

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL

FACULDADE DE ENGENHARIA

IUBERI CARSON ZWETSCH

PROCESSAMENTO DO SINAL DE VOZ PARA O DIAGNÓSTICO  
DIFERENCIAL DE ALTERAÇÕES LARÍNGEAS

Dissertação de Mestrado

PORTO ALEGRE

2006

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

IUBERI CARSON ZWETSCH

PROCESSAMENTO DO SINAL DE VOZ PARA O DIAGNÓSTICO  
DIFERENCIAL DE ALTERAÇÕES LARÍNGEAS

Dissertação apresentada a Pontifícia  
Universidade Católica de Porto Alegre,  
faculdade de Engenharia para obtenção do  
título de mestre em Engenharia Biomédica

Área de concentração: Processamento Digital  
de Sinais

Orientador: Professor Doutor Rubem  
Dutra Ribeiro Fagundes, Ph.D.

Co-orientador: Professora Doutora Thais  
Russomano, Ph.D.

PORTO ALEGRE

2006

## **DEDICATÓRIA**

Dedico este trabalho a minha família pela compreensão e apoio nesta importante etapa de minha vida profissional.

## **AGRADECIMENTOS**

Agradeço ao meu orientador pela confiança e dedicação e pela Instituição Faculdades IPA pelo estímulo e pela bolsa de ajuda financeira que me permitiram concluir este trabalho .

## SUMARIO

Lista de Figuras.....	7
Lista de Tabelas.....	8
<b>Resumo.....</b>	<b>9</b>
<b>Abstract.....</b>	<b>11</b>
<b>1 Introdução.....</b>	<b>12</b>
1.1 Apresentação do tema.....	12
1.2 Objetivos.....	14
1.3 Estrutura da Dissertação.....	14
<b>2 Fundamentação teórica.....</b>	<b>16</b>
2.1 Anatomia e fisiologia do aparelho fonador.....	16
2.1.1 Cartilagens e musculatura da laringe.....	16
2.1.2 Fisiologia da laringe.....	21
2.1.3 Anatomia das Pregas Vocais.....	23
2.1.4 Vibração das pregas vocais.....	25
2.2 Alterações das pregas vocais.....	26
2.2.1 Nódulos de prega vocal.....	26
2.2.2 Pólipos de prega vocal.....	27
2.2.3 Cistos de prega vocal.....	28
2.2.4 Edema de reinke.....	29
2.2.5 Sulco de prega vocal.....	30
2.3 Mecanismos de Produção da voz.....	31
2.3.1 Bases da análise acústica da voz.....	34
2.3.2 Classificação das Formas de Excitação dos Sons da Voz.....	35
2.3.3 Sons Vocálicos.....	35
2.3.4 Sons Não Vocálicos.....	36
2.3.5 Sons Plosivos.....	37
2.3.6 Sons de Excitação Mista.....	38
2.4 Modelo de produção do sinal de voz.....	39
2.5 Processamento dos sinais de voz.....	45
2.5.1 Pré-ênfase.....	45
2.5.2 Janelamento.....	46
2.6 Análise Homomórfica .....	49

2.6.1	Análise Cepstral.....	50
2.6.2	Liftering.....	52
<b>3</b>	<b>Proposta de trabalho.....</b>	<b>54</b>
3.1	Técnica de Processamento de Sinais na identificação de alterações laríngeas.....	54
3.2	Diagrama de blocos de um sistema de reconhecimento de padrões de voz.....	55
3.3	Análise das alterações laríngeas .....	55
3.4	Características e diferenciação das alterações laríngeas.....	56
3.5	Técnica proposta de identificação das alterações laríngeas.....	56
<b>4</b>	<b>Resultados.....</b>	<b>57</b>
4.1	A interpretação dos gráficos obtidos.....	57
4.2	Aplicação da interpretação desenvolvida nos gráficos obtidos.....	57
<b>5</b>	<b>Conclusões e Trabalhos Futuros.....</b>	<b>59</b>
<b>6</b>	<b>Referências Bibliográficas.....</b>	<b>63</b>
	<b>ANEXO A - Parecer do Comitê de Ética em Pesquisa.....</b>	<b>64</b>
	<b>ANEXO B - Consentimento Pós-informado.....</b>	<b>65</b>
	<b>ANEXO C - Declaração de aceitação do artigo.....</b>	<b>67</b>
	<b>ANEXO D - Artigo.....</b>	<b>68</b>

## Lista de Figuras

Figura 1 - Estruturas anatômicas da laringe.....	17
Figura 2 - Estruturas anatômicas da laringe.....	18
Figura 3 - Estruturas anatômicas da laringe.....	19
Figura 4 - Laringe com pregas vocais em posição aduzidas (A) e com pregas vocais em Posição abduzidas(B).....	23
Figura 5 - Desenho esquemático de um corte frontal, representando as estruturas de uma prega vocal humana adulta.....	24
Figura 6 - Nódulo de prega vocal.....	27
Figura 7 - Pólipo de prega vocal.....	28
Figura 8 - Cisto de prega vocal.....	29
Figura 9 - Edema de Reinke.....	30
Figura 10 - Sulco vocal.....	31
Figura 11 - Estruturas envolvidas na produção da voz humana.....	32
Figura 12 - Diagrama em blocos do mecanismo de produção de voz humana.....	33
Figura 13 - Segmento de vogal /a/.....	36
Figura 14 - Forma de onda do fonema /s/.....	37
Figura 15 - Forma de onda do fonema /p/ na fonação de “pa”.....	38
Figura 16 - Forma de onda do fonema /z/.....	38
Figura 17 - Forma de onda do fonema /b/ na fonação de “ba”.....	39
Figura 18 - Diagrama de blocos de modelo de excitações.....	40
Figura 19 - Modelo de tubos acústicos concatenados.....	41
Figura 20 - Diagrama de blocos do modelo de produção da voz.....	44
Figura 21 - Exemplo de sinal de voz segmentado com diferentes comprimentos de janela.....	47
Figura 22 - Função da janela(a) e o espectro de algumas janelas.....	48
Figura 23 - Modelo simplificado da produção da voz no domínio do tempo.....	50
Figura 24 - Diagrama de fluxo da estimação do Cepstro.....	51
Figura 25 - Cepstro de um segmento de fala da vogal /a/.....	52
Figura 26 - Exemplo de uma função de <i>liftering</i> .....	53
Figura 27 - Esquema do exame de videolaringoscopia.....	58

Figura 28 - Esquema de obtenção do sinal acústico.....	58
Figura 29 - Blocos de análise do sinal.....	59

### **Lista de Tabelas**

Tabela 1 - Resumo das principais características de cada método.....	81
--	----

## Resumo

Este trabalho trata do desenvolvimento de técnica de processamento de sinais que usa como princípio a Análise Cepstral do sinal da voz para o desenvolvimento de um sistema de reconhecimento de alterações das pregas vocais, visando obter alternativas para as técnicas habitualmente usadas.

O objetivo é obter um sistema de análise de voz eficiente, que proporcione ferramenta auxiliar no processo diagnóstico e terapêutico, intercambiando áreas da saúde como medicina e fonoaudiologia com a engenharia, para aprimorar a aplicabilidade na obtenção e manipulação dos dados deste tema em comum a estas profissões.

Após uma revisão dos tópicos considerados importantes, para o entendimento de análise do sinal da voz, é realizada uma implementação em duas etapas: uma de análise dos sinais com a técnica Cepstral e após, outra etapa onde as características Cepstrais encontradas em cada caso são identificadas para que se obtenha padrões de reconhecimento das alterações laríngeas.

Na primeira fase os sinais acústicos das alterações vocais são separados em grupos, onde previamente se obteve o diagnóstico exame endoscópico de 61 pessoas, todos adultos, sendo diagnosticados como 17 apresentando cisto de prega vocal, 9 com sulco de prega vocal, 8 com pólipos de prega vocal, 5 com edema de Reinke, 9 com nódulos e 13 normais.

As alterações laríngeas em estudo representam a maior parte dos diagnósticos dos pacientes que consultam com queixas de rouquidão nos consultórios dos otorrinolaringologistas.[23] O método de processamento do sinal utilizado foi o de análise Cepstral como parâmetro de comparação entre os achados de cada alteração laríngea e o isolamento das características significativas do sinal para a aplicação nos métodos de identificação.

Na etapa de identificação foram realizadas a separação das características e a classificação das alterações laríngeas de uma maneira que foi possível ser obtido um protocolo de diagnóstico. Com a técnica aplicada neste trabalho e a similaridade dos padrões obtidos dos sinais das vozes, foi possível obter um reconhecimento automático das alterações laríngeas com taxa de identificação elevada (89 %).

O projeto de identificação das alterações das pregas vocais é, portanto, capaz de discriminar, de forma eficiente, as alterações laríngeas a partir das suas características acústicas apresentadas, independentemente do sexo do paciente e de pequenas variações na alteração laríngea e na obtenção do sinal satisfazendo o que é proposto neste trabalho.

## Abstract

This work deals with the development processing technique of signals that uses as principle the Cepstral Analysis of the voice signal for the development of a recognition system to vocal folds diseases, to be getting at to alternative for the habitually used techniques.

The objective is getting a efficient and reliable voice analysis system, that it provides tool auxiliary in the diagnostic and therapeutical process, interchanging areas of the health as medicine and fonoaudiology with engineering, to improve the applicability in the attainment and manipulation of the common data of this subject to these professions.

After a revision of the considered important topics, for the agreement of the signal voice analysis, an implementation in two stages is carried through: signals analysis with the Cepstral technique, and after, another stage where Cepstrais characteristics joined of each disease are identified so that if it gets of diseases recognition standards.

At the first phase, the vocal folds diseases acoustic signals are separate in groups, where previously the diagnosis through endoscopic examination of 61 individuals was gotten, all adults, being diagnosed as 17 presenting vocal fold cyst, 9 with vocal fold ridge, 8 with vocal fold polyp, 5 with Reinke edema, 9 with nodules and 13 normal ones.

These diseases represent most of the patient diagnostics who consult with complaints of hoarseness in the doctor's offices of the otorrinolaringologists. The method of signal processing used was the Cepstral analysis as comparison parameter between the alterations of each disease and the isolation of the signal significant characteristics for the application in the identification methods.

In the stage of identification it was carried through the characteristics separation and the diseases classification in a way that was possible to be gotten a diagnosis protocol. With the technique applied in this work and the similarity of the voices signals gotten standards, it was possible to get an automatic recognition of the disease with high identification tax (89%).

The project of vocal folds disease identification is, therefore, capable to discriminate, of efficient form, the disease from its characteristics presented acoustics, independently of the patient sex and small variations in the disease and the signal acquisition satisfying what it is considered in this work.

# 1 Introdução

## 1.1 Apresentação do tema

A voz tem relevante importância no processo na comunicação humana, pois é um componente da linguagem oral e responsável por um dos principais vínculos do relacionamento entre os seres humanos.

Quando a voz apresenta um distúrbio, prontamente o locutor, ou o ouvinte, se preocupam, pois não conseguir falar corretamente pode levar a uma dificuldade de comunicação ou até de expressar sentimentos .

O som conhecido como voz humana é originada na laringe, onde se situam as pregas vocais que, com a passagem do ar vindo dos pulmões, vão vibrar e gerar um som que percorrerá a o tubo fonatório até sua exteriorização ao nível dos lábios.

A voz humana apresenta características próprias que variam de acordo com o sexo, a pessoa e idade, além de expressar o estado emocional, reflete o estado e o comportamento da laringe, caracterizