	-5.7599	-9.1272
293.1	9.7692	0.2515	10.08	12.61	-2.9409	-13.7707
196.7	13.6147	0.3752	1.31	15.74	-4.623	-12.1425
231.4	0.4453	-0.0029	5.81	5.24	-3.3627	-6.8247
175.1	2.3621	0.1143	10.92	17.57	-2.4741	-3.4725
327.1	0.8539	0.2774	14.27	11.02	-3.8779	-9.2125
329.3	2.2969	0.0882	13.92	18.23	-2.6705	-8.7555
209.4	1.5413	0.1455	3.29	5.78	-2.854	-10.4295
224.3	203.9389	0.1398	0.43	7.91	-2.8596	-16.1596
296.1	1.4656	0.3417	3.48	13.28	-3.3901	-9.6886
197.1	4.5257	0.6125	11.93	17.8	-6.1309	-3.375
133.9	5.723	0.6159	0.49	7.6	-5.7268	-4.5475
229.6	2.3692	0.2062	9.56	13.09	-3.4523	-8.0095
380.8	57.7607	0.049	14.28	25.35	-0.0806	-9.4407
281.8	0.8281	0.0694	10.46	17.99	-1.5371	-9.5142
264.6	5.7431	0.221	3.47	31.66	-3.6321	-9.1427
254.4	1.0756	-0.0047	12.58	15.15	-2.2454	-12.3205
143.2	8.2562	0.2402	1.35	6.87	-3.1842	-9.1087
239.3	0.7543	0.4377	1.22	14.47	-4.7228	-10.9842
310.8	0.8623	0.1465	4.17	14.54	-2.9498	-11.312
192.3	2.4828	0.2769	12.8	18.15	-4.3846	-10.8111
134.5	8.8979	0.624	1.27	6.2	-6.216	-13.1122
100	9.4398	0.469	0.05	2.69	-1.871	-17.0212
206.8	5.1514	0.5913	9.05	11.14	-7.0123	-6.975
100.2	11.505	0.6822	0.36	10.96	-6.375	-5.8935
125.3	7.4567	0.645	3.45	10.34	-6.962	-9.5412