

UNIVERSIDADE CATÓLICA DE PELOTAS  
ESCOLA DE INFORMÁTICA  
CIÊNCIA DA COMPUTAÇÃO

## **Aplicações da Filtragem Homomórfica a Sinais de Voz**

Por

Diego Ferreira dos Santos

Projeto de Graduação submetido como requisito parcial à obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Cláudio Villela

Co-Orientadores: Prof. Márcio Costa  
Prof. Alexandre Pino

Pelotas, Julho de 2000

## **Agradecimentos**

Gostaria de agradecer aos professores Cláudio Villela, Márcio Holsbach Costa e Alexandre Pino por seus conselhos e pela participação que foi decisiva e fundamental para elaboração dessa monografia.

A Clínica Castagno pelo fornecimento dos sinais de voz dos pacientes e a todos os meus colegas e professores que de uma maneira direta ou indireta participaram deste projeto.

# SUMÁRIO

<b>1</b>	<b><i>Introdução</i></b>	<b>11</b>
1.1	<b>Preâmbulo</b>	<b>11</b>
1.2	<b>Objetivos</b>	<b>11</b>
1.2.1	Gerais	11
1.2.2	Específicos	11
1.3	<b>Justificativa</b>	<b>11</b>
1.4	<b>Motivação</b>	<b>12</b>
1.5	<b>Organização</b>	<b>12</b>
<b>2</b>	<b><i>Fundamentos Teóricos</i></b>	<b>13</b>
2.1	<b>Anatomia e Princípios Funcionais do Aparelho Fonador</b>	<b>13</b>
2.1.1	Descrição de Patologias	18
2.2	<b>Descrição e Caracterização de Sinais de Voz</b>	<b>20</b>
2.2.1	Caracterização da Voz	20
2.2.1.1	Informação Lingüística	20
2.2.1.2	Mecanismos de Produção de Voz	20
2.2.1.3	Características Acústicas da Voz	22
2.2.1.4	Características Estatísticas da Voz	23
2.2.2	Modelos de Produção de Voz	23
2.3	<b>Aquisição de Sinais de Voz</b>	<b>24</b>
2.3.1	Conversão Analógico-Digital	24
2.3.1.1	Amostragem de Sinais Contínuos	24
2.3.1.2	Quantização	26
2.3.2	Condicionamento do Sinal	27
2.4	<b>Técnicas de Processamento de Sinais Aplicadas à Voz</b>	<b>28</b>
2.4.1	Filtragem Digital	28
2.4.2	Análise de Fourier	29
2.4.3	Filtragem Homomórfica	31
<b>3</b>	<b><i>Análise dos Sinais de Voz</i></b>	<b>34</b>

<b>3.1</b>	<b>Aquisição dos Sinais de Voz</b>	<b>34</b>
3.1.1	Descrição do <i>Hardware</i> utilizado	34
3.1.2	Descrição da Metodologia de Aquisição	34
3.1.3	Plataforma da Análise dos Sinais	35
<b>3.2</b>	<b>Considerações da Análise</b>	<b>35</b>
<b>3.3</b>	<b>Análise do Espectro de Voz</b>	<b>36</b>
<b>3.4</b>	<b>Análise do Cepstrum Real</b>	<b>38</b>
<b>3.5</b>	<b>Análise das Formantes Utilizando o Cepstrum Real</b>	<b>38</b>
<b>3.6</b>	<b>Resposta ao Impulso do Trato Vocal</b>	<b>40</b>
<b>3.7</b>	<b>Espectro da Resposta ao Impulso do Trato Vocal</b>	<b>41</b>
<b>4</b>	<b><i>Protótipo para um Sistema de Auxílio Clínico ao Diagnóstico de Patologias Vocais</i></b>	<b>43</b>
4.1	Características da Linguagem para a Aplicação	43
4.2	Bibliotecas Utilizadas	44
4.3	Características do Usuário	45
4.4	Fluxograma do Protótipo	46
4.5	Banco de Dados	46
4.6	Descrição dos Módulos	47
4.6.1	Módulo para Leitura do Formato de Arquivo .WAV	47
4.6.2	Rotinas para o Processamento de Sinais Digitais	49
4.7	Descrição da Funcionalidade do <i>Software</i>	49
<b>5</b>	<b><i>Conclusão</i></b>	<b>58</b>
5.1	Possibilidades Futuras	58
<b>6</b>	<b><i>Bibliografia</i></b>	<b>59</b>
<b>7</b>	<b><i>Anexo</i></b>	<b>60</b>
7.1	Anexo A – Espectro de Voz	61
7.2	Anexo B – Cepstrum Real	62
7.3	Anexo C – Formantes	65

<b>7.4</b>	<b>Anexo D - Resposta ao Impulso do Trato Vocal</b>	<b>68</b>
<b>7.5</b>	<b>Anexo E – Espectro da Resposta Impulsiva do Trato Vocal</b>	<b>71</b>
<b>7.6</b>	<b>Anexo F – Biblioteca para Leitura do Formato de Arquivo .WAV</b>	<b>72</b>
<b>7.7</b>	<b>Anexo G – Rotinas em Delphi para Análise de Sinais</b>	<b>75</b>
7.7.1.1	Espectro de Potência	75
7.7.1.2	Cepstrum Real	76
7.7.1.3	Análise das Formantes	78

## Lista de Figuras

<i>Figura 2-1 Diagrama dos órgãos que compõem o aparelho fonador humano</i>	13
<i>Figura 2-2 Laringe em várias posições</i>	15
<i>Figura 2-3 Cordas vocais aduzidas</i>	16
<i>Figura 2-4 Cordas vocais abduzidas</i>	16
<i>Figura 2-5 Modelo de produção de voz. Extraído de (OPPENHEIM, 1989)</i>	21
<i>Figura 2-6 Exemplos de sons-vozeados e não-vozeados. Extraído de (PELTON, 1992)</i>	22
<i>Figura 2-7 Modelo da máquina húngara - extraído de (ROCHA, 1997)</i>	23
<i>Figura 2-8 Sinal analógico no domínio tempo (a) - Sinal discreto no domínio tempo (b). Extraído de (PELTON,1992)</i>	24
<i>Figura 2-9 Aliasing. a) Sequência de valores discretos no tempo. b) Duas senóides diferentes que passam pelos mesmos pontos da sequência discreta. Extraído de (LYONS, 1999)</i>	26
<i>Figura 2-10 Exemplo de um sinal digitalizado (quantização). Extraído de (SMITH,1999)</i>	27
<i>Figura 2-11 Diagrama de um DSP. Extraído de (SMITH, 1999)</i>	27
<i>Figura 2-12 Exemplo de Transformada de Fourier. Extraído de (OPPENHEIM, 1989)</i>	29
<i>Figura 2-13 Exemplos de janelamento</i>	30
<i>Figura 2-14 Cepstrum de um sinal de fala</i>	32
<i>Figura 2-15 Filtragem homomórfica</i>	33
<i>Figura 3-1 Paciente 1 antes do tratamento</i>	36
<i>Figura 3-2 Paciente 1 após o tratamento</i>	36
<i>Figura 3-3 Espectro de voz (Paciente 1). Vermelho = sinal antes do tratamento; Azul = sinal após o tratamento</i>	37
<i>Figura 3-4 Espectro de voz (Paciente 2). Vermelho = sinal antes do tratamento; Azul = sinal após o tratamento</i>	37
<i>Figura 3-5 Paciente 2 antes do tratamento</i>	38
<i>Figura 3-6 Paciente 2 após o tratamento</i>	38
<i>Figura 3-7 Paciente 1 antes do tratamento</i>	40
<i>Figura 3-8 Paciente 1 após o tratamento</i>	40

<i>Figura 3-9 Paciente 1 antes do tratamento</i>	41
<i>Figura 3-10 Paciente 1 após o tratamento</i>	41
<i>Figura 3-11 Vermelho = Paciente 1 antes do tratamento. Azul = Paciente 1 após o tratamento</i>	42
<i>Figura 3-12 Vermelho = Paciente 5 antes do tratamento. Azul = Paciente 5 após o tratamento</i>	42
<i>Figura 4-1 Biblioteca ZoomDisplay</i>	45
<i>Figura 4-2 Fluxograma do protótipo</i>	46
<i>Figura 4-3 Modelo ER do banco de dados</i>	46
<i>Figura 4-4 Tela inicial do protótipo</i>	49
<i>Figura 4-5 Menu arquivo do protótipo</i>	49
<i>Figura 4-6 Janela cadastro</i>	50
<i>Figura 4-7 Janela de cadastro de patologias</i>	50
<i>Figura 4-8 Janela de cadastro de sinal (aquisição do sinal de voz)</i>	51
<i>Figura 4-9 Mensagem explicando como fazer a gravação</i>	51
<i>Figura 4-10 Sinal do paciente</i>	52
<i>Figura 4-11 Caixa de diálogo dizendo que tudo ocorreu sem problemas</i>	52
<i>Figura 4-12 Janela de análise de voz</i>	53
<i>Figura 4-13 Sinais apresentados na ListBox do paciente DIEGO com LARINGITE AGUDA</i>	53
<i>Figura 4-14 Visualização do gráfico para análise</i>	54
<i>Figura 4-15 Zoom do sinal</i>	54
<i>Figura 4-16 No gráfico baixo está o trecho de 1 segundo selecionado no gráfico acima</i>	55
<i>Figura 4-17 Espectro de potência do trecho selecionado</i>	56
<i>Figura 4-18 Cepstrum real do trecho selecionado</i>	56
<i>Figura 4-19 Análise das formantes do trecho selecionado</i>	57
<i>Figura 4-20 Visualização das 3 primeiras formantes</i>	57
<i>Figura 7-1 Paciente 3. Vermelho = sinal antes do tratamento. Azul = sinal após o tratamento</i>	61
<i>Figura 7-2 Paciente 4. Vermelho = sinal antes do tratamento. Azul = sinal após o tratamento</i>	61
<i>Figura 7-3 Paciente 5. Vermelho = sinal antes do tratamento. Azul = sinal após o tratamento</i>	61

<i>Figura 7-4 Paciente 1 antes do tratamento</i>	62
<i>Figura 7-5 Paciente 1 após o tratamento</i>	62
<i>Figura 7-6 Paciente 3 antes do tratamento</i>	62
<i>Figura 7-7 Paciente 3 após o tratamento</i>	63
<i>Figura 7-8 Paciente 4 antes do tratamento</i>	63
<i>Figura 7-9 Paciente 4 após o tratamento</i>	63
<i>Figura 7-10 Paciente 5 antes do tratamento</i>	64
<i>Figura 7-11 Paciente 5 após o tratamento</i>	64
<i>Figura 7-12 Paciente 2 antes do tratamento</i>	65
<i>Figura 7-13 Paciente 2 após o tratamento</i>	65
<i>Figura 7-14 Paciente 3 antes do tratamento</i>	65
<i>Figura 7-15 Paciente 3 após o tratamento</i>	66
<i>Figura 7-16 Paciente 4 antes do tratamento</i>	66
<i>Figura 7-17 Paciente 4 após o tratamento</i>	66
<i>Figura 7-18 Paciente 5 antes do tratamento</i>	67
<i>Figura 7-19 Paciente 5 após o tratamento</i>	67
<i>Figura 7-20 Paciente 2 antes do tratamento</i>	68
<i>Figura 7-21 Paciente 2 após o tratamento</i>	68
<i>Figura 7-22 Paciente 3 antes do tratamento</i>	68
<i>Figura 7-23 Paciente 3 após o tratamento</i>	69
<i>Figura 7-24 Paciente 4 antes do tratamento</i>	69
<i>Figura 7-25 Paciente 4 após o tratamento</i>	69
<i>Figura 7-26 Paciente 5 antes do tratamento</i>	70
<i>Figura 7-27 Paciente 5 após o tratamento</i>	70
<i>Figura 7-28 Vermelho = Paciente 2 antes do tratamento. Azul = Paciente 2 após o tratamento</i>	71
<i>Figura 7-29 Vermelho = Paciente 3 antes do tratamento. Azul = Paciente 3 após o tratamento</i>	71
<i>Figura 7-30 Vermelho = Paciente 4 antes do tratamento. Azul = Paciente 4 após o tratamento</i>	71



## Resumo

Diversas patologias do trato vocal são caracterizadas pela diminuição da complacência e/ou elasticidade. A avaliação clínica da evolução do tratamento geralmente é realizada de forma visual e baseada na diminuição da sintomatologia apresentada, visto não existirem métodos quantitativos padronizados. A técnica de filtragem homomórfica permite uma estimativa da função de transferência do trato vocal e tem sido utilizada com sucesso na área de processamento de voz. O presente trabalho apresenta os resultados preliminares de um estudo sobre a potencialidade do uso de filtragem homomórfica como uma ferramenta para a avaliação das condições anatômicas do trato vocal. São apresentados resultados comparativos referentes à análise do *cepstrum*, resposta ao impulso do trato vocal e espectro de potência filtrado, bem como suas particularidades e vantagens sobre os métodos convencionais de análise espectral. Finalizando, é apresentada a proposta de um protótipo de programa para análise de sinais de voz utilizando as referidas técnicas.

**Palavras-Chave:** Filtragem Homomórfica, Cepstrum, Processamento de Sinais

## **Abstract**

Several pathologies of the vocal tract are characterized by the decrease of the complacency and/or elasticity. The clinical evaluation of the evolution of the treatment is generally accomplished in a visual way and based on the decrease of the presented symptomatology, since standardized quantitative methods doesn't exist. The technique of homomorphic filtering allows an estimate of the transfer function of the vocal tract and it has been used with success in the area of speech processing. This work presents preliminary results of a study about the potentiality of the use of homomorphic filtering as a tool for the evaluation of the anatomical conditions of the vocal tract. Results are presented for the cepstrum analysis, impulse response of the vocal tract and filtered power spectrum, as well as their particularities and advantages over the conventional methods of spectrum analysis. Concluding, the proposal of a prototype is presented for analysis of speech signals using the referred techniques.

**Keywords:** Homomorphic Filtering, Cepstrum, Signal Processing

# **Introdução**

## **1.1 Preâmbulo**

Nas últimas décadas, a informática vem ganhando uma maior importância no desenvolvimento de sistemas que auxiliam a Medicina. Atualmente a utilização de computadores na área da saúde (consultórios, clínicas, laboratórios, etc.) é um dos segmentos do mercado que mais crescem. O baixo custo do *hardware* e a oferta de diversos *softwares* especializados vêm despertando a comunidade médica para a importância da informática e de como ela pode aumentar sua produtividade e conseqüentemente seus benefícios.

Hoje é raro um profissional que não disponha do seu computador, que a cada dia que passa se torna mais potente, para auxiliar na sua profissão. Nós podemos ver a informática auxiliando médicos na cardiologia, na ultra-sonografia, na fisioterapia, etc.

## **1.2 Objetivos**

### **1.2.1 Gerais**

O objetivo deste trabalho é realizar um estudo sobre a potencialidade do uso de técnicas de filtragem como uma ferramenta para a avaliação das condições clínicas do trato vocal.

### **1.2.2 Específicos**

Os objetivos específicos para este projeto são pesquisar as características do sinal de voz, pesquisar técnicas sobre aquisição de sinais de voz, pesquisar técnicas de filtragem para fazer a análise das vozes dos pacientes antes e após o tratamento, realizar testes utilizando sinais reais de pacientes, implementar um protótipo para auxiliar no diagnóstico de patologias do trato vocal.

## **1.3 Justificativa**

O diagnóstico clínico da evolução do tratamento de patologias vocais é realizado pelo médico baseando-se na diminuição da sintomatologia apresentada e de

sua experiência. O desenvolvimento de métodos quantitativos permitiria ao clínico uma análise mais consistente no auxílio ao diagnóstico.

Diversos trabalhos abordam este tema, podemos citar: (ROSA, 2000), (ZHU, 2000)

A informática na medicina é uma área em pleno desenvolvimento e a satisfação de ver um trabalho sendo usado para ajudar na cura de doenças compensa todo o esforço dispensado para a realização do mesmo.

#### **1.4 Motivação**

Este documento poderá contribuir para pesquisas posteriores na área de processamento de sinais de voz, de técnicas de filtragem, mais especificamente a filtragem homomórfica, filtragem digital, e a Transformada Discreta de Fourier. Além disso, poderá auxiliar no desenvolvimento de novas técnicas e equipamentos na área da medicina para o tratamento vocal.

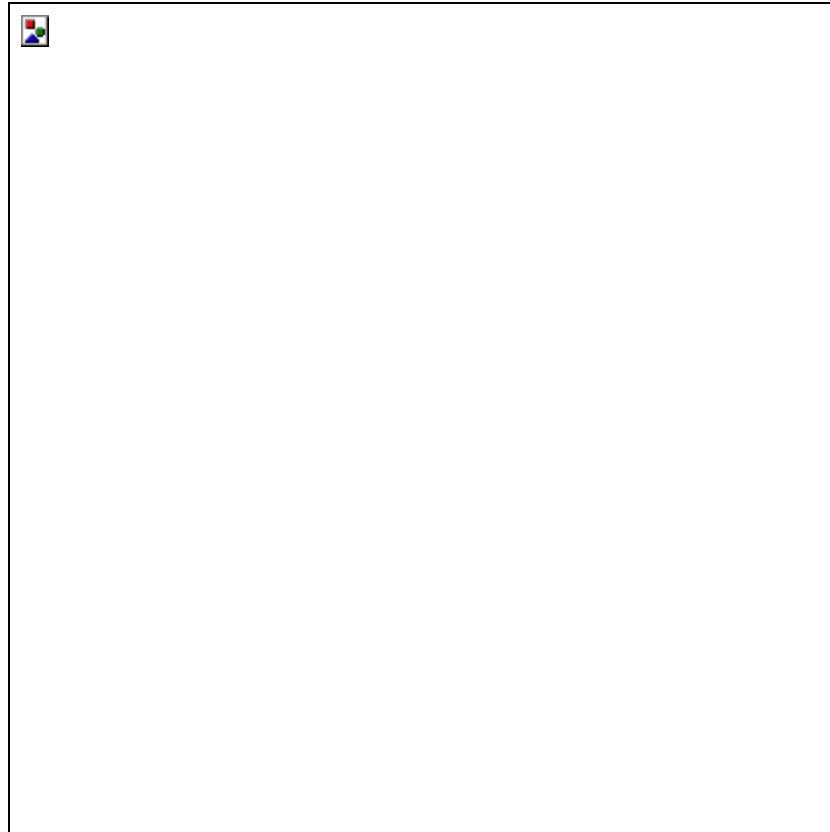
#### **1.5 Organização**

O projeto está dividido em cinco partes:

A Introdução apresenta uma visão geral sobre o assunto do projeto bem como os objetivos, as justificativas e motivação para a realização do mesmo. Os Fundamentos Teóricos apresentam uma visão teórica sobre o processamento de sinais de voz. Como é a anatomia do trato vocal; descrição das patologias relacionadas à voz; características do processamento de voz; métodos de aquisição de sinais de voz e descrição de técnicas de processamento de sinais. A Análise dos Sinais de Voz apresenta os resultados obtidos através da análise de voz de pacientes antes e após o tratamento. O Protótipo apresenta uma base para um possível *software* de auxílio clínico ao diagnóstico de patologias vocais desenvolvido na linguagem de programação Delphi. Na Conclusão são apresentadas as conclusões sobre a análise de sinais dos pacientes, assim como as possibilidades futuras de análise e de desenvolvimento de um *software*, tendo como base o protótipo.

## 2 Fundamentos Teóricos

### 2.1 Anatomia e Princípios Funcionais do Aparelho Fonador



**Figura 2-1 Diagrama dos órgãos que compõem o aparelho fonador humano**

Vários órgãos são responsáveis pela produção da voz humana (Figura 2-1). Do ponto de vista fonatório, encontramos, no falante: o produtor, o condutor, o vibrador, o ressoador e o articulador.

**O Produtor:** é constituído pelos pulmões e músculos respiratórios, que provocam a inspiração e expiração.

A caixa torácica é formada posteriormente pela coluna vertebral e anteriormente, pelo osso esterno e doze pares de costelas que ligam à coluna e ao esterno, formando arcos. Essa armação é móvel, pois as doze costelas articulam-se para trás com a ajuda das vértebras dorsais e para frente, por intermédio das cartilagens costais.

Os pulmões estão contidos na caixa torácica, e além de outras funções, eles são essenciais para a respiração. O volume de ar dos pulmões varia de falante para falante, mas sempre aumenta na inspiração e diminui, na expiração. No interior de cada um dos pulmões estão os brônquios intrapulmonares.

O diafragma é um músculo que se abaixa durante a inspiração e sobe na expiração. Quando abaixa, forma uma pressão negativa e o ar entra nos pulmões. Na expiração, o diafragma sobe, diminuindo o volume pulmonar e o ar contido nos pulmões sai ativamente.

**O Condutor:** é formado por um conjunto de tubos do aparelho fonador que são responsáveis pelo percurso de corrente de ar. São eles: brônquios, traquéia, laringe, faringe, boca e fossas nasais.

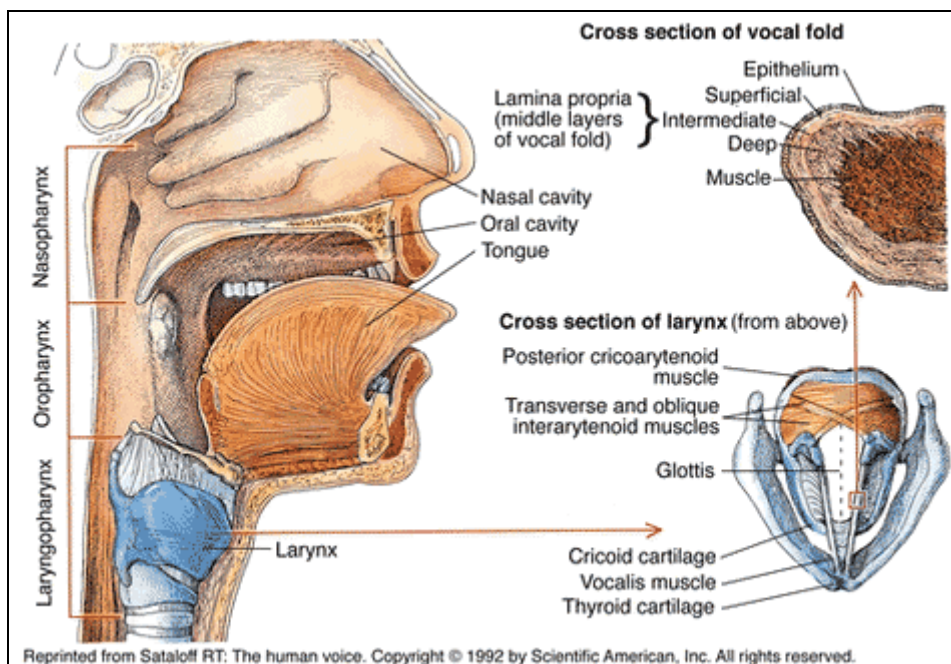
Os brônquios são resultantes da bifurcação da traquéia, distinguindo-se um à direita e outro à esquerda. Ambos dirigem-se obliquamente para baixo até atingirem o hilo do pulmão correspondente.

A traquéia é um tubo que para a fonação funciona como condutor de corrente de ar dos brônquios para a laringe. Em seu trajeto, ela é um canal único, formado por anéis cartilagosos, que a tornam irregular e coloca-se, assim como a laringe, à frente do canal alimentar. Inicialmente, a traquéia ocupa a parte inferior do pescoço e depois penetra no tórax. Como condutor na inspiração, a traquéia conduz o ar da laringe para os brônquios, e na expiração o trajeto é inverso. Por ser elástica, a traquéia acompanha a laringe em seus movimentos.

Como condutor, a laringe é um tubo músculo-cartilaginoso, que leva a corrente de ar da faringe para a traquéia. Para a fonação, a laringe é o vibrador.

A faringe, como condutor, é um tubo músculo-membranoso, situado no pescoço, ligado interiormente por um lado à laringe, e por outro ao esôfago. Na parte superior comunica-se com as cavidades nasal e bucal. Para a fonação, a faringe é ressoador.

**O Vibrador:** é formado pela laringe e cordas vocais.



**Figura 2-2 Laringe em várias posições**

A laringe tem o papel de vibrador, pois é aí que se produz a marca da sonoridade sobre os sons surdos, diferenciando-se assim os sons do idioma Português, que se caracteriza por ser bastante sonoro, pois temos poucas articulações surdas, apenas: [P, T, K, F, S, Š, TŠ, TS], todas as demais são sonoras. Ela limita-se anteriormente com a cartilagem tiróide e posteriormente com a cricóide (Figura 2-2). Sua fixação é devida à traquéia, na parte inferior; à faringe, na superior; e por um certo número de músculos e ligamentos está presa de um lado ao osso hióide e de outra à base do tórax.

A laringe é composta pelas cartilagens: cricóide, aritenóides, tiróide, epiglote e outras cartilagens, que para a produção de sonoridade, não são tão importantes.



**Figura 2-3 Cordas vocais aduzidas**



**Figura 2-4 Cordas vocais abduzidas**

As cordas vocais são formadas pelo músculo vocal e recobertas por ligamentos, elas são as duas pregas inferiores (Figura 2-3 e Figura 2-4) de dois pares de pregas situados um de cada lado da face interna da laringe. Na frente, as cordas vocais estão localizadas no ângulo da cartilagem tiróide, e atrás na apófise vocal das aritenóides. Ao vibrarem, pelo movimento dos músculos laríngenos, abrem e fecham a glote, vibrando vertical e horizontalmente. As pregas superiores (dos dois pares) são denominadas falsas cordas vocais ou pregas vestibulares e não tem papel importante para a fonação.

A Glote é o espaço delimitado entre as cordas vocais (Figura 2-4). Essa espécie de fenda tem uma função respiratória e outra vocal. Sua forma é variada e depende da posição tomada pelas cordas vocais. A glote estando aberta, o som é surdo; estando fechada (Figura 2-3), o som é sonoro.



Para a produção do som, as cordas vocais contraem-se sob o efeito do influxo nervoso que vem ativar as fibras musculares do tiro-aritenoideo interno. Então as fibras contraem-se e por conseqüência auxiliam a corrente de ar a separar as cordas vocais, permitindo que a pressão do ar subglótico insinue-se entre elas. A reação elástica que nasce pela contração dos músculos, leva-as a unirem-se novamente. Assim permanecem até que um novo influxo do recorrente as atinja. Em síntese, o mecanismo de produção de sonoridade é dado pela freqüência dos influxos do nervo recorrente, fazendo com que as fibras do músculo vocal contraíam-se e descontraíam-se sobre as inserções contidas na margem livre das cordas vocais, levando-as a vibrar com maior ou menor intensidade.

**O Ressonador:** é constituído por ventrículos, faringe, a boca e as fossas nasais. No ato fonatório, o som foi formado na laringe, pelas vibrações das cordas vocais. Mas para que haja ressonância, é necessário que a freqüência do corpo ressoador seja a mesma do corpo vibrador.

Os ventrículos estão localizados na região supraglótica. Eles são duas pequenas cavidades uma de cada lado da laringe, entre o par de pregas superior e inferior. Estas pequenas cavidades por vezes podem ter o papel de ressoadores.

A faringe é a parte profunda da boca e constitui-se como um tubo com direção vertical, tendo na parte inferior ligação com o aparelho respiratório, pela laringe e com o aparelho digestivo, pelo esôfago. O controle entre o digestivo e o respiratório é realizado pela epiglote, que é situada na parte superior da laringe.

A cavidade bucal está separada da cavidade faríngea pelos movimentos da língua, auxiliada pelos órgãos que limitam a faringe: véu palatino e úvula. É formada por duas porções ósseas, chamadas maxilares superior e inferior, revestidas por paredes elásticas e móveis, as bochechas e na frente, os lábios. Os graus de abertura da boca articulados com os movimentos da língua, do véu palatino e lábios possibilitam as mudanças constantes desta cavidade irregular, provocando assim ressonâncias distintas. Caso a corrente de ar ressoe apenas na cavidade bucal, teremos os sons orais. Caso ressoe na boca e fossas nasais, temos os sons nasais.

As cavidades nasais situam-se acima da cavidade bucal, tendo comunicação com o tubo faríngeo pela cavidade rino-faríngea. Este orifício está aberto ou fechado dependendo dos movimentos dos músculos do véu palatino que levantam ou abaixam a úvula.

**O Articulador:** compreende os órgãos situados na boca e que são responsáveis pela modificação do som, são eles: maxilar inferior e dentes, língua, lábios, úvula, véu palatino, maxilar superior e dentes, alvéolos e palato.

Dependendo dos movimentos de abertura e fechamento da cavidade bucal, os sons do Português se classificam como: abertos, ligeiríssimamente fechados, ligeiramente fechados, menor fechamento, maior fechamento.

A língua é um órgão musculoso e muito ágil, situado no espaço parabólico que as arcadas dentárias descrevem. É o órgão articulador principal para a fonação.

Os lábios são pregas músculo-membranosas situadas na parte anterior da boca em número de dois: superior e inferior. Como articuladores os lábios são responsáveis pelos sons labiais do Português.

O maxilar superior é um osso muito resistente e fixo, atravessado por um grande número de nervos e apresenta, em sua borda livre, os dentes superiores, inseridos em seus respectivos alvéolos e gengivas.

Os dentes superiores formam regiões de articulação dos sons, classificando-os como dentais, labiodentais e dentais alveolares. Os dentes incisivos formam a região do articulador, os demais dentes superiores servem de ponto de apoio para as bordas laterais da língua durante a articulação de vários outros sons.

O véu palatino é um conjunto músculo-membranoso que na parte anterior está ligado ao pós-palato e na posterior é inclinado e termina verticalmente numa porção chamada úvula. A região do véu palatino é móvel e contrai-se em movimentos para baixo e para cima. Quando abaixa ele entra em contato com o pós-dorso da língua e intercepta a comunicação da cavidade bucal com a faríngea.

A úvula é o prolongamento vertical do bordo posterior do véu palatino. Ela articula a vibrante múltipla classificada como uvular.

### **2.1.1 Descrição de Patologias**

Existem várias patologias vocais e serão descritas abaixo algumas delas:

**Nódulo em Pregas Vocais:** É um processo inflamatório localizado na túnica mucosa da parte membranácea da prega vocal, entre seu terço anterior e médio, geralmente bilaterais, conseqüentes ao abuso vocal. Ocorre predominantemente em meninos, até os 14 anos, e em adultos principalmente do sexo feminino. (HUNGRIA, 1960) (ALBERNAZ, 1963)

**Laringite Aguda:** os processos inflamatórios agudos da laringe traduzem-se por sintomatologia decorrente de distúrbios das funções laríngeas, isto é, perturbações da fonação (disfonia discreta até a afonia), da respiração (simples dispnéia até a asfixia) e dos reflexos naturais de defesa. (HUNGRIA, 1960) (ALBERNAZ, 1963)

**Laringite Crônica:** caracteriza-se pela rouquidão permanente, em maior ou menor intensidade, geralmente acompanhada de expectoração muco-catarral, principalmente pela manhã, obrigando o paciente a tossir e a rapar a garganta. As causas de laringite crônica mais comuns são: surtos freqüentes de laringite aguda, uso exagerado de tabaco e álcool, inalação de substâncias irritantes nos ambientes profissionais, exposição continuada ao calor ou ao frio. (HUNGRIA, 1960) (ALBERNAZ, 1963)

**Pólipo:** são geralmente tumores benignos únicos localizados ao nível da borda livre, ou mais raramente, nas faces superior e inferior das pregas vocais. A formação dos pólipos é secundária a laringites agudas ou, mais freqüentemente, a laringites crônicas. A localização mais comum na borda livre das cordas vocais, impossibilitando a coaptação das mesmas, determina o sintoma mais freqüente do pólipo laríngeo, a rouquidão. (HUNGRIA, 1960) (ALBERNAZ, 1963)

**Papiloma:** são tumores benignos que se caracterizam por acentuada tendência recidivante. Apresentam cor branco-acinzentado ou róseo, com aspecto vegetante, tipo couve-flor, séssil ou pediculado, implantado ora na corda vocal, ora na comissura anterior, ora na faixa ventricular e, mais raramente, na epiglote e prega aríteno-epiglótica. A sintomatologia traduz-se por rouquidão e dificuldade respiratória, esta última excepcional no adulto e freqüente na criança. (HUNGRIA, 1960) (ALBERNAZ, 1963)

**Câncer de Laringe:** a localização mais freqüente do câncer laríngeo é ao nível do terço médio ou do terço anterior da prega vocal. O primeiro sintoma a surgir é a rouquidão, que é duradoura. (HUNGRIA, 1960) (ALBERNAZ, 1963)

## **2.2 Descrição e Caracterização de Sinais de Voz**

Nesta seção os sinais de voz serão descritos, não sob aspectos clínicos ou anatômicos, mas sim sob uma visão matemática e como um processo físico. Este enfoque interdisciplinar possibilita a compreensão das técnicas e artifícios utilizados para extrair as informações clínicas desejadas.

### **2.2.1 Caracterização da Voz**

#### **2.2.1.1 Informação Lingüística**

A voz humana é um sistema complexo, ela guarda várias informações que são necessárias para o entendimento da fala, como por exemplo: emoção, acentuação, tom da fala, timbre (informação individual – de quem está falando), etc.

A fala é transmitida através de sentenças que são construídas utilizando unidades básicas chamadas de palavras. Cada palavra contém sílabas e cada sílaba, fonemas, que podem ser decompostos em vogais e consoantes.

#### **2.2.1.2 Mecanismos de Produção de Voz**

Quando o ser humano fala, ele faz uma determinada pressão sobre os pulmões que expulsam o ar. O ar então passa pela traquéia chegando na glote. A glote estreita-se e as cordas vocais vibram com a passagem do som em uma determinada frequência. No homem, a frequência de vibração das cordas vocais é de 110 Hz a 170 Hz enquanto que na mulher é frequentemente de 180 Hz a 280 Hz e nas crianças de 220 Hz a 320 Hz. (DIAS, 1960)

O ar então prossegue nas cavidades refletindo nas paredes e saindo pelos lábios como uma forma de onda de som analógica.

Os órgãos articuladores são usados pelo homem para manipular o som. A língua, os lábios, os dentes, são usados para mudar o formato da cavidade oral.

O som produzido pode ser vozeado ou não-vozeado. Durante uma fala vozeada, as cordas vocais vibram em uma frequência conhecida como frequência fundamental, mais conhecido como *pitch*. Os sons não-vozeados são criados quando não ocorre vibração das cordas vocais. O ar então passa pelo trato vocal sem ocorrer a vibração das cordas vocais ou então ocorre uma turbulência devido a alguma constrição ocasionada pela língua, dentes, lábios ou glote.

Os sons vozeados podem ser demonstrados pela produção de vogais ou consoantes – conhecido como consoantes vozeadas, enquanto que os sons não-vozeados pela produção exclusivamente de consoantes. Existem dois tipos de consoantes: as fricativas e as plosivas. As fricativas (/s/,/f/) são produzidas por um fluxo turbulento que ocorre quando o ar passa por uma constrição no trato vocal. As plosivas (/p/, /t/, /k/) são sons impulsivos que ocorrem quando há uma mudança repentina de pressão produzida por retenção e liberação de ar.

Um modelo de produção de voz pode ser visto na Figura 2-5. Conforme modelo, um sinal de voz é produzido excitando o trato vocal, ou seja, as fontes de excitação e a forma do trato vocal são independentes (*Impulse from Generator* e *Random Number Generator*). Outro fator que afeta a fala é a amplitude e o tom da fala (*pitch*).

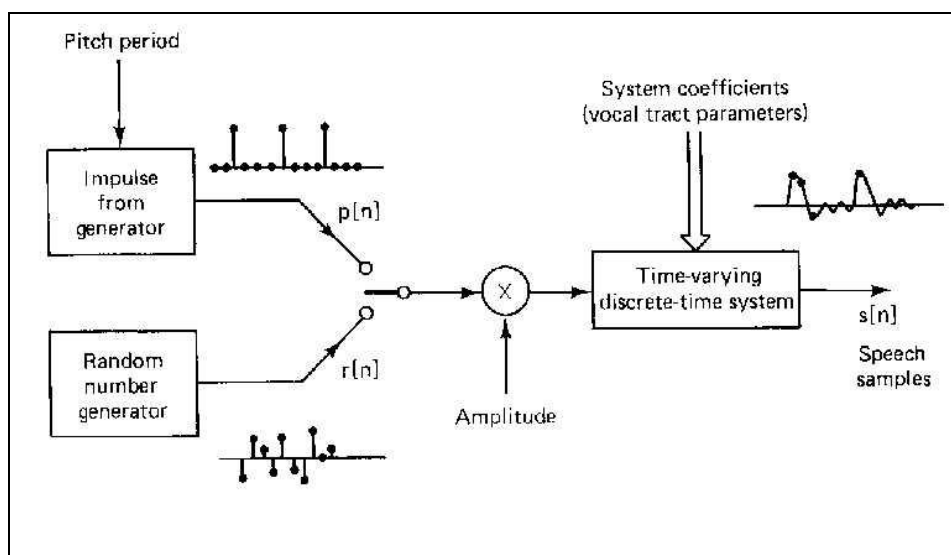
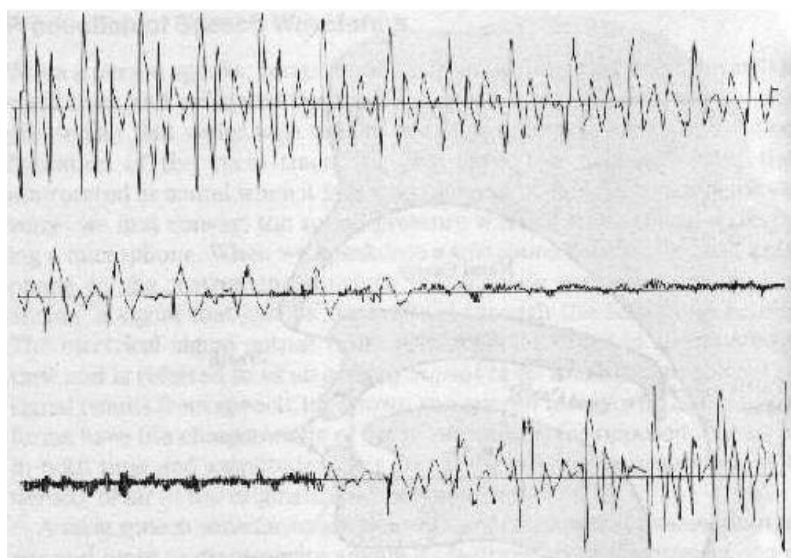


Figura 2-5 Modelo de produção de voz. Extraído de (OPPENHEIM, 1989)

### 2.2.1.3 Características Acústicas da Voz

O som produzido por uma fala pode ser visto como um sinal analógico contínuo. Observando um trecho de fala (Figura 2-6), podemos ver que existem componentes vozeados e não-vozeados. Durante o período vozeado a forma da onda consiste de ciclos que parecem cópias exatas dos vizinhos, apresentando ondas periódicas de amplitudes elevadas. Durante o período não-vozeado a fala se apresenta como ondas aleatórias de pequena amplitude. (PELTON, 1992)



**Figura 2-6 Exemplos de sons-vozeados e não-vozeados. Extraído de (PELTON, 1992)**

Ao se falar, o ar sobre pressão dos pulmões passa pelo trato vocal e entra em contato com várias obstruções: língua, dentes, lábios. Algumas de suas energias são absorvidas por essas obstruções, muitas são refletidas pelas paredes. As reflexões ocorrem em todas as direções.

Algumas ondas ressonam dentro do trato vocal de acordo com sua frequência e com o formato do trato vocal do momento – este fenômeno é conhecido como formante. Uma onda de uma fala vozeada contém inúmeras formantes mas somente as primeiras são importantes para a inteligibilidade. As frequências das formantes variam de pessoa para pessoa de acordo com o tamanho da cabeça e outros fatores. (PELTON, 1992)

#### 2.2.1.4 Características Estatísticas da Voz

A análise do espectro de voz de homens e mulheres mostra que não existem diferenças significativas, exceto em baixas frequências (100Hz – 300Hz). Para os homens o espectro se mostra aproximadamente plano até 500Hz, baixando 10 dB (decibéis) por oitava a partir desse ponto. Para as mulheres a curva é aproximadamente a mesma, mas com uma queda de 20 dB por oitava para frequências menores que 200 Hz.

#### 2.2.2 Modelos de Produção de Voz

Muitas foram as tentativas de se produzir artificialmente a voz humana, também conhecida como síntese da voz. Entre essas tentativas, alguns se destacaram e entre eles podemos citar o trabalho de um húngaro que apresentou em 1791 uma máquina capaz de gerar todas as vogais e a maior parte das consoantes do alemão e latim, bem como pronunciar frases curtas nessas línguas. (ROCHA, 1997)

O modelo dessa máquina pode ser visto na Figura 2-7.

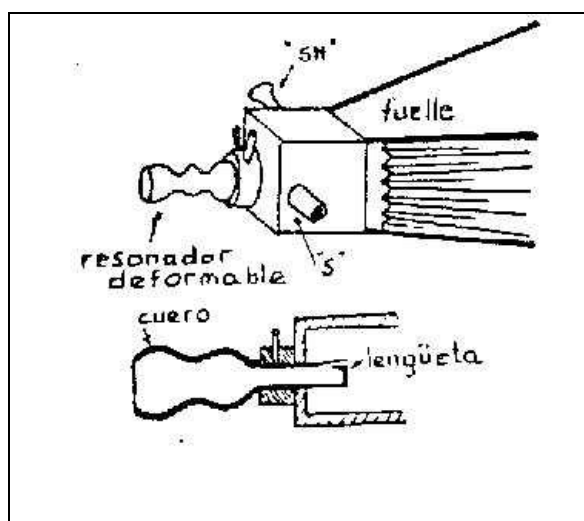


Figura 2-7 Modelo da máquina húngara- extraído de (ROCHA, 1997)

A máquina utilizava basicamente um fole para impulsionar o ar através da língua vibratória que ao mover-se produzia um som semelhante às cordas vocais. A mão direita do operador descansava sobre a palanca de fole enquanto que a esquerda ficava em uma câmara ressonante que simulava uma cavidade bucal. (ROCHA, 1997)

## 2.3 Aquisição de Sinais de Voz

### 2.3.1 Conversão Analógico-Digital

Os sinais analógicos resultantes da fala humana precisam ser convertidos para uma forma que o computador os “entenda”, pois os computadores só trabalham em modo binário. Essa conversão precisa ser feita de maneira que não ocorram perdas significativas de informação no sinal.

#### 2.3.1.1 Amostragem de Sinais Contínuos

A técnica de amostragem de sinais contínuos permite que o sinal analógico seja convertido para uma forma discreta ao longo do tempo. (Figura 2-8).

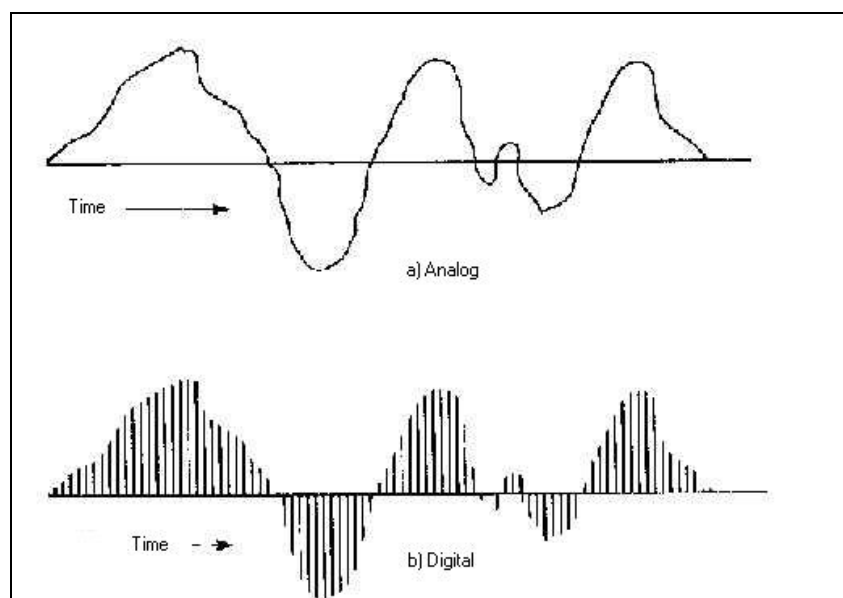


Figura 2-8 Sinal analógico no domínio tempo (a)- Sinal discreto no domínio tempo (b). Extraído de (PELTON,1992)



Um método típico de obter uma representação discreta a partir de um sinal contínuo no tempo é a amostragem periódica, dada pela relação:

$$x[n] = x_c(nT), \quad -\infty < n < \infty \quad \text{Eq. 2.3.1}$$

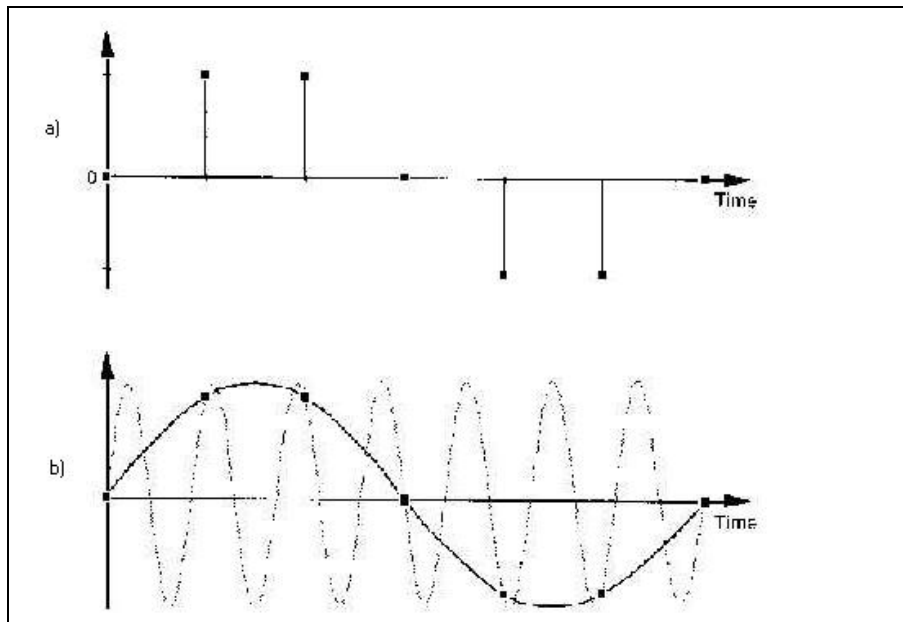
Onde:  $T$  é o período da amostragem.

A frequência de amostragem então é dada por:

$$f_s = 1 / T \quad \text{Eq. 2.3.2}$$

Na prática a operação de amostragem é geralmente implementada por um dispositivo denominado conversor analógico-digital (A/D).

Mas um problema surge quando se quer o inverso, ou seja, tem-se o sinal discreto e se quer o sinal contínuo no tempo. Como pode ser visto na Figura 2-9, pelos pontos do sinal discreto podem passar um ou mais sinais com diferentes frequências, mas qual desses sinais é realmente o sinal original? Esse problema é conhecido como *aliasing*.



**Figura 2-9 Aliasing. a) Sequência de valores discretos no tempo. b) Duas senóides diferentes que passam pelos mesmos pontos da sequência discreta. Extraído de (LYONS, 1999)**

A solução para esse problema é o Teorema de Nyquist, que determina que para sinais limitados em banda (passa-baixas) a taxa de amostragem deve ser no mínimo duas vezes a maior frequência presente no sinal analógico. Por exemplo:

Um sinal contínuo que possua componentes até 200 Hz deve ser amostrado utilizando-se uma taxa de pelo menos 400 Hz para que se possa reconstruí-lo corretamente.

$$\text{Teorema de Nyquist: } f_{\text{max}} = \frac{1}{2} f_s$$

onde:  $f_s$  é a máxima componente em frequência do sinal

### 2.3.1.2 Quantização

O processo de quantização corresponde à discretização da amplitude do sinal. A informação contínua é aproximada a um nível finito correspondente. Isto é necessário, da mesma forma que na operação de amostragem no tempo, descrita anteriormente, devido ao fato dos computadores possuírem memória finita.

Um exemplo de um sinal quantizado é apresentado na Figura 2-10.

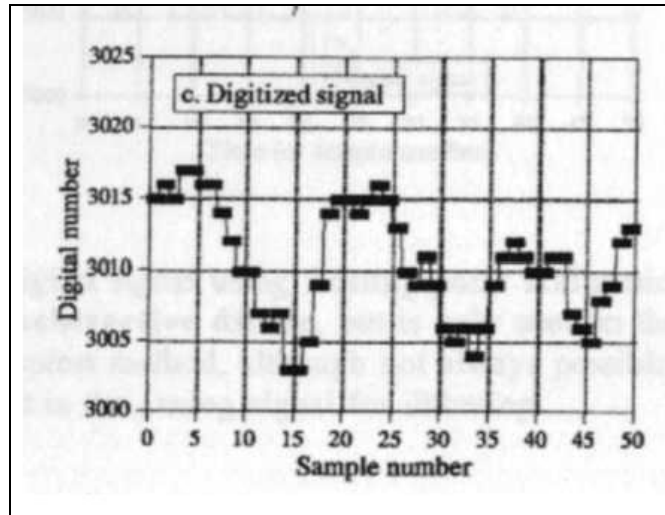


Figura 2-10 Exemplo de um sinal digitalizado (quantização). Extraído de (SMITH,1999)

Ou seja, quanto maior o número de níveis de quantização (*bits*) usado para a quantização, mais próximo do original o sinal ficará.

### 2.3.2 Condicionamento do Sinal

A Figura 2-11 mostra um diagrama em blocos de um sistema básico de processamento digital de sinais (DSP – Digital Signal Processing).

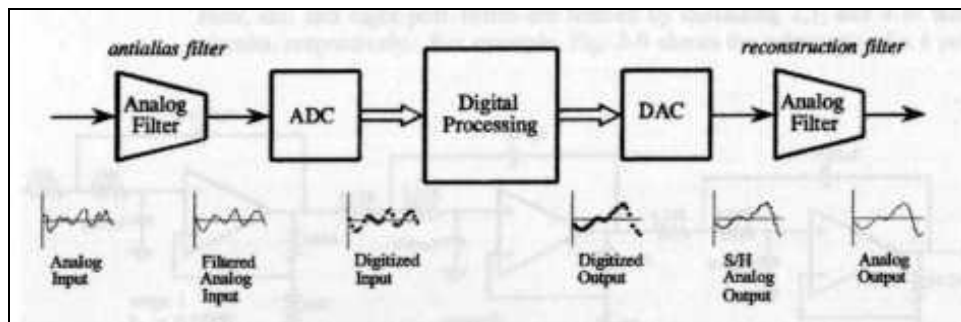


Figura 2-11 Diagrama de um DSP. Extraído de (SMITH, 1999)

Inicialmente, o sinal analógico é processado por um filtro analógico passa-baixas para limitar o sinal em frequência. Isto é feito para prevenir o *aliasing* durante a amostragem, e é chamado de filtro Anti-Alias. A seguir, o sinal passa por um conversor digital-analógico procedendo assim a sua discretização tanto ao longo do tempo como em amplitude. A partir deste momento o sinal de interesse pode ser modificado por um processador digital, alterando, enfatizando ou restringindo suas características através de algum tipo de algoritmo numérico. Após este processo, o sinal contínuo é reconstruído por um conversor digital-analógico e sujeito a um segundo filtro passa-baixas para remover as altas frequências decorrentes da quantização. Este filtro é denominado de Filtro de Reconstrução.

Os tipos de Filtros Analógicos mais usados são: *Chebyshev*, *Butterworth* e *Bessel*. Cada um deles possui distintas características de filtragem. (SMITH, 1999)

## 2.4 Técnicas de Processamento de Sinais Aplicadas à Voz

### 2.4.1 Filtragem Digital

A filtragem digital é uma importante ferramenta para o processamento de sinais. Uma das maneiras pelas quais podemos verificar seus efeitos é através das modificações que ocorrem no conteúdo espectral do sinal após seu processamento. Podemos fazer uma analogia da filtragem com uma peneira ou um filtro de água, ou seja, nós extraímos de um sinal somente o que nos interessa.

A filtragem digital nada mais é do que a aplicação de um algoritmo numérico sobre a seqüência de valores discretos que caracteriza um sinal amostrado. Os procedimentos matemáticos relacionados ao algoritmo numérico podem ser realizados através de um programa de computador (*software*) ou através de realizações em *hardware* implementadas em um circuito integrado.

Os filtros podem ser classificados, basicamente, quanto à sua influência no conteúdo espectral do sinal resultante relacionado ao sinal original. Portanto, usualmente costuma-se definir quatro classificações: passa-baixas, passa-altas, passa-faixa e rejeita faixa. O filtro passa-baixas atenua as frequências altas e passa as frequências baixas. Enquanto que o filtro passa-altas é o inverso, ou seja, atenua as frequências baixas e

passa as frequências altas. Da mesma forma, os tipos passa-faixa e rejeita faixa enfatizam ou atenuam determinadas bandas de frequência do sinal de interesse.

### 2.4.2 Análise de Fourier

A análise de Fourier também é uma importante ferramenta utilizada em processamento de sinais. Esta técnica permite a representação de um sinal em função de suas componentes espectrais. Sua principal vantagem é permitir que comportamentos oscilatórios periódicos possam ser facilmente determinados e quantificados. Basicamente, a transformada de Fourier expressa o sinal de interesse em termos de sinusóides de diferentes frequências, explicitando suas amplitudes e fases. (OPPENHEIM, 1989)

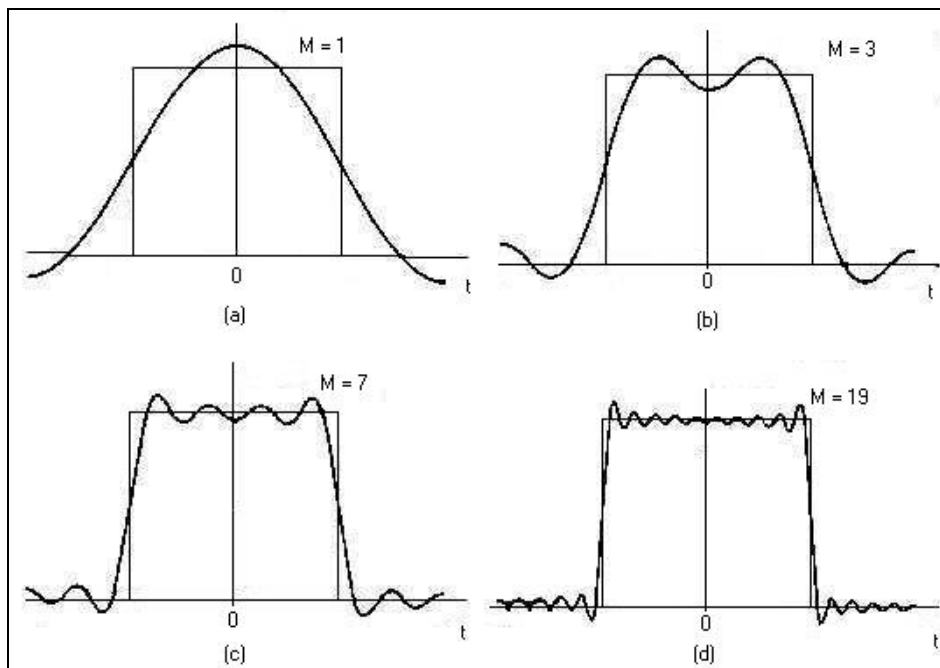


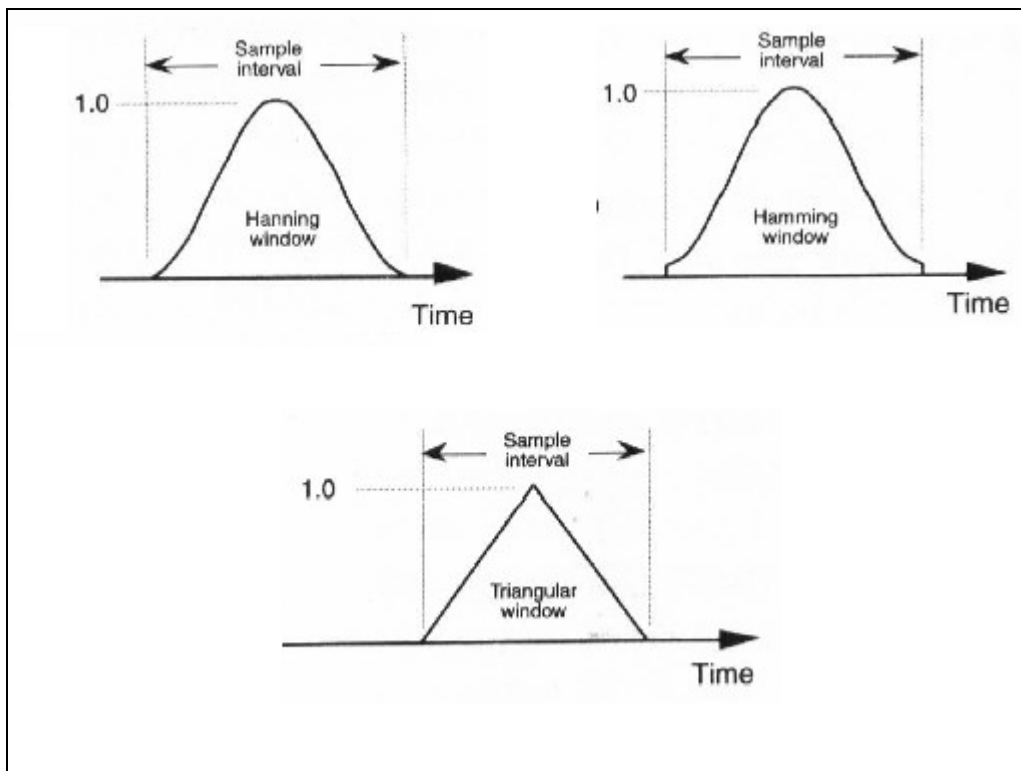
Figura 2-12 Exemplo de Transformada de Fourier. Extraído de (OPPENHEIM, 1989)

O sinal original da Figura 2-12 é quadrado, e no primeiro gráfico observamos uma aproximação utilizando apenas uma senoide ( $M=1$ ). À medida em que se aumenta o número de sinusóides, o sinal fica cada vez mais próximo do sinal original, como podemos ver no último gráfico.

Para sinais discretos, é utilizada a Transformada Discreta de Fourier (DFT – *discrete Fourier transform*), cuja equação é dada por:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j \frac{2\pi}{N} k \cdot n}, k = 0, 1, \dots, N - 1$$

A DFT aplica-se a sinais periódicos no tempo. Quando se utiliza sinais não periódicos ou intervalos de tempo que não apresentem múltiplos inteiros de seu período, ocorre o efeito que chamamos de espalhamento (*leakage*). Ou seja, ocorre uma deformação do espectro do sinal, através de erros na estimação das amplitudes e fases das sinusóides que o compõe. Para diminuir este problema, utiliza-se uma técnica chamada de Janelamento (*Windowing*). (OPPENHEIM, 1989) O sinal de interesse é multiplicado por um outro sinal cuja amplitude decai ao longo do tempo. Existem vários tipos de janelas, podemos citar entre : hanning, hamming, e triangular. Figura 2-13



**Figura 2-13 Exemplos de janelamento**

Quando se tem um sinal no domínio frequência e se deseja retornar ao domínio tempo, é usada a Transformada Discreta de Fourier Inversa (IDFT – Inverse Discrete Fourier Transform), cuja equação é a seguinte:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot e^{+j\frac{2\pi}{N} \cdot k \cdot n}; n = 0, 1, \dots, N - 1$$

A DFT é um procedimento que exige um esforço computacional bastante grande e por isso seu uso sempre foi limitado em implementações práticas. Entretanto, a partir do trabalho de Cooley e Tukey (1965), que introduziram os algoritmos chamados de Transformada Rápida de Fourier (FFT - Fast Fourier Transform), esta técnica de estimação espectral passou a ser considerada quase que um padrão. A grande vantagem dos algoritmos de FFT é que estes são métodos otimizados para o cálculo exato da Transformada Discreta de Fourier com um número bem menor de operações aritméticas.

Existem hoje muitos algoritmos mas podemos classificar os métodos em 2 grandes grupos:

- a) Métodos de sub-dividir a DFT em DFTs menores, também chamado de "Método de dividir para conquistar";

Por sua vez, o grupo “a” pode ser subdividido em 2 sub-grupos:

- Decimação no tempo (DIT) ;
  - Decimação na frequência (DIF).
- b) Métodos baseados em filtragem linear, também denominados de métodos convolutivos.

### 2.4.3 Filtragem Homomórfica

As técnicas de filtragem e processamento anteriormente apresentadas podem ser vistas basicamente como operações matemáticas lineares aplicadas sobre os sinais de interesse. Entretanto, na natureza, diversas informações (sinais) apresentam-se misturadas de forma não-linear com ruídos ou fenômenos associados. Entre as diversas

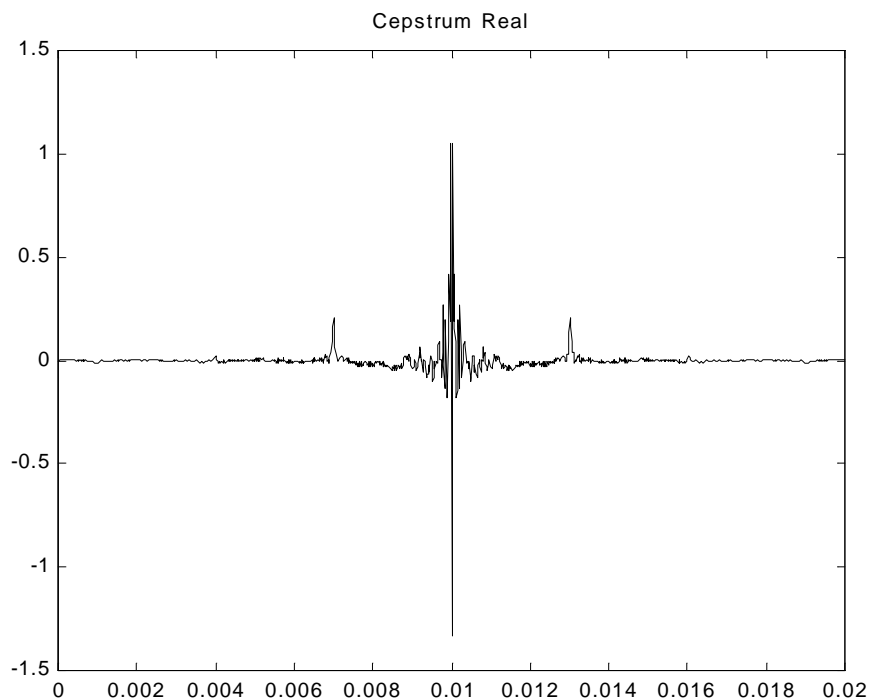
formas de “misturas” não-lineares podemos citar a modulação e a multiplicação de sinais. Nestes casos específicos, a filtragem homomórfica apresenta-se como uma interessante ferramenta a ser explorada.

Para explicar o funcionamento da filtragem homomórfica, é preciso definir o Cepstrum.

Em 1983, Bogert, Healy, e Tukey observaram que o logaritmo do espectro de potência de um sinal contendo um eco possuía um componente periódico a mais, devido ao eco, e assim a Transformada de Fourier do logaritmo do espectro de potência apresenta um pico. Eles chamaram esta função de Cepstrum. (OPPENHEIM, 1989)

O Cepstrum nada mais é do que a inversa do logaritmo da FFT, ou seja:

Sinal  $\rightarrow$  FFT  $\rightarrow$  LOG  $\rightarrow$  IFFT

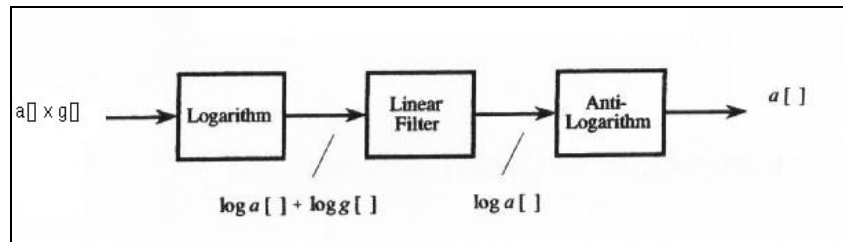


**Figura 2-14 Cepstrum de um sinal de fala**

Podemos ver na Figura 2-14 um exemplo de cepstrum de um trecho de 20ms de fala. Os picos menores correspondem ao componente vozeado.



Na análise da voz humana no domínio “cepstral”, o trecho central, entre 0,008 e 0,012 segundos, representa as características do aparelho fonador, enquanto que os trechos laterais representam o sinal de excitação produzido pelos pulmões (veja Figura 2-14). Removendo-se, portanto, esta última informação através de técnicas de filtragem, pode-se obter um sinal representativo das condições funcionais do trato vocal. Podemos ver um diagrama de como é feita a filtragem homomórfica na Figura 2-15.



**Figura 2-15 Filtragem homomórfica**

### 3 Análise dos Sinais de Voz

#### 3.1 Aquisição dos Sinais de Voz

##### 3.1.1 Descrição do *Hardware* utilizado

Para fazer a aquisição dos sinais de voz dos pacientes, foi utilizado um microfone Sony Cardioide modelo F-V07T, e uma Placa de Som 16 bit Creative Sound Blaster AWE64.

##### 3.1.2 Descrição da Metodologia de Aquisição

Diversos sinais de voz de pacientes foram adquiridos na Clínica Castagno, em Pelotas. Foram obtidos 5 sinais de vozes de pacientes, sendo 3 homens e 2 mulheres. Como a maior parte deles ainda está em tratamento, neste trabalho foram utilizados sinais de cinco pacientes obtidos a partir do banco de dados do serviço de Engenharia Biomédica do Hospital de Clínicas de Porto Alegre. Todos os casos foram devidamente diagnosticados por um médico e são pessoas do sexo feminino.

Os pacientes com as respectivas patologias estão na tabela abaixo:

Paciente	Patologia
1	Edema de Prega Vocal - é um acúmulo de líquido nos tecidos afetados, decorrentes de inúmeras causas, entre as quais: irritação, inflamação e tumor.
2	Nódulo em Prega Vocal - é um aumento de volume localizado nas pregas vocais, geralmente de origem tumoral benigna.
3	Cisto - é um tumor geralmente benigno composto por uma cápsula e de conteúdo líquido.
4	Hiperemia de Prega Vocal - é um aumento da circulação capilar, decorrente de inúmeras causas, entre elas: irritação, tumor e principalmente processos inflamatórios agudos.
5	Hiperemia de Prega Vocal

Os sinais das vozes dos pacientes foram obtidos a partir de uma taxa de aquisição de 44100Hz e de uma quantização de 16 bits. Foi pedido para os pacientes pronunciarem a vogal “e” aberta (/ɛ/) durante mais ou menos 4 segundos a uma distância de mais ou menos 5cm do microfone. Os sinais foram armazenados no computador no formato de arquivo WAV não comprimido.

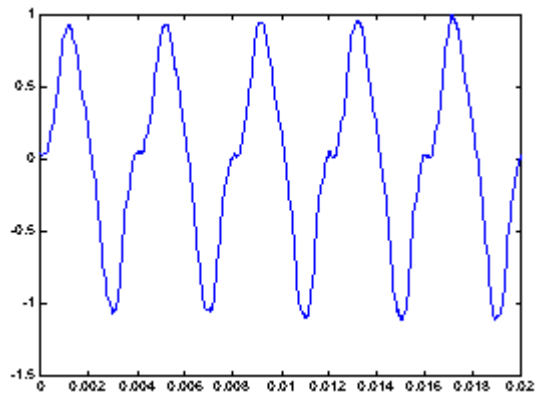
### **3.1.3 Plataforma da Análise dos Sinais**

Os sinais de voz foram analisados utilizando o *software* Matlab. Este é um programa matemático que integra programação, visualização e computação e que possui várias funções próprias para o processamento de sinais digitais.

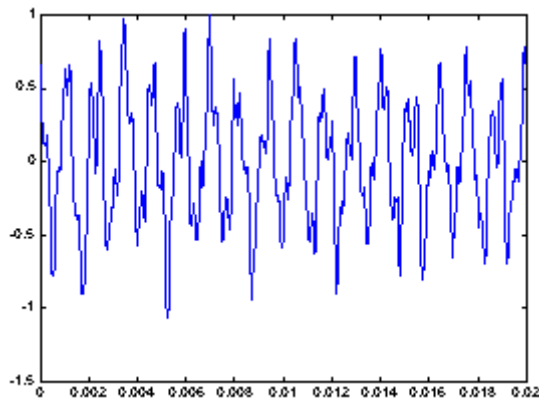
## **3.2 Considerações da Análise**

Para realizar a análise dos sinais, foi escolhido um trecho de 1s de cada sinal de voz dos pacientes que se mostrava mais constante, com menos variações de amplitude. Nesse um segundo, foram separados 50 trechos de 20ms cada um (pois em 20ms o sinal de voz é considerado sem variações estatísticas, ou seja, aproximadamente estacionário), processados individualmente e então realizada a média sobre o resultado das operações sobre estes trechos. A análise foi realizada na média dos 50 trechos. Todos os sinais foram normalizados pelo seu valor máximo.

Observando os sinais de voz no domínio tempo, podemos verificar que após o tratamento o sinal apresenta componentes de maior frequência. Isto pôde ser observado mais claramente nos pacientes 1,2,3,4. O gráfico do paciente 5 se mostrou semelhante antes e após o tratamento. A Figura 3-1 e Figura 3-2 mostra essa observação.



**Figura 3-1 Paciente 1 antes do tratamento**



**Figura 3-2 Paciente 1 após o tratamento**

### 3.3 Análise do Espectro de Voz

A análise foi realizada utilizando a técnica de periodograma, com sobreposição (*overlap*) de 50% e utilizada a janela de Hanning. Os resultados obtidos foram normalizados também pelo valor de pico.

O espectro de Voz dos Pacientes foi realizado utilizando-se a seguinte equação:

$$S_x = 20 \cdot \log(|FFT(x)|)$$

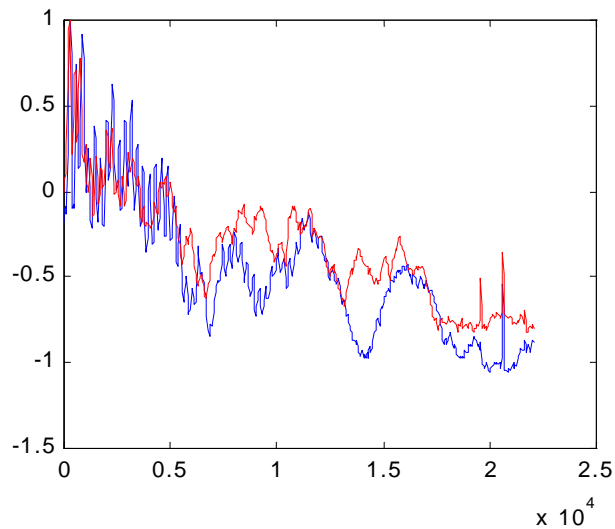
Onde:

x = sinal original

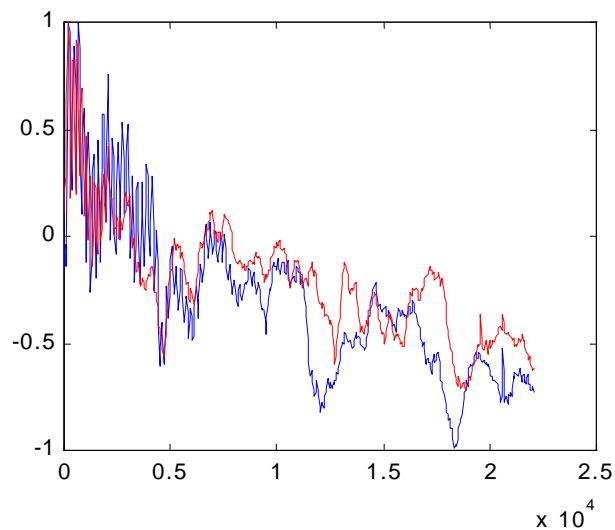
Os sinais foram normalizados também pelo máximo pico.

Após realizar a análise dos sinais dos 5 pacientes, foi verificado que os sinais antes e após o tratamento apresentam características muito semelhantes, portanto, não é possível fazer distinção entre eles.

Os gráficos abaixo mostram o espectro de potência para os pacientes 1 e 2. O eixo horizontal é a frequência e o vertical o módulo (em decibéis).



**Figura 3-3 Espectro de voz (Paciente 1). Vermelho = sinal antes do tratamento; Azul = sinal após o tratamento**



**Figura 3-4 Espectro de voz (Paciente 2). Vermelho = sinal antes do tratamento; Azul = sinal após o tratamento**

Os gráficos dos demais pacientes podem ser encontrados no Anexo A.

### 3.4 Análise do Cepstrum Real

A análise do cepstrum real se mostrou bastante eficaz. Notamos na maioria dos pacientes (1,2,3,4) um aumento considerável da componente vozeada (que se encontra, normalmente, entre 4ms e 6ms) antes e após o tratamento. Já no paciente 5 esse aumento não foi verificado.

Os eixos horizontais estão no domínio tempo e os verticais são a amplitude.

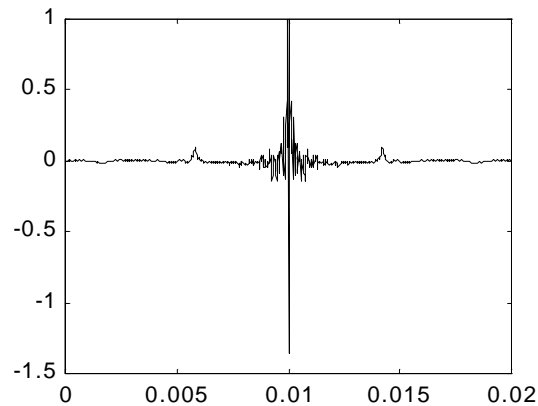


Figura 3-5 Paciente 2 antes do tratamento

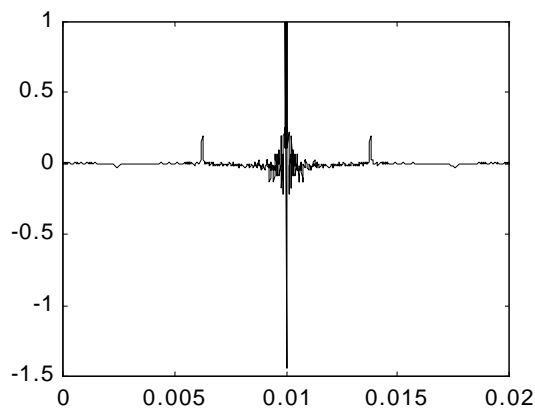


Figura 3-6 Paciente 2 após o tratamento

Os gráficos dos demais pacientes podem ser encontrados no Anexo B.

### 3.5 Análise das Formantes Utilizando o Cepstrum Real

A análise das formantes foi realizada através da filtragem do cepstrum real com uma janela de 100 amostras, para filtrar a excitação do trato vocal.

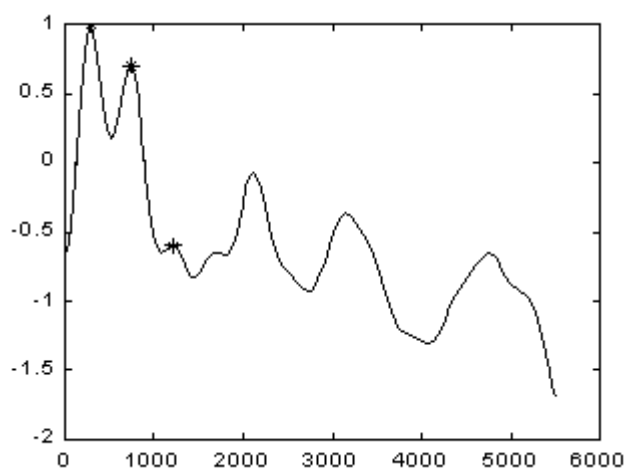
Foi calculada a média e o desvio padrão das formantes em cada um dos 5 pacientes, e obtidos os seguintes resultados:

Formante	Paciente	Antes do Tratamento		Após o Tratamento	
		Média	Desvio Padrão	Média	Desvio Padrão
1	1	207	10.81	216	3.06
1	2	212	8.76	217	3.06
1	3	214	6.57	218	5.20
1	4	201	9.84	217	0
1	5	218	5.94	217	0
2	1	694	11.16	746	18.07
2	2	696	150.29	673	10.27
2	3	752	175.01	728	123
2	4	678	16.29	678	12.25
2	5	703	121.66	696	9.03
3	1	1305	159.64	1372	88.61
3	2	1546	247.37	1197	136.25
3	3	1638	336.12	1222	137
3	4	1446	213.9	1328	230.15
3	5	1614	244.25	1224	21.54

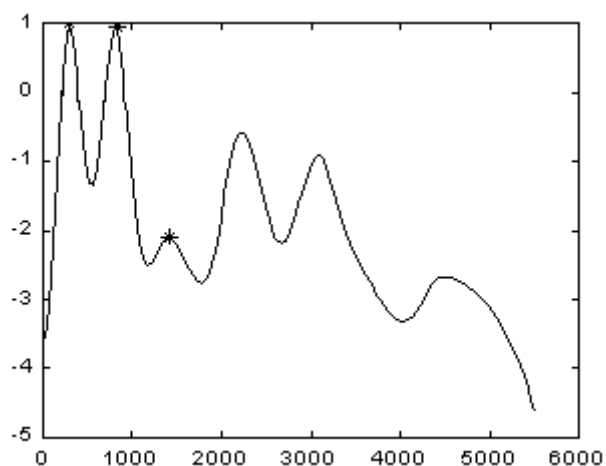
Esta tabela nos mostra que existe um pequeno aumento da frequência das formantes após o tratamento, mas não foi possível qualquer identificação através da análise de formantes das patologias analisadas, pois ocorre uma variação aparentemente mínima em seus valores.

Esse resultado não é o que se esperava da análise, pois as formantes nada mais são do que uma ressonância do sinal de voz no trato vocal, ou seja, uma amplificação do sinal da voz seletivamente à frequência. Se o paciente apresenta o trato vocal rígido, teoricamente as estruturas do trato vocal deveriam sofrer algum tipo de alteração e isso deveria aparecer no gráfico. Por outro lado sabe-se que as 3 primeiras formantes estão associadas à inteligibilidade da fala e portanto, formantes de ordem mais elevada poderiam fornecer melhores informações.

Abaixo está o gráfico da análise de formantes do paciente 1 antes e depois do tratamento. O eixo horizontal é a frequência e o vertical é a amplitude.



**Figura 3-7 Paciente 1 antes do tratamento**



**Figura 3-8 Paciente 1 após o tratamento**

Os gráficos dos demais pacientes podem ser encontrados no Anexo C.

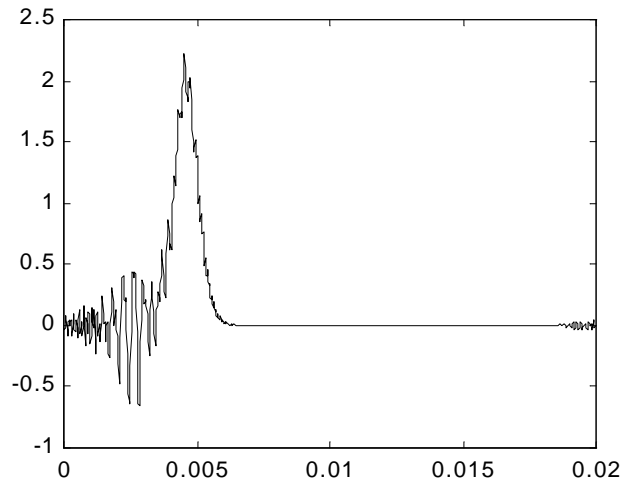
### **3.6 Resposta ao Impulso do Trato Vocal**

A resposta ao Impulso do Trato Vocal foi calculada utilizando um filtro invariante no tempo com uma janela de 20 amostras.

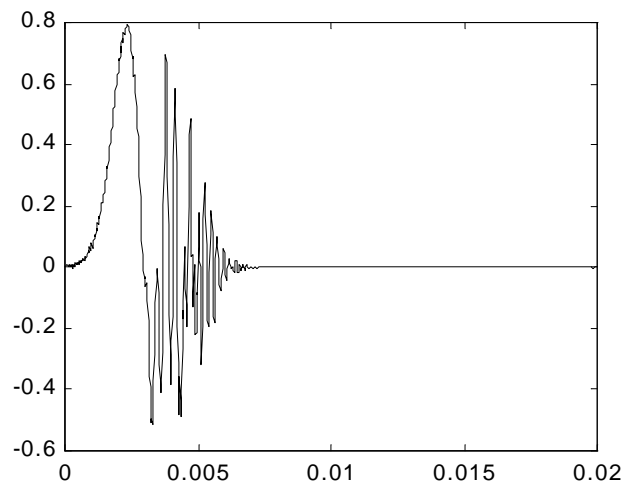
Verificou-se nos pacientes 1, 2 e 4 que após o tratamento o gráfico apresenta uma grande oscilação após o maior pico, o que não acontece nos sinais patológicos. Já no restante dos pacientes, o gráfico não apresenta condições de análise visuais.



A figura abaixo mostra um exemplo do gráfico do paciente 1. O eixo horizontal encontra-se no domínio tempo e o vertical é a amplitude.



**Figura 3-9 Paciente 1 antes do tratamento**



**Figura 3-10 Paciente 1 após o tratamento**

Os gráficos do estante dos pacientes podem ser encontrados no Anexo D.

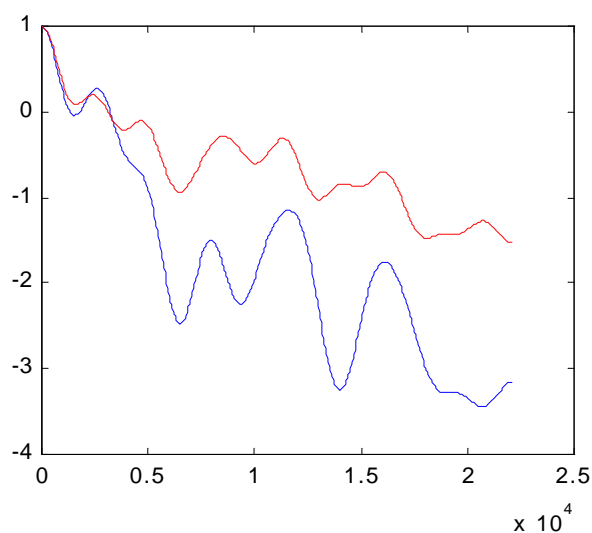
### **3.7 Espectro da Resposta ao Impulso do Trato Vocal**

Podemos observar pelos gráficos, que antes do tratamento o sinal possui menor potência em altas frequências, enquanto que após o tratamento ocorre um aumento.

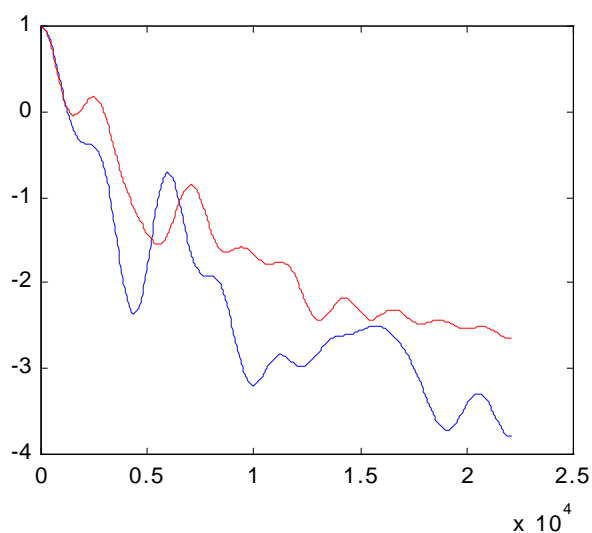
Essa característica já era prevista, pois devido a rigidez do trato vocal antes do tratamento, a capacidade de resposta a altas frequências deve ser diminuída.

O filtro utilizado possui as mesmas características do item anterior.

Os gráficos abaixo mostram o espectro para o paciente 1 e 5, respectivamente. O eixo horizontal é a frequência e o vertical o módulo (em decibéis).



**Figura 3-11 Vermelho = Paciente 1 antes do tratamento. Azul = Paciente 1 após o tratamento**



**Figura 3-12 Vermelho = Paciente 5 antes do tratamento. Azul = Paciente 5 após o tratamento**

Os gráficos dos demais pacientes podem ser encontrados no Anexo E.

## 4 Protótipo para um Sistema de Auxílio Clínico ao Diagnóstico de Patologias Vocais

Foi proposto um protótipo para demonstrar como seria um possível *software* para auxiliar o médico a diagnosticar patologias do trato vocal. Esse protótipo foi desenvolvido no ambiente Windows 98, devido ao fato de ser o sistema operacional mais utilizado atualmente.

Sendo um dos objetivos deste trabalho a implementação prática de um protótipo, não se pretende fazer uma especificação formal do processo de desenvolvimento e sim uma descrição sucinta considerando a apresentação do protótipo implementado.

### 4.1 Características da Linguagem para a Aplicação

Para a elaboração do protótipo, foi utilizada a linguagem de programação Delphi versão 4.

Delphi é uma linguagem orientada a objetos e eventos que possui um ambiente de desenvolvimento fácil de usar, com uma grande biblioteca de componentes visuais (VCL – Visual Component Library). Ele trabalha com eventos que dão início a alguma rotina de trabalho.

As características apresentadas pelo ambiente de programação são muitas, dentre elas cita-se:

- a)* a modularidade;
- b)* a facilidade de criar rotinas visuais;
- c)* a variedade de componentes visuais encontrados na VCL
- d)* o grande suporte a banco de dados;
- e)* a disponibilidade do código-fonte da VCL
- f)* a linguagem *Object Pascal*

A base de toda a programação do Delphi é a linguagem *Object Pascal*. Ela contém todo o conceito de orientação a objetos, incluindo encapsulamento, herança e polimorfismo. Algumas extensões foram incluídas para facilitar o uso, tais como o

conceito de propriedade (particular e pública), tipos de informações em modo *run-time* (em execução), gerenciamento de exceções e referências de classes. O resultado de toda essa junção faz com que o *Object Pascal* consiga suportar as facilidades de um baixo nível de programação, tais como:

- Controle e acesso de subclasses do Windows (API)
- Interceptar mensagens de laço do Windows
- Gerar Mensagens semelhantes as do Windows
- Gerar Código puro da linguagem Assembler

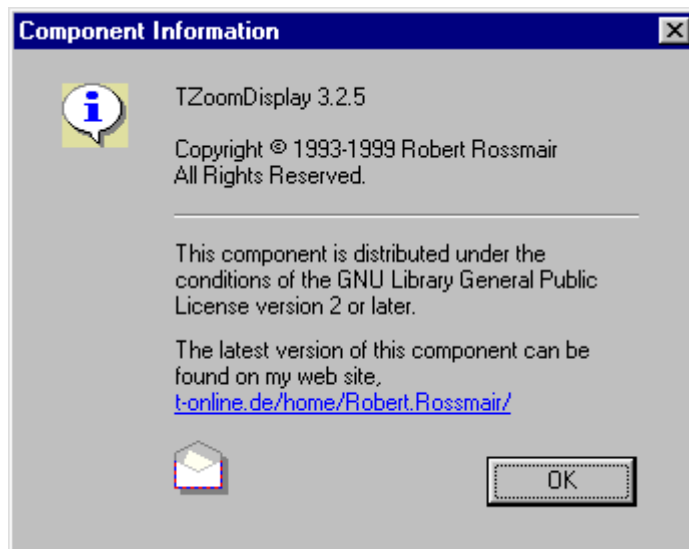
#### **4.2 Bibliotecas Utilizadas**

Foram utilizadas três bibliotecas para elaboração do protótipo, são elas:

- NSP – Intel Signal Processing Library
- ZoomDisplay
- Audio

A biblioteca de Processamento de Sinais da *Intel*, versão 4.3, distribuída gratuitamente contém um conjunto de rotinas para o processamento de sinais otimizados para os processadores *Intel* de última geração que permite processar áudio, vídeo e comunicações de dados utilizando somente software, sem ter que utilizar um hardware específico para DSP. A biblioteca inclui uma DLL que detecta o processador que está sendo utilizado e carrega uma DLL com funções específicas para o processador.

A biblioteca *ZoomDisplay* versão 3.2.5, possui rotinas e componentes visuais para a utilização de gráficos, como fazer *Zoom In* e *Zoom Out* em um sinal, desenhar um sinal resultante de alguma função, como por exemplo a FFT, etc. As informações sobre essa biblioteca podem ser vistas na Figura 4-1.



**Figura 4-1** Biblioteca ZoomDisplay

A biblioteca Audio desenvolvida por Hakan Bergzen permite que os programas desenvolvidos no ambiente Delphi acessem dispositivos de som, como por exemplo, a placa de som SoundBlaster, através do uso de rotinas de baixo-nível da API do Windows.

### **4.3 Características do Usuário**

O usuário é a parte mais importante para a realização do *software*, pois um *software* só tem valor se ele preenche todas as necessidades pretendidas pelo usuário. Por isso, conhecer o usuário é um dos principais pontos que devem ser valorizados na construção de um *software*. Desta maneira, pode-se produzir um sistema que atenda a maioria dos estilos cognitivos, associados a sua utilização.

Este protótipo foi desenvolvido especificamente para um tipo de usuário, o médico, pois se faz necessário o diagnóstico da patologia pelo médico e o *software* o auxiliará no acompanhamento da evolução do tratamento, verificando as condições de melhora do quadro clínico do paciente.

#### 4.4 Fluxograma do Protótipo

A seguir é apresentado o fluxograma do sistema:

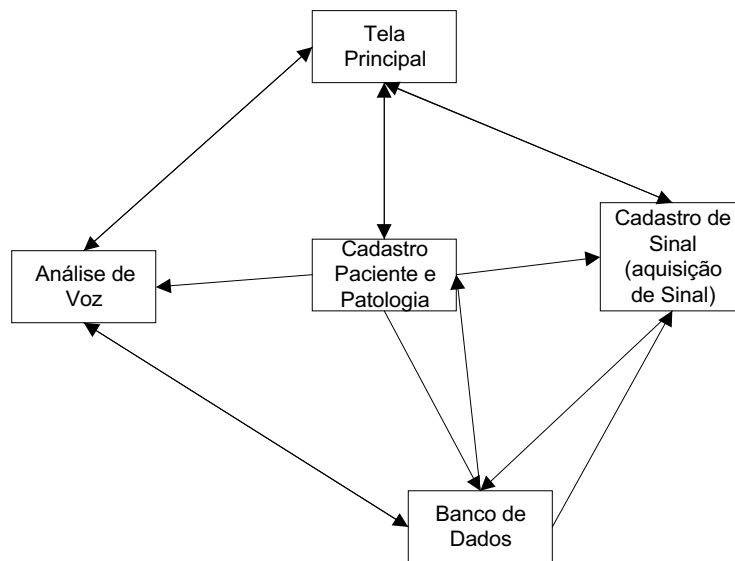


Figura 4-2 Fluxograma do protótipo

#### 4.5 Banco de Dados

Foi utilizado um banco de dados Paradox para armazenar os dados dos pacientes, das patologias e os sinais de vozes dos pacientes.

Abaixo, tem-se um modelo ER do banco de dados utilizado no protótipo:

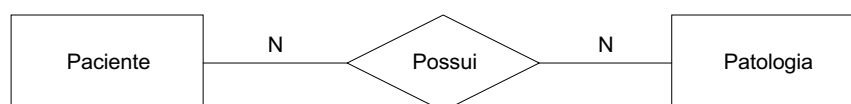


Figura 4-3 Modelo ER do banco de dados

Onde:

<b>Paciente</b>
#CodigoPac
Nome
Endereço
Sexo

E outros dados referentes ao cadastro do Paciente.

Patologia
#codigoPato
Patologia
Descrição

A seguir, a tabela “possui” foi denominada de sinal no protótipo, já que ela irá armazenar os sinais referentes aos dados do paciente e patologia.

Possui
#codigoSinal
CodigoPac
CodigoPato
Data
Hora
Sinal

#### **4.6 Descrição dos Módulos**

Serão explicados os principais módulos criados para a elaboração do protótipo, incluindo rotinas para o desenvolvimento das funções de processamento de sinais digitais e rotinas para a leitura do arquivo WAV.

##### **4.6.1 Módulo para Leitura do Formato de Arquivo .WAV**

Foi criada uma classe para leitura de arquivos em formato .WAV. A biblioteca criada foi limitada a arquivos WAV não comprimidos, gravados em modo Mono, 16bits e com uma taxa de quantização de 44100Hz. Se por acaso o arquivo lido fugir desses valores, a biblioteca acusará um erro.

O arquivo em formato WAV é dividido em blocos (*chunks*). Cada bloco tem um cabeçalho de 8 bytes usado para a identificação do bloco e informando o comprimento em bytes dos dados que seguem o cabeçalho.

Em Delphi, o cabeçalho dos blocos apresenta-se da seguinte forma:

```
// Chunk Header - Para todos os chunks
WChunk_Hdr = record
    id : array[1..4] of char;
    len: DWORD;
end;
```

O formato WAV tem no seu início um *Chunk Header* obrigatoriamente com o ID='RIFF', e logo em seguida um campo com a palavra 'WAVE' Se o arquivo não possuir essas características, ele não é um arquivo WAV válido.

Logo após a identificação, a estrutura mínima de um arquivo WAV é a seguinte:

- *Chunk* Formato
- *Chunk* Data

Esses dois blocos devem aparecer obrigatoriamente nessa ordem.

O Bloco Formato possui as características do sinal gravado, como por exemplo: número de bits, taxa de amostragem, número de canais, etc. No protótipo em Delphi, ficou assim:

```
// Chunk Format do Wav
WFormat = record
    wFormatTag      : WORD;
    wChannels       : WORD;
    dwSamplesPerSec : DWORD;
    dwAvgBytesPerSec : DWORD;
    wBlockAlign    : WORD;
    wBitsPerSample  : WORD;
end;
```

O Bloco Data é um vetor com os dados do sinal. No protótipo em Delphi é o seguinte:

```
wavdata:PsmallIntArray;
```

Onde

```
TSmallIntArray = array[0..0] of SmallInt;
PSmallIntArray  = ^TSmallIntArray;
```

A Estrutura completa do arquivo WAV no protótipo é a seguinte:

```
Inf_Wave = record
    riff_header  : WChunk_Hdr;
    wavid       : array[1..4] of char;
    formato_header : WChunk_Hdr;
    formato     : WFormat;
    data_header  : WChunk_Hdr;
end;
```



A biblioteca completa para a leitura de arquivos em formato .WAV está no anexo F.

#### 4.6.2 Rotinas para o Processamento de Sinais Digitais

Foram criadas rotinas para o desenvolvimento de funções como Espectro de Potência, Cepstrum Real e Análise de Formantes. Essas rotinas utilizam a biblioteca da Intel para o Processamento de Sinais Digitais e podem ser encontradas no Anexo G.

#### 4.7 Descrição da Funcionalidade do *Software*

A tela de apresentação do protótipo é apresentada na Figura 4-4

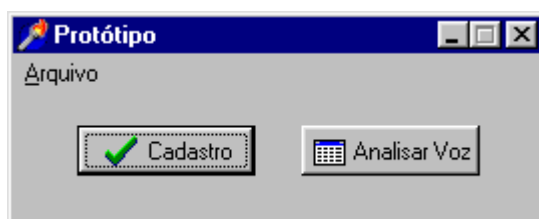


Figura 4-4 Tela inicial do protótipo

O usuário tem duas opções, o cadastro de pacientes, de patologias e de sinal ou então fazer a análise da voz já armazenada no Banco de Dados.

Caso o usuário deseje entrar direto no cadastro de sinal, paciente ou patologia, ele poderá acessar o menu Arquivo, conforme mostra a Figura 4-5.

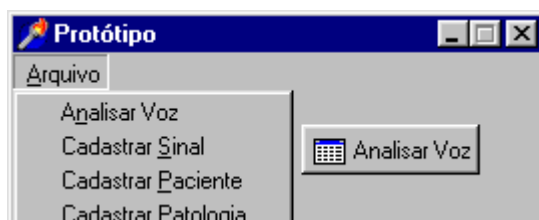
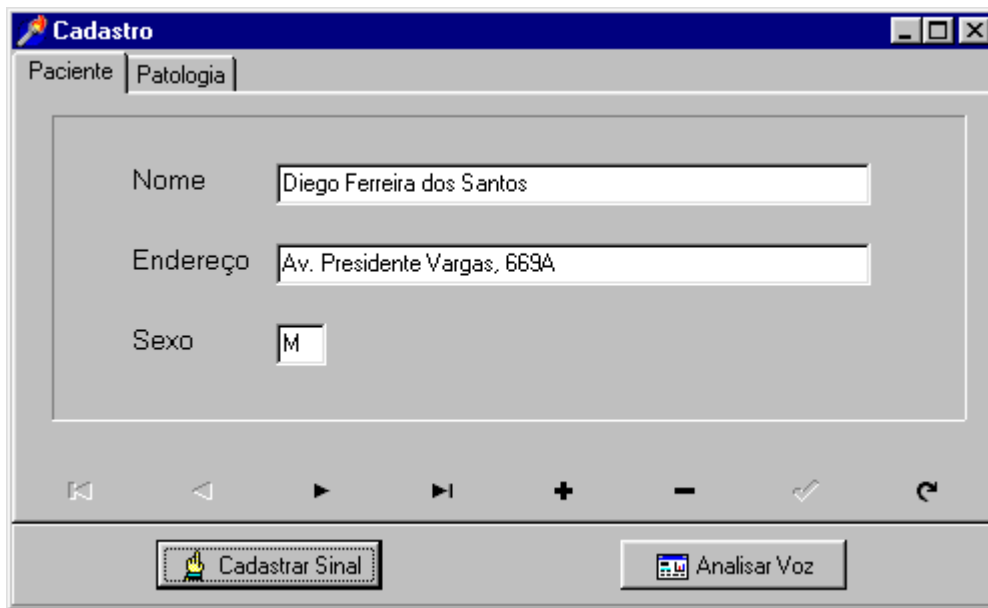


Figura 4-5 Menu arquivo do protótipo

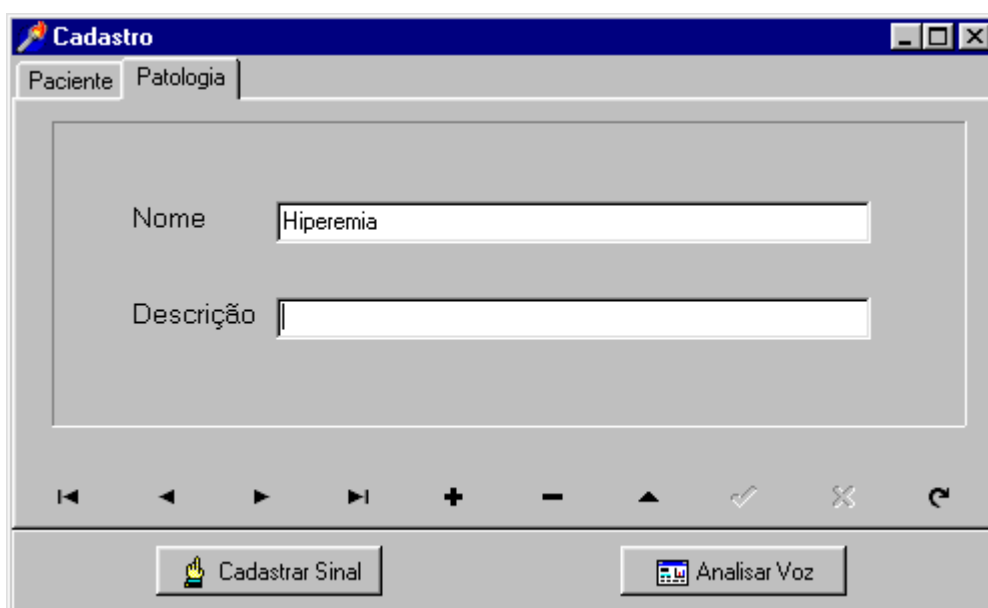
Se o médico não possui nenhum paciente cadastrado, ele clica então no botão Cadastro ou então acessa o menu Cadastrar Paciente. Aparecerá a seguinte janela:



**Figura 4-6 Janela cadastro**

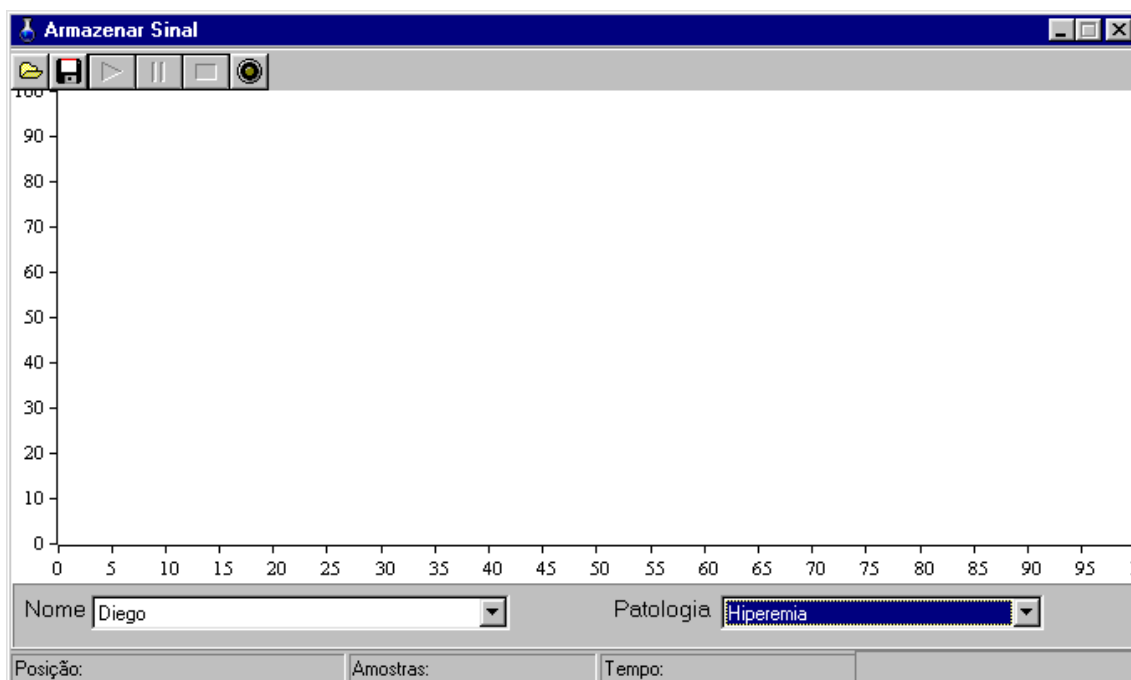
Onde o médico tem a opção de cadastrar os pacientes no banco de dados.

Se o médico já possui um paciente cadastrado e deseja cadastrar uma patologia, ele clica no controle Patologia, conforme mostra a Figura 4-7.



**Figura 4-7 Janela de cadastro de patologias**

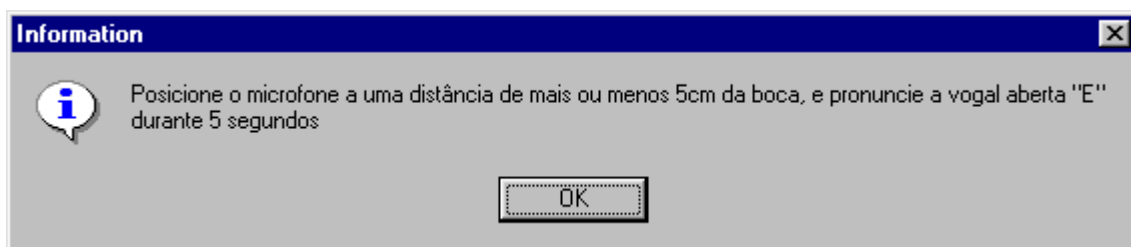
Se o médico já possui o cadastro do paciente e da patologia e deseja fazer a aquisição da voz do paciente para posterior análise, ele clica no botão Cadastrar Sinal.



**Figura 4-8** Janela de cadastro de sinal (aquisição do sinal de voz)

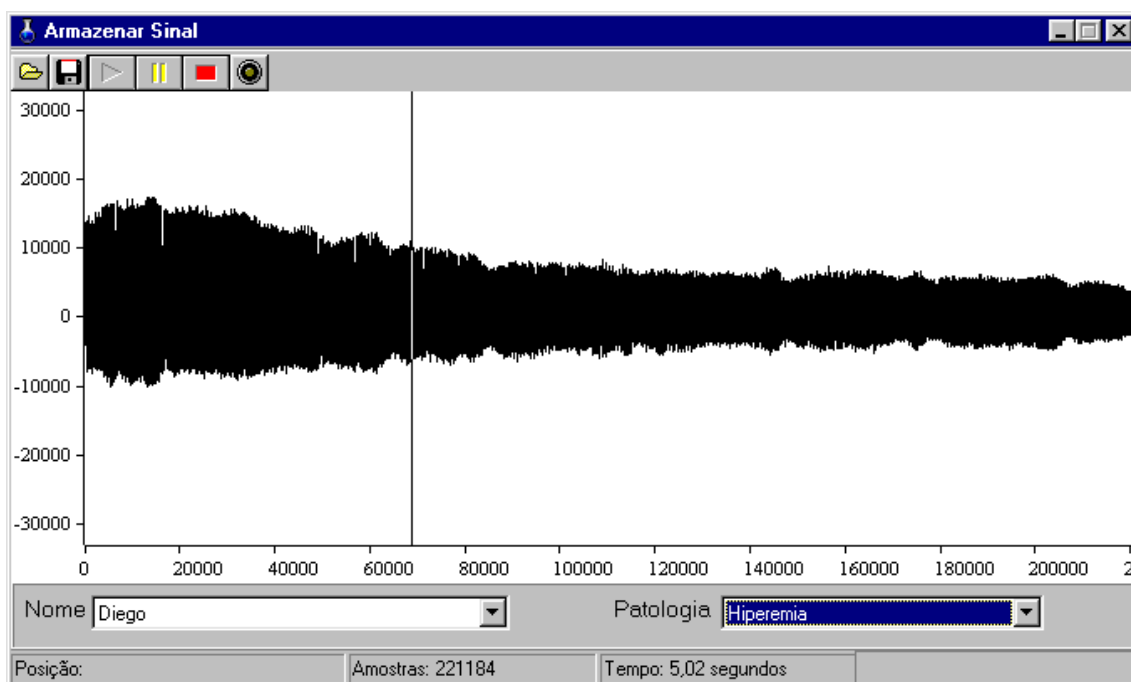
O médico pode então, conforme Figura 4-8, fazer a aquisição da voz do paciente relacionado com uma patologia através do microfone. Se por acaso o médico já possuir a voz do paciente armazenada em um arquivo formato WAV, ele poderá abrir esse arquivo para cadastrar o sinal na base de dados.

Se o médico optar pela aquisição da voz através do microfone, o paciente terá que pronunciar a vogal aberta “/ε/” durante 5 segundos, conforme mostra a Figura 4-9.



**Figura 4-9** Mensagem explicando como fazer a gravação

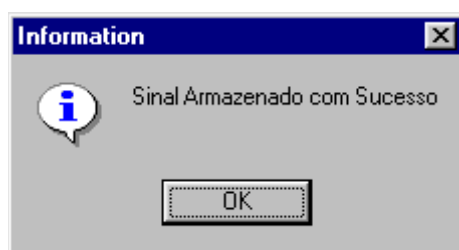
O Sinal do paciente então pode ser visto no gráfico, onde tem as opções de PLAY, PAUSE, STOP no sinal.



**Figura 4-10 Sinal do paciente**

O médico pode observar as características do sinal e verificar o número de amostras, o tempo total e a posição do cursor em relação ao sinal.

Se o médico determinar que o sinal está em boas condições (mais ou menos constante, sem ruídos externos), ele então armazena o sinal na base de dados pressionando o botão SALVAR da barra de ferramentas. Caso tudo ocorra sem problemas, aparecerá a seguinte mensagem:



**Figura 4-11 Caixa de diálogo dizendo que tudo ocorreu sem problemas**

Para realizar a análise dos sinais dos pacientes, o médico clica no botão (da tela inicial) Analisar Voz. Aparecerá a seguinte tela:

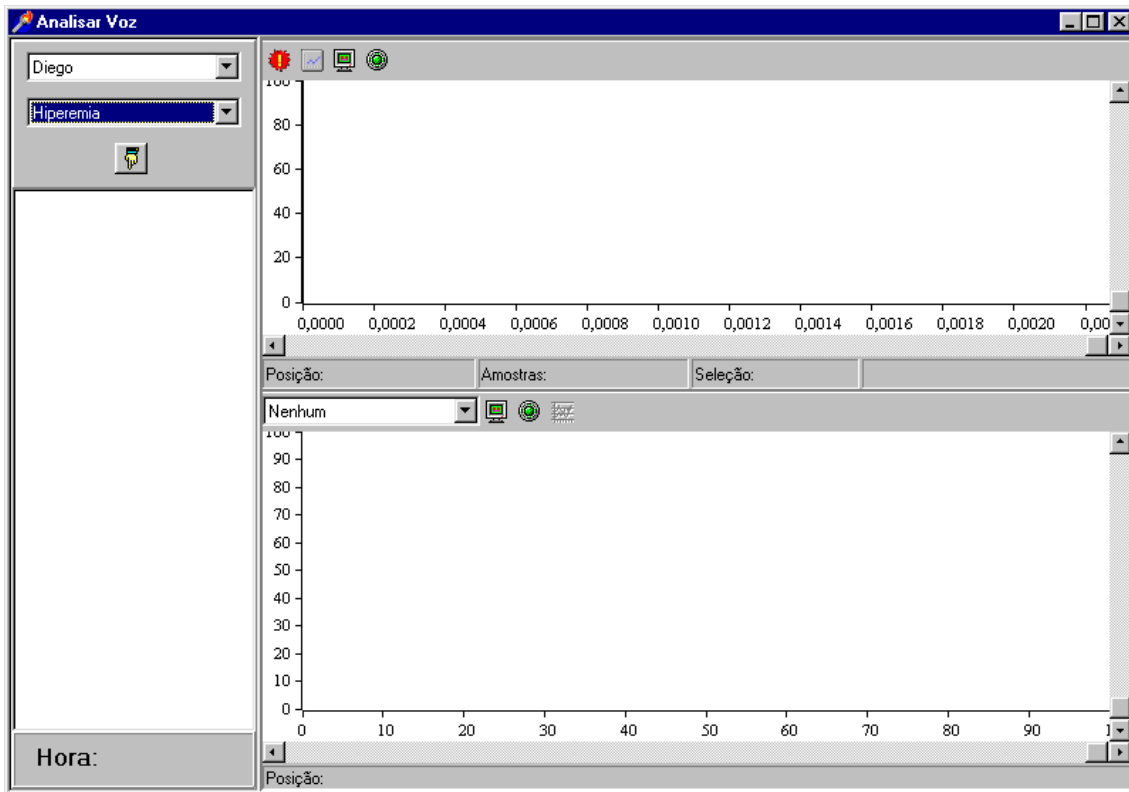


Figura 4-12 Janela de análise de voz

O médico então escolhe o paciente e a patologia para fazer a análise, o resultado da pesquisa será apresentado na *listbox*, conforme:

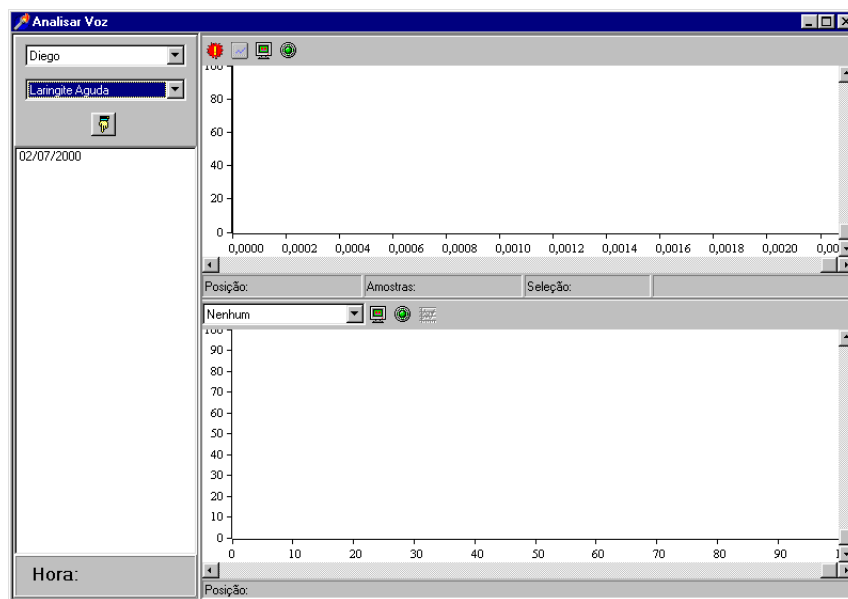
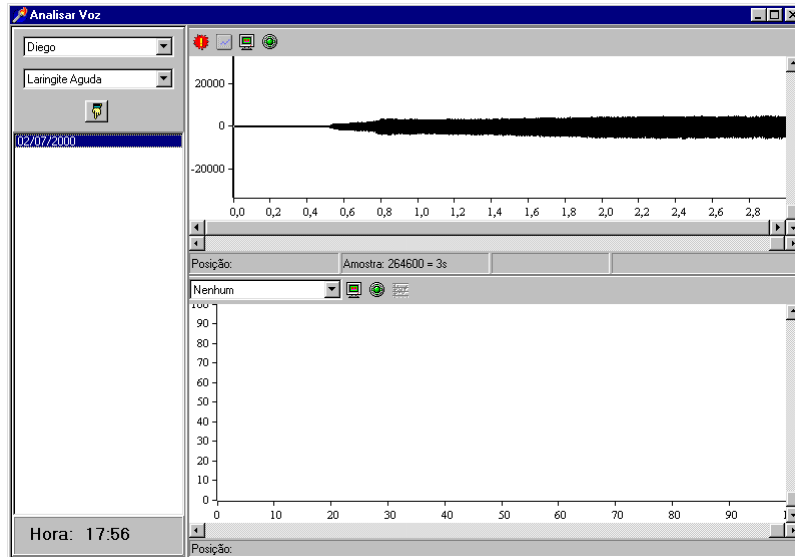


Figura 4-13 Sinais apresentados na ListBox do paciente DIEGO com LARINGITE AGUDA

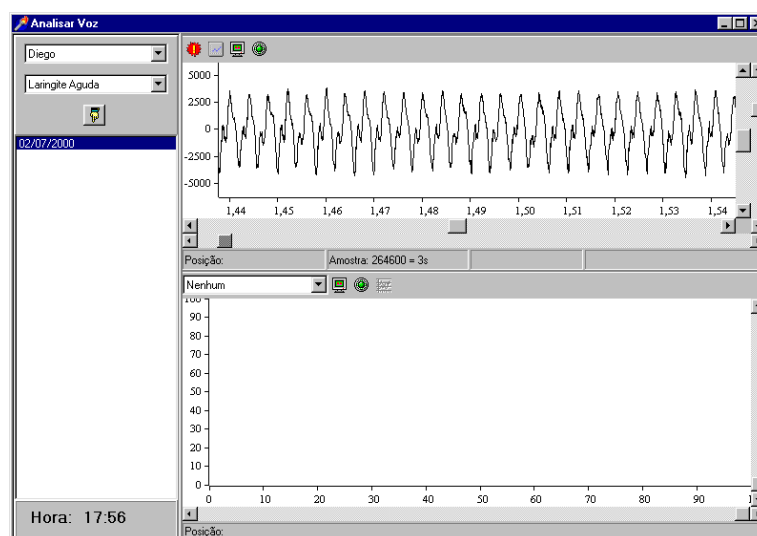
Os sinais aparecem na listbox de acordo com a data com que foi feita a aquisição. Após selecionar uma data, o sinal aparecerá no gráfico que está localizado na parte de cima. Figura 4-14



**Figura 4-14** Visualização do gráfico para análise

Em todos os gráficos o médico tem a opção de copiar a imagem do sinal para um arquivo no formato BMP ou então copiar para a área de transferência para usar a imagem em outros aplicativos.

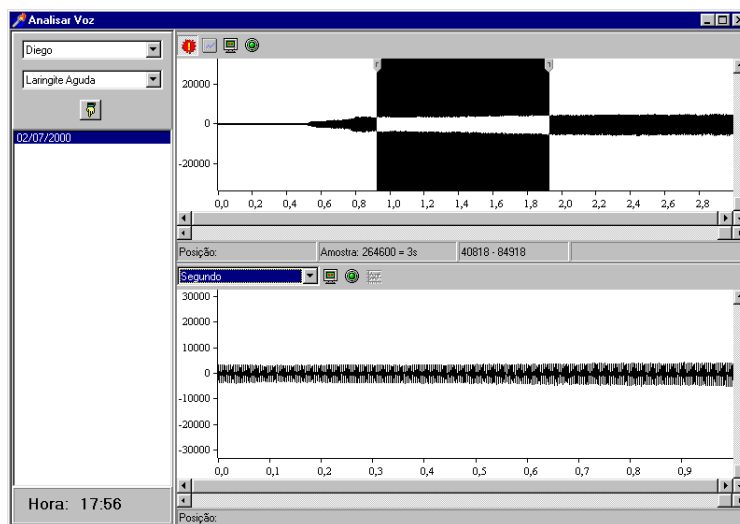
Uma outra característica que pode ser atribuída a todos os gráficos é a opção de fazer um *Zoom* no eixo horizontal ou então no eixo vertical. Figura 4-15



**Figura 4-15** Zoom do sinal

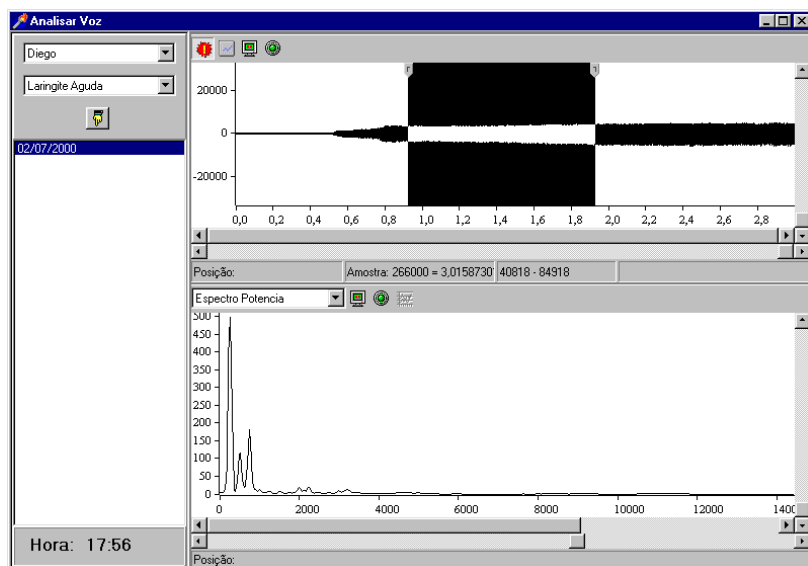
O médico então, para realizar a análise, tem que selecionar um trecho de 1s do sinal que pareça ser mais estável, ou seja, mais constante em amplitude. Para isso, ele pressiona o primeiro botão da barra de ferramentas e na *combobox* do gráfico localizado mais abaixo, ele seleciona o tipo de análise que deseja fazer. As opções são:

- Tempo → verifica somente o trecho selecionado no gráfico
- Espectro de Potência → gráfico do espectro de potência do trecho selecionado
- Cepstrum Real → cepstrum real do trecho selecionado
- Formantes → análise das 3 primeiras formantes do trecho selecionado
- Espectro da Resposta ao Impulso do Trato Vocal → gráfico do espectro da resposta impulsiva do trato vocal do trecho selecionado



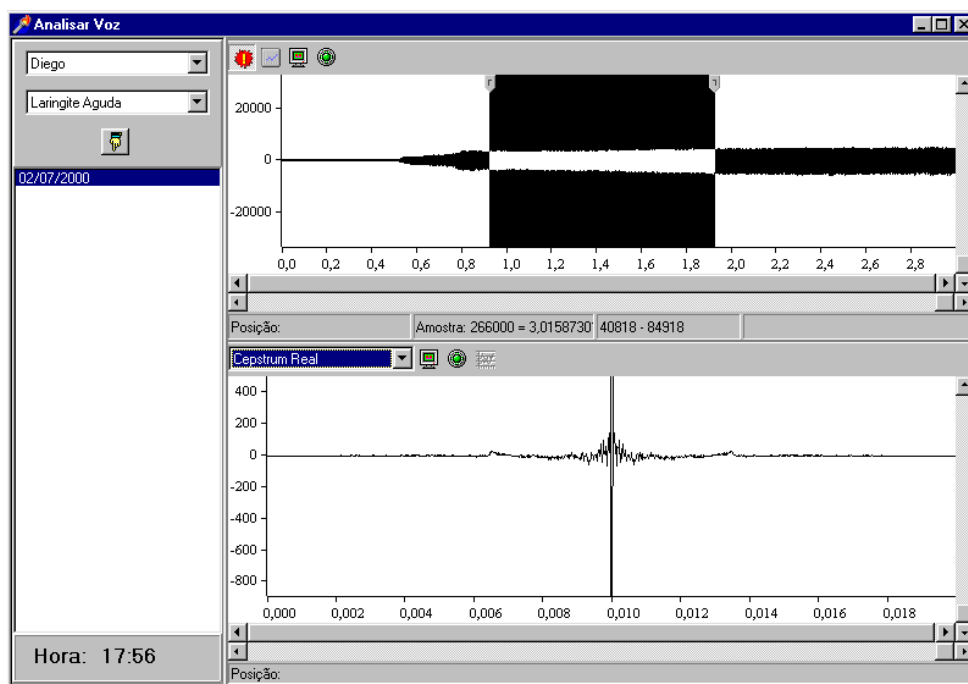
**Figura 4-16** No gráfico baixo está o trecho de 1 segundo selecionado no gráfico acima

Se o médico deseja fazer a análise do espectro de potência, então ele seleciona na *combobox* o Espectro de Potência, e o gráfico aparecerá logo em seguida:



**Figura 4-17 Espectro de potência do trecho selecionado**

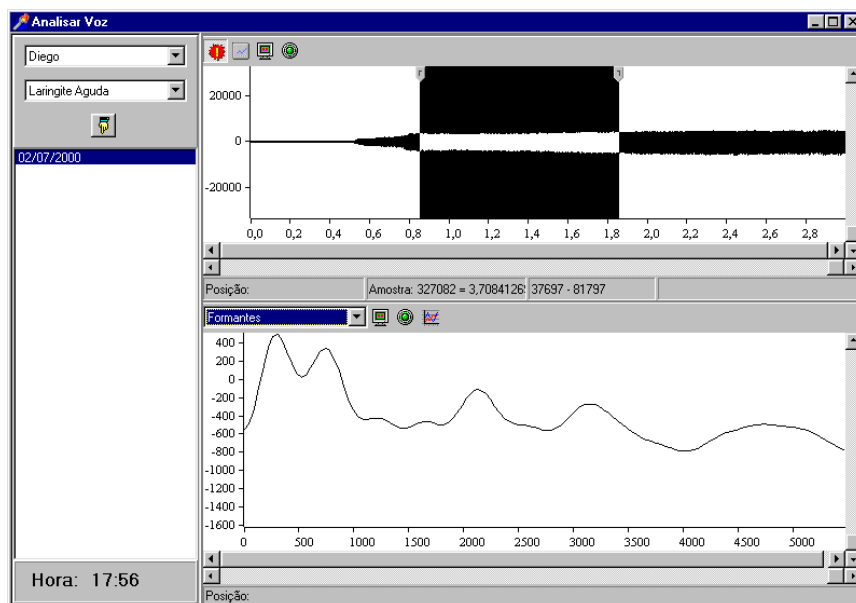
Caso o médico opte pelo cepstrum real, então:



**Figura 4-18 Cepstrum real do trecho selecionado**

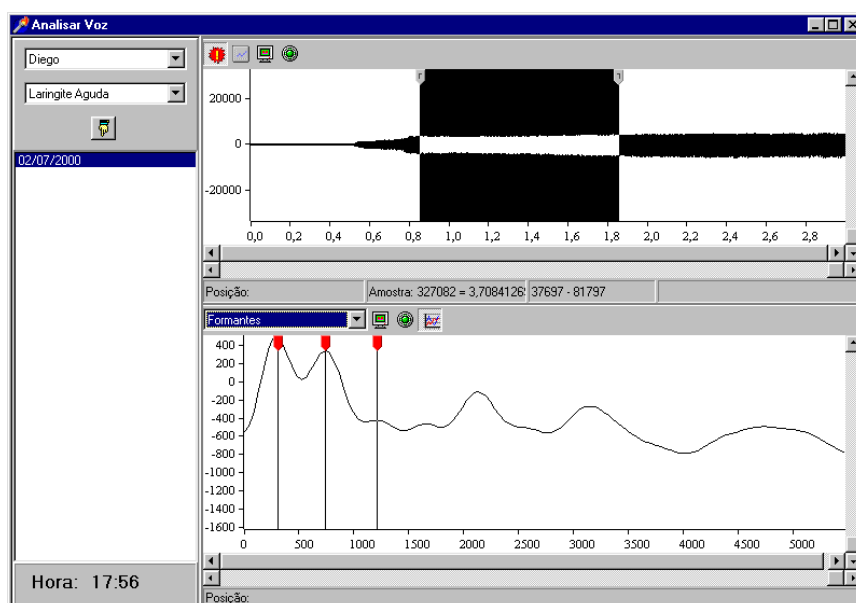
Se análise escolhida foi a análise de formantes, então:





**Figura 4-19 Análise das formantes do trecho selecionado**

Se o médico desejar ver onde estão os picos das 3 primeiras formantes, ele deverá clicar no último botão da barra de ferramentas, e as 3 primeiras formantes serão informadas, conforme Figura 4-20:



**Figura 4-20 Visualização das 3 primeiras formantes**

## **5 Conclusão**

A filtragem homomórfica, além da análise do cepstrum real, se mostrou uma ferramenta interessante para auxiliar no diagnóstico de patologias do trato vocal. É possível correlacionar melhoras do estado clínico do paciente com as alterações do cepstrum real, da resposta ao impulso do trato vocal e do espectro da resposta ao impulso do trato vocal. Já o espectro de voz e a análise das formantes se mostraram ineficientes para fazer essa análise, visto que as características dos sinais são muito semelhantes antes e após o tratamento do paciente.

### **5.1 Possibilidades Futuras**

Devido ao pouco tempo, o número de pacientes e patologias analisados foi muito pequeno. Para chegar a um resultado melhor e mais eficaz, poderá ser feita a aquisição de vozes de um número maior de patologias e de um número maior de pacientes com a mesma patologia para realizar-se uma comparação.

Quanto ao protótipo, evidentemente, não está completo, ele é apenas uma base para posterior elaboração de um sistema completo de auxílio ao diagnóstico de patologias do trato vocal.

Serão acrescentadas as funções de espectro de potência e espectro da resposta ao impulso do trato vocal. Poderão ser modificados os campos das tabelas da base de dados, acrescentando mais informações. Poderá ser feito um sistema de busca de sinal, com a visualização do mesmo.

## 6 Bibliografia

ALBERNAZ, Paulo. M. *Otorrinolaringologia Prática*. São Paulo: Fundo Editorial Prociex, 1963.

COOLEY, J. W., TUKEY, J. W. *An Algorithm for the machine computation of complex Fourier series*. Mathematics of Computation, v.19, p.297-301, abr. 1965

COSTA, Cristiano A. *Normas para apresentação do Projeto de Graduação na Escola de Informática da UCPEL*. Pelotas: Educat, 1999.

DIAS, Mario U. V. *Fisiologia Humana*. Rio de Janeiro: Editora Guanabara, 1960.

HUNGRIA, H. *Manual de Oto-rinolaringologia*. Rio de Janeiro: Editora Guanabara, 1960.

LYONS, R.G. *Understanding Digital Signal Processing*. Estados Unidos: Addison Wesley, 1999.

OPPENHEIM, A.V., SCHAFER, R.W. *Discrete-Time Signal Processing*. Londres: Prentice-Hall Internacional, 1989.

PELTON, G.E. *Voice Processing*. Singapura: McGraw-Hill, 1992.

ROCHA, Luis. *Procesamiento de Voz*. Buenos Aires: Kapelusz, 1997.

ROSA, Marcelo de O., PEREIRA, José C., GRELLET, M. *Adaptative Estimation of Residue Signal for Voice Pathology Diagnosis*. IEE Transactions Biomed. Eng., v. 47, p.96-104, Jan. 2000

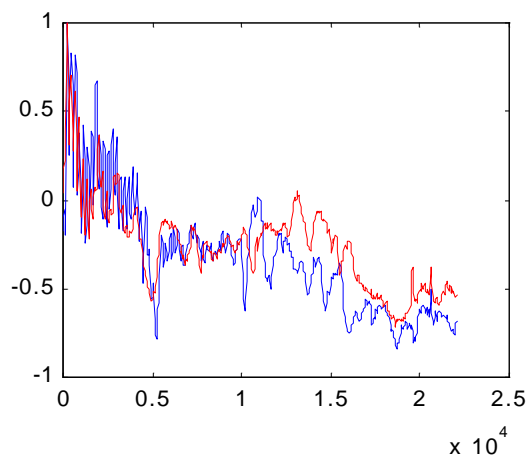
SMITH, Steven W. *The Scientist and Engineer's Guide to Digital Signal Processing*. San Diego: California Technical Publishing, 1999.

ZHU, B., TZANAKOU, E. M. *Nonstationary Speech Analysis Using Neural Prediction*. IEEE Engineering in Medicine and Biology, v.19, p.102-105, Jan/Fev. 2000.

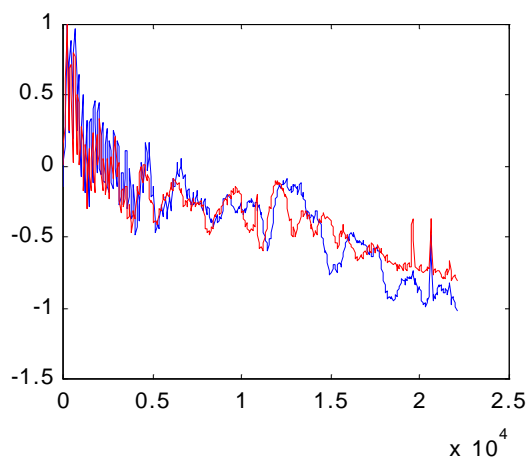
## **7 Anexo**

A seguir, encontram-se em anexo os gráficos das análises dos pacientes que não foram apresentados no documento, as rotinas utilizadas para elaboração da análise e as rotinas utilizadas para elaboração do protótipo.

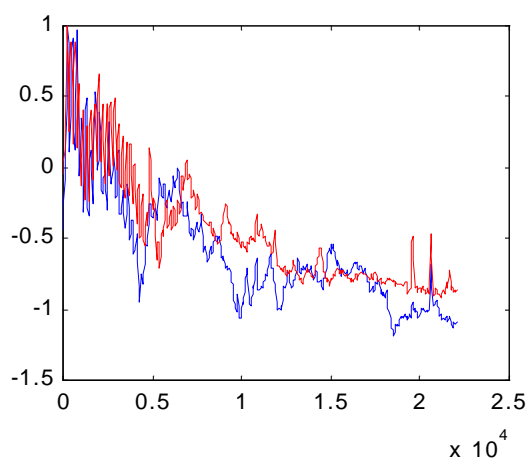
## 7.1 Anexo A – Espectro de Voz



**Figura 7-1 Paciente 3. Vermelho = sinal antes do tratamento. Azul = sinal após o tratamento**

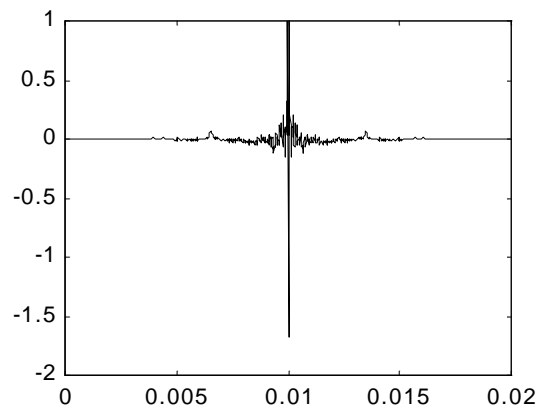


**Figura 7-2 Paciente 4. Vermelho = sinal antes do tratamento. Azul = sinal após o tratamento**

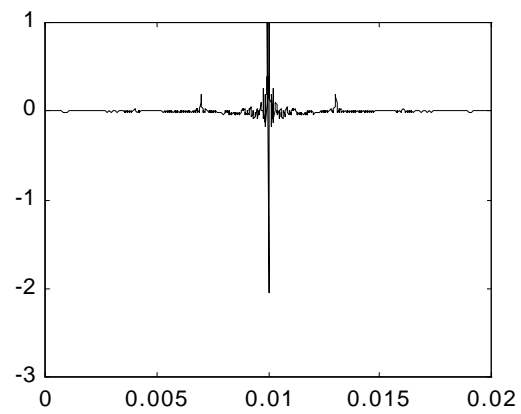


**Figura 7-3 Paciente 5. Vermelho = sinal antes do tratamento. Azul = sinal após o tratamento**

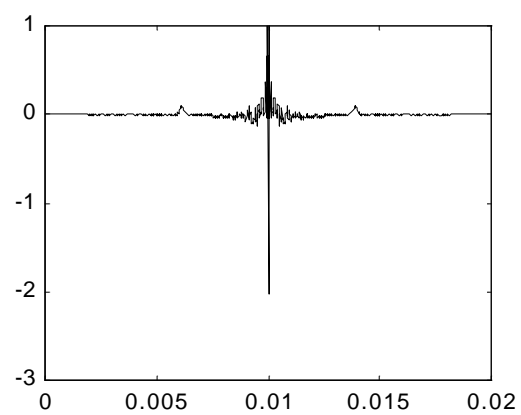
## 7.2 Anexo B – Cepstrum Real



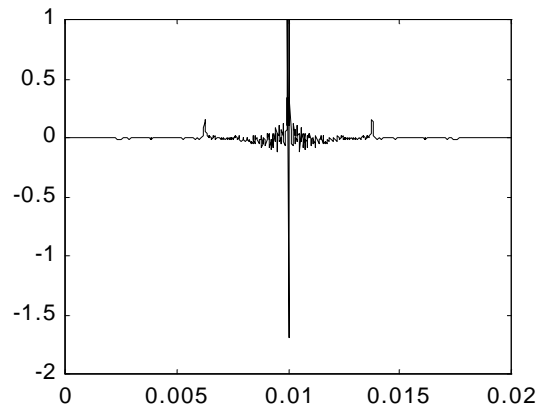
**Figura 7-4 Paciente 1 antes do tratamento**



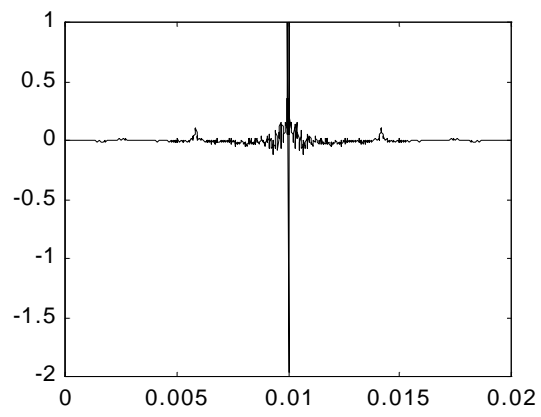
**Figura 7-5 Paciente 1 após o tratamento**



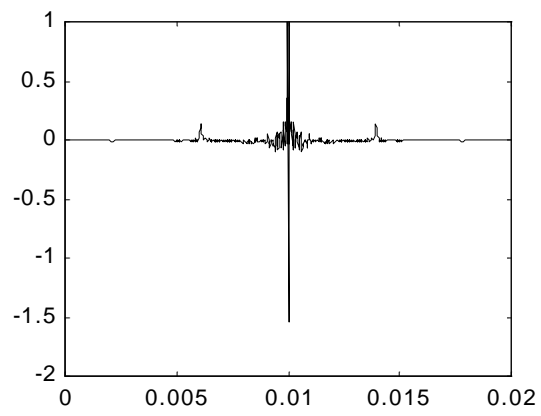
**Figura 7-6 Paciente 3 antes do tratamento**



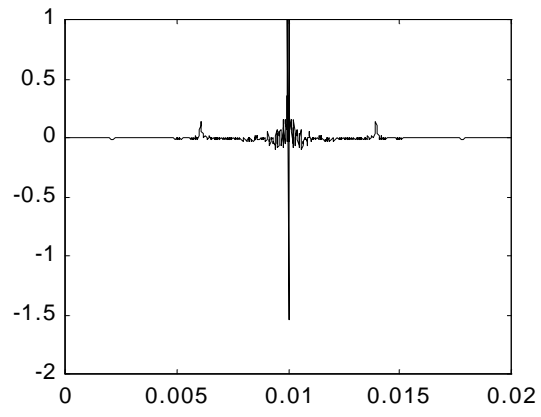
**Figura 7-7 Paciente 3 após o tratamento**



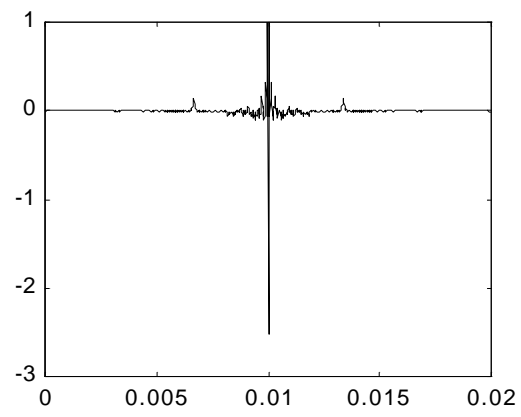
**Figura 7-8 Paciente 4 antes do tratamento**



**Figura 7-9 Paciente 4 após o tratamento**



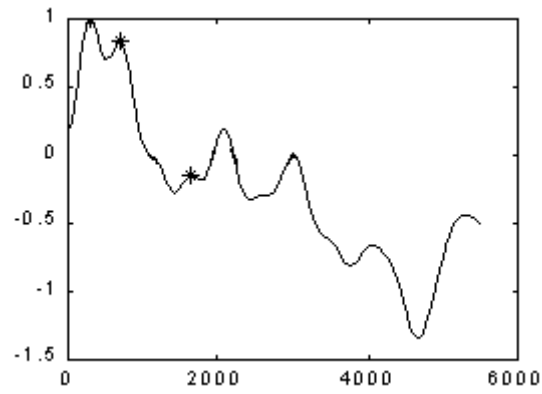
**Figura 7-10 Paciente 5 antes do tratamento**



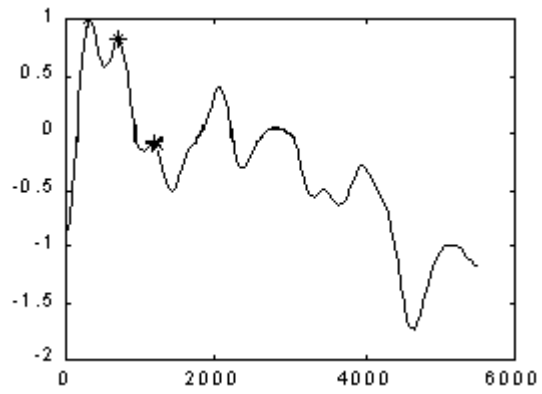
**Figura 7-11 Paciente 5 após o tratamento**



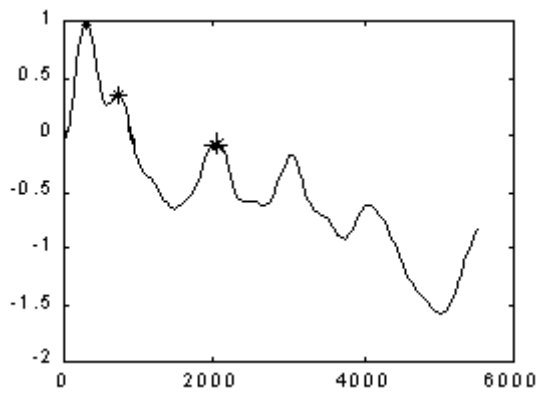
### 7.3 Anexo C – Formantes



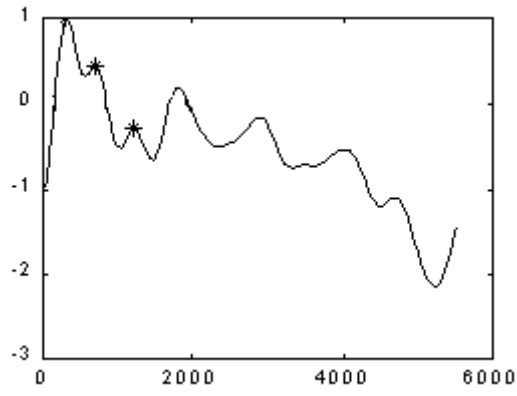
**Figura 7-12 Paciente 2 antes do tratamento**



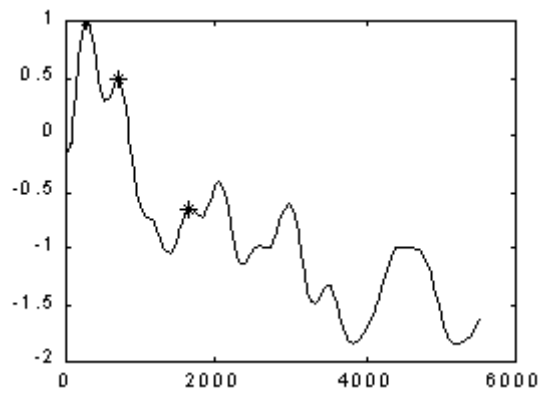
**Figura 7-13 Paciente 2 após o tratamento**



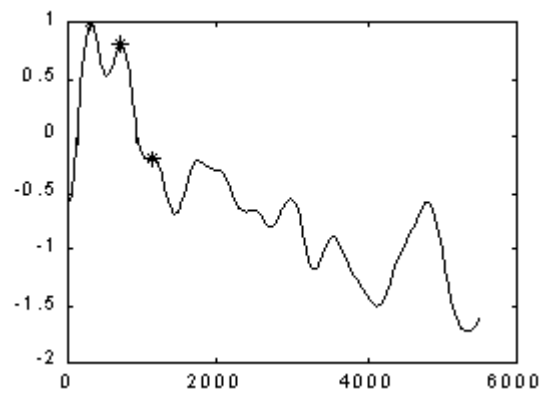
**Figura 7-14 Paciente 3 antes do tratamento**



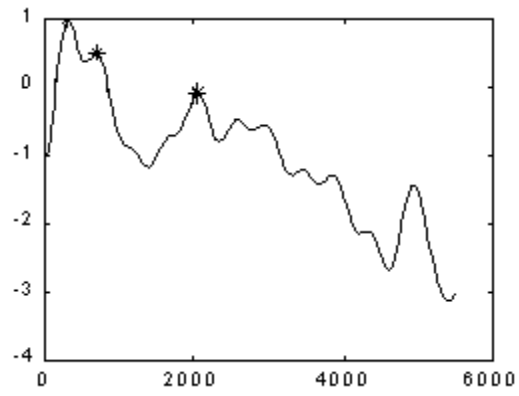
**Figura 7-15 Paciente 3 após o tratamento**



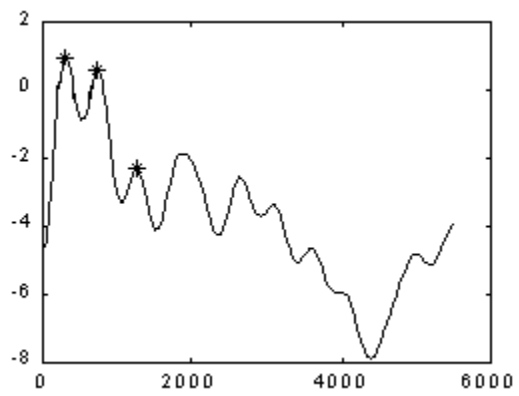
**Figura 7-16 Paciente 4 antes do tratamento**



**Figura 7-17 Paciente 4 após o tratamento**

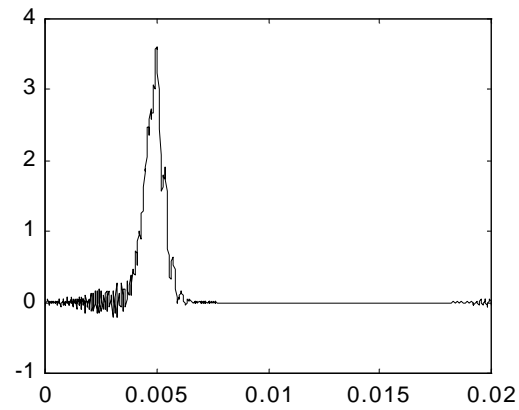


**Figura 7-18 Paciente 5 antes do tratamento**

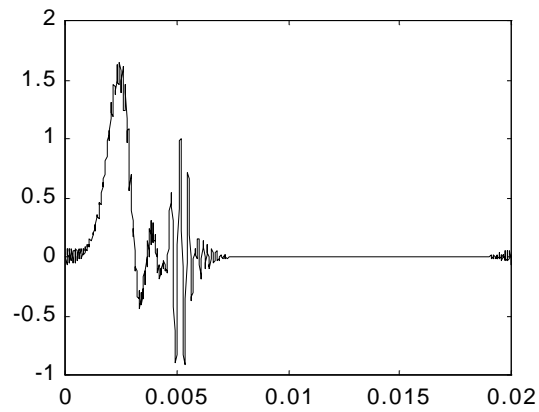


**Figura 7-19 Paciente 5 após o tratamento**

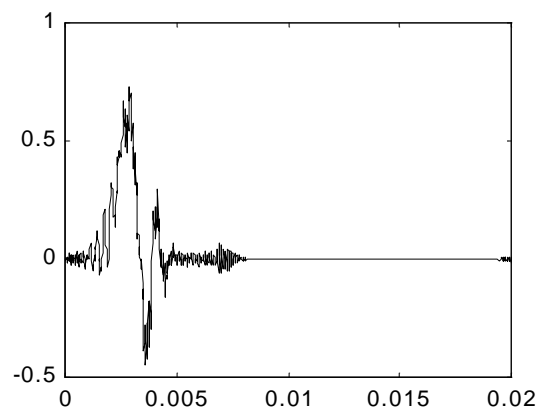
#### 7.4 Anexo D - Resposta ao Impulso do Trato Vocal



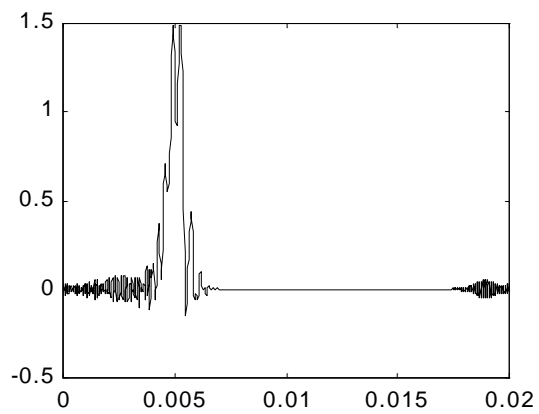
**Figura 7-20 Paciente 2 antes do tratamento**



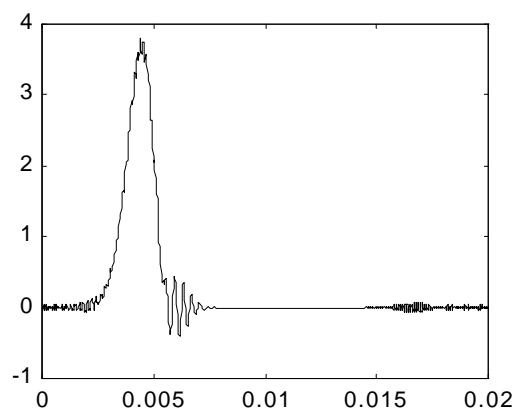
**Figura 7-21 Paciente 2 após o tratamento**



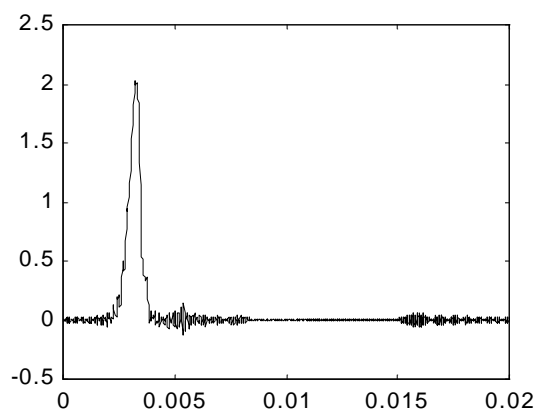
**Figura 7-22 Paciente 3 antes do tratamento**



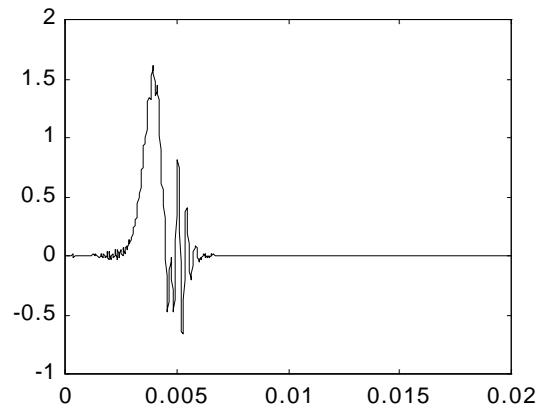
**Figura 7-23 Paciente 3 após o tratamento**



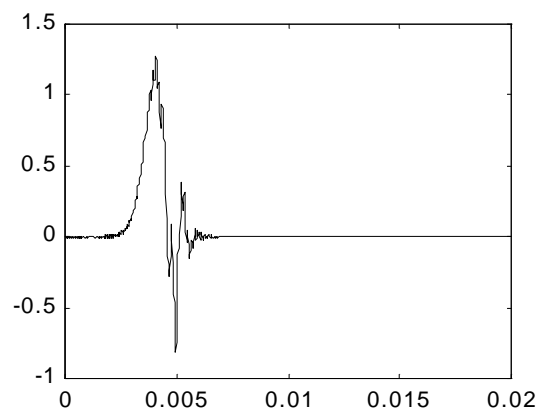
**Figura 7-24 Paciente 4 antes do tratamento**



**Figura 7-25 Paciente 4 após o tratamento**

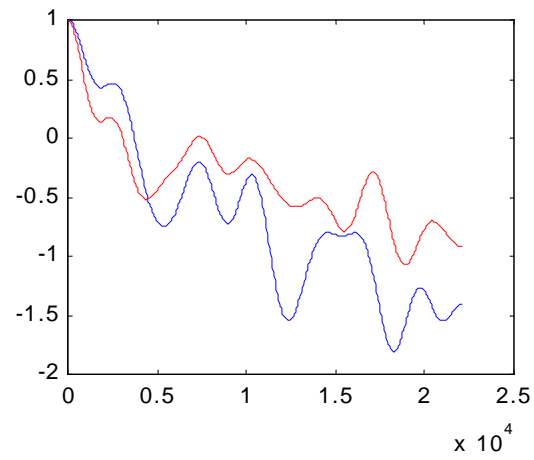


**Figura 7-26 Paciente 5 antes do tratamento**

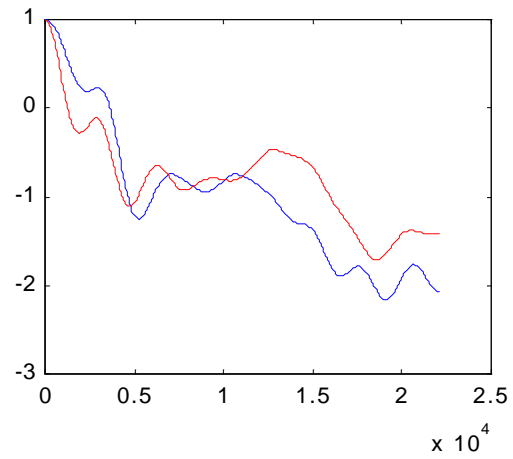


**Figura 7-27 Paciente 5 após o tratamento**

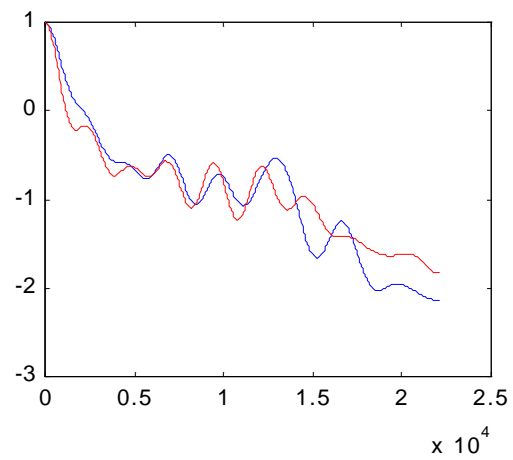
## 7.5 Anexo E – Espectro da Resposta Impulsiva do Trato Vocal



**Figura 7-28 Vermelho = Paciente 2 antes do tratamento. Azul = Paciente 2 após o tratamento**



**Figura 7-29 Vermelho = Paciente 3 antes do tratamento. Azul = Paciente 3 após o tratamento**



**Figura 7-30 Vermelho = Paciente 4 antes do tratamento. Azul = Paciente 4 após o tratamento**

## 7.6 Anexo F – Biblioteca para Leitura do Formato de Arquivo .WAV

```
unit WaveLib;

interface

Uses Windows, Sysutils, LibUtil;

//const SampleRate = 44100; //soh serah permitido essa taxa de
aquisicao
//const segundos = 10; //Esse será o tempo máximo permitido em
//um arquivo .WAV

Type

// Valores para retorno de algum erro
ERRO = (
        arquivo,
        header,
        chunk_formato,
        compression,
        canal,
        samples_sec,
        bits_sample,
        memoria,
        tamanho_arq,
        OK
    );

// Chuck Header - Para todos os chunks
WChunk_Hdr = record
    id : array[1..4] of char;
    len: DWORD;
end;

// Chunk Format Wav
WFormat = record
    wFormatTag      : WORD;
    wChannels       : WORD;
    dwSamplesPerSec : DWORD;
    dwAvgBytesPerSec : DWORD;
    wBlockAlign    : WORD;
    wBitsPerSample  : WORD;
end;

// Conteudo completo das informações do Arquivo wav
Inf_Wave = record
    riff_header      : WChunk_Hdr;
    wavid            : array[1..4] of char;
    formato_header   : WChunk_Hdr;
    formato          : WFormat;
    data_header      : WChunk_Hdr;
end;

//Classe para leitura de Arquivos Wave
TWave = class
Private
```



```

        WaveFile : File;
        Info      : Inf_Wave; //guarda as informações do arquivo
wave
        H          : THandle; //handle para uma área da memória
    Public
        Function Open_Wave (Filename: String) : Boolean;
        Function read_wave(var wavdata:PSmallIntArray):ERRO;
        Function n_samples : DWORD;
        Procedure Libera;

    end;

implementation

{ Wave }

Procedure TWave.Libera;
Begin
    GlobalFree(h);
End;

/////////////////////////////////////////////////////////////////
// Funcao: Abre um arquivo Wave para leitura
// Entrada: F = Arquivo para leitura
//          Filename = Nome do arquivo
/////////////////////////////////////////////////////////////////
function TWave.Open_Wave(Filename: String): Boolean;

begin
    {$I-}
    AssignFile(WaveFile, Filename);
    FileMode := 0;

    Reset(WaveFile,1);
    {$I+}
    if IOResult <> 0 then
        Open_Wave := False
    else
        Open_Wave := True;
end;

Function TWave.read_wave(var wavdata : PSmallIntArray):ERRO;
Begin
    wavdata := NIL;
    //carrega toda a informação contida no arquivo wave
    blockread(wavefile,info,sizeof(inf_wave));

    //verifica se é um arquivo wave
    if (info.riff_header.id <> 'RIFF') OR (info.wavid <> 'WAVE') then
    begin
        Read_wave := Header;
        exit;
    end;

    //verifica se existe corretamente o chunk formato
    if info.formato_header.id <> 'fmt ' then
    begin
        Read_Wave := Chunk_Formato;
        exit;
    end;
end;

```

```

//verifica se o arquivo wave está comprimido
if info.formato.wFormatTag <> 1 then
begin
    Read_Wave := Compression;
    exit;
end;

//verifica se o som foi gravado em mono
if info.formato.wChannels <> 1 then
begin
    Read_Wave := canal;
    exit;
end;

//verifica se a taxa de transferencia é de 44100
if info.formato.dwSamplesPerSec <> 44100 then
begin
    Read_Wave := samples_sec;
    exit;
end;

//verifica se o número de bits usado é de 16
if info.formato.wBitsPerSample <> 16 then
begin
    Read_Wave := bits_sample;
    exit;
end;

//aloca memoria para os dados do wave
h := GlobalAlloc(GMEM_FIXED, info.data_header.len);

//se der erro de memoria !
if h = NULL then
begin
    Read_Wave := memoria;
    exit;
end;

//atribui a área de memoria à um ponteiro !
wavdata := GlobalLock(h);

//le os dados do arquivo wave
blockread(WaveFile,wavdata^,info.data_header.len);

CloseFile(WaveFile);

Read_Wave := OK;

End;

// retorna o número de amostras do arquivo WAV.
Function TWave.n_samples : DWORD;
begin
    Result := (info.data_header.len div 2);
end;

end.

```

## 7.7 Anexo G – Rotinas em Delphi para Análise de Sinais

### 7.7.1.1 Espectro de Potência

A seguir está a rotina utilizada para a realização do Espectro de Potência:

```
////////////////////////////////////  
// Funcao: Calcular o espectro de potencia de um trecho de 1s//  
// Entrada: dados = Trecho de 1 segundo normalizado  
// Saida : trecho de 1024 samples com o espectro de potencia  
////////////////////////////////////  
Procedure TDSP.espectro_potencia(dados: PFloatArray;var  
Normalizado:PFloatArray);  
var  
    trechos : integer;  
    s : PFloatArray; //sinal  
    m : PFloatArray; //média  
    resultado : PFloatArray; //espectro de potencia  
    magfft : PFloatArray; //magnitudo da fft  
    salto : Integer;  
    cont : integer;  
    sfft : PSCplx; //fft do sinal  
  
begin  
  
    salto := 0 ;  
  
    GetMem(s,1024*SizeOf(Float));  
    GetMem(m,1024*SizeOf(Float));  
    GetMem(resultado,1024*SizeOf(Float));  
  
    //50 trechos de 20ms cada, fazendo com q fique 1 s no final  
    for trechos:=1 to 50 do  
        begin  
            salto := (trechos-1) * 882;  
            for cont:=0 to 881 do  
                s^[cont] := dados^[cont + salto]; //copia os trechos  
//para o S  
  
                //faz o janelamento  
                nspsWinHann(PFloat(s),882);  
  
                getmem(magfft,1024*Sizeof(Float));  
                modulo_fft(s,magfft); //essa função retorna a magnitudo da  
//FFT  
  
                //faz a soma para fazer o calculo da média  
                nspsbAdd2(PFloat(magfft),PFloat(m),1024);  
            end;  
            //calcula a media dividindo os valores por 50 (trechos)  
            nspsbNormalize(PFloat(m),PFloat(resultado),1024,0,50);  
  
            //normaliza o resultado  
            nspsbNormalize(PFloat(resultado),  
                PFloat(normalizado),1024,0,nspsMax(PFloat(resultado),1024));
```

```

//faz o valor máximo = 500
nspsbMpy1(500,PFloat(normalizado),1024);

FreeMem(s,1024*Sizeof(Float));
FreeMem(magfft,1024*Sizeof(Float));
FreeMem(m,1024*SizeOf(Float));
FreeMem(resultado,1024*Sizeof(Float));

end;

```

### 7.7.1.2 Cepstrum Real

A seguir estão as rotinas utilizada para a realização do Cepstrum Real:

```

////////////////////////////////////
// Função: Calculo do cepstrum real do sinal
// Entrada: dados = trecho para fazer o cepstrum real
//          tipo = 1 se for normal
//          tipo = 2 se for para fazer a montagem do cepstrum tbm
// Saida : cepstrum_real
////////////////////////////////////
Function TDsp.rceps(dados:PFloatArray; Tipo: Byte):PFloatArray;
Var
  ifft : PSCplx;
  magfft : PFloatArray;
  partereal : PFloatArray;
  cepstrum : PFloatArray;
  c : integer;
begin
  GetMem(ifft,1024*SizeOf(TSCplx));
  GetMem(partereal,1024*SizeOf(Float));

  if tipo = 2 then
    GetMem(cepstrum,1024*SizeOf(Float));

    //magnitudo da fft
    getmem(magfft,1024*Sizeof(Float));
    modulo_fft(dados,magfft);

    //calcula o log da fft
    nspsbLn1(PFloat(magfft),1024);

    //calcula o inverso da FFT
    nspsRealFFTNip(PFloat(magfft),ifft,10,NSP_INV);
    //Retorna a Parte Real
    nspcbReal(ifft,PFloat(partereal),1024);

    //faz a simetria
    for c:=0 to 511 do
      partereal^[1023-c] := partereal^[c];

    //faz a montagem do cepstrum
    if tipo = 2 then
      for c:=0 to 511 do
        Begin
          cepstrum^[c] := partereal^[c+512];
          cepstrum^[c+512] := partereal^[c];
        End;

```

```

FreeMem(iff,1024*SizeOf(TSCplx));
FreeMem(magFft,1024*SizeOf(Float));
if tipo = 2 then
    FreeMem(partereal,1024*SizeOf(Float));

if tipo = 2 then
    Result := @cepstrum[0]
else Result := @partereal[0];

end;

////////////////////////////////////
// Função: Faz o cálculo do Cepstrum_real da Voz no trecho de 1s
// Será analisado 50 trechos de 1s da voz e calculada a media
// do cepstrum real
// Entrada: dados = trecho de 1 segundo
// Saída: Trecho contendo a media do cepstrum real normalizado
////////////////////////////////////
Function TDsp.Cepstrum_Real (dados:PFloatArray):PFloatArray;
var
    Trechos : Integer; //50 trechos
    s       : PFloatArray; //Sinal
    Resultado: PFloatArray; //Cepstrum_Real
    media   : PFloatArray; //retorna a media do cepstrum real
    Normaliza : PFloatArray;
    salto    : integer;
    c        : integer;
    cont     : integer;
    cepstrum : PFloatArray;

Begin
    GetMem(s,1024*Sizeof(Float));
    GetMem(media,1024*SizeOf(Float));
    GetMem(resultado,1024*SizeOf(Float));
    GetMem(Normaliza,1024*Sizeof(Float));

    For trechos :=1 to 50 do
        Begin
            salto := (trechos-1) * 882;
            for cont:=0 to 881 do
                s^[cont] := dados^[cont + salto];

                // faz o janelamento
                nspsWinHann(PFloat(s),882);

                //calcula o cepstrum real
                cepstrum:=rceps(s,2);

                //soma para fazer o calculo da media
                nspsbAdd2(PFloat(cepstrum),PFloat(media), 1024);
            End;

            //faz a divisao para calulo da media
            nspsbNormalize(PFloat(media), PFloat(resultado),1024,0,50);

            //normaliza o resultado
            nspsbNormalize(PFloat(resultado),
                PFloat(normaliza),1024,0,nspsMax(PFloat(resultado),1024));
        End;
    End;

```

```

    nspsbMpy1(500, PFloat(normaliza), 1024);

    FreeMem(s, 1024*Sizeof(Float));
    FreeMem(media, 1024*SizeOf(Float));
    FreeMem(cepstrum, 1024*SizeOf(Float));
    Freemem(resultado, 1024*Sizeof(Float));

    Result := @Normaliza[0];

End;

```

### 7.7.1.3 Análise das Formantes

A seguir está a rotina utilizada para a realização da Análise das Formantes do Sinal.

```

//Função: Faz o calculo das 3 primeiras formantes do espectro do cepstrum filtrado
// Entrada: dados = Trecho de 1s com o sinal a ser processado
// Saída: pos = posição das 3 primeiras formantes no sinal

Function TDsp.formante (dados : PFloatArray; var pos : TPOS):
PFloatArray;
var
    salto : integer;
    trecho : integer;
    Resultado : PFloatArray;
    s : PFloatArray;
    cont : integer;
    rcepstrum : PFloatArray; //cepstrum real
    filtro : PfloatArray; //cepstrum filtrado
    rfftceps : PFloatArray; //parte real da fft do cepstrum real
    rfftfceps : PFloatArray; //parte real da fft do cepstrum filtrado
    sfft : PSCplx; //parte complexa do cepstrum real
    sfftf : PSCplx; //parte complexa do cepstrum filtrado
    maior : Float; //maior pico
    Form : PFloatArray; //vetor com as formantes
    c : Integer; //contador
    media : PFloatArray; //para calculo da media
begin
    GetMem(s, 1024*Sizeof(Float));
    GetMem(filtro, 1024 * SizeOf(Float));
    GetMem(sfft, 1024*Sizeof(TSCplx));
    GetMem(rfftceps, 1024*Sizeof(Float));
    GetMem(sfftf, 1024*Sizeof(TSCplx));
    GetMem(rfftfceps, 1024*Sizeof(Float));
    GetMem(media, 1024*Sizeof(Float));

    for trecho := 1 to 50 do
    begin
        salto := (trecho - 1) * 882;

        for cont:=0 to 881 do
            s^[cont] := dados^[cont + salto];

            //janelamento
            nspsWinHann(PFloat(s), 882);

```

```

//cepstrum real
rcepstrum := rceps(s,1);

//copia os primeiros 100 samples para o filtro
for cont:=0 to 99 do
    filtro^[cont] := rcepstrum^[cont];

//fft do cepstrum real
nspRealFFTNip(PFloat(rcepstrum),sfft,10,NSP_FORW);

//retorna a parte real
nspcbReal(sfft,PFloat(rfftceps),1024);

//faz a simetria
for cont:=0 to 511 do
    rfftceps^[1023-cont] := rfftceps^[cont];

//fft do cepstrum filtrado
nspRealFFTNip(PFloat(filtro),sfft,10,NSP_FORW);

//retorna a parte real
nspcbReal(sfft,PFloat(rfftceps),1024);

//faz a simetria
for cont:=0 to 511 do
    rfftceps^[1023-cont] := rfftceps^[cont];

nspsbAdd2(PFloat(rfftceps),PFloat(media), 1024);
end; //do primeiro FOR

FreeMem(rfftceps,1024*Sizeof(Float));
GetMem(rfftceps,1024*Sizeof(Float));
GetMem(resultado,1024*Sizeof(Float));
//faz a divisao para calculo da media
nspsbNormalize(PFloat(media), PFloat(resultado),1024,0,50);

//normaliza o resultado
nspsbNormalize(PFloat(resultado),
PFloat(rfftceps),1024,0,nspMax(PFloat(resultado),1024));

//faz amplitude máxima = 500
nspsbMpy1(500,PFloat(rfftceps),1024);

//libera as variáveis que não serão mais utilizadas
FreeMem(S, 1024*Sizeof(Float));
FreeMem(rcepstrum,1024*Sizeof(float));
FreeMem(filtro,1024*Sizeof(Float));
FreeMem(sfft,1024*Sizeof(TSCplx));
FreeMem(sfft,1024*Sizeof(TSCplx));
FreeMem(media,1024*Sizeof(Float));
FreeMem(resultado,1024*Sizeof(Float));
//faz o calculo das formantes

GetMem(Form,1024*Sizeof(Float));

c:=1;
maior := rfftceps^[c];
c := 0; //indice da formante

for cont:=0 to 1023 do

```

```

        form^[c] := 0;

    for cont:=0 to 2 do
        pos[cont] := 0;

        //procura pelos 3 maiores picos no sinal
    for cont:=2 to 100 do
    begin
        if (rfftfceps^[cont] < maior) AND (rfftfceps^[cont-2] <
maior) then
            begin
                form^[c] := rfftfceps^[cont-1];
                maior := rfftfceps^[cont];
                if c <= 2 then
                    pos[c] := cont - 1;
                    Inc(c);
                end
            else
                maior := rfftfceps^[cont];
            end;

    FreeMem(rfftfceps,1024*Sizeof(Float));
    FreeMem(Form,1024*Sizeof(float));

    Result := @rfftfceps[0];

end;

```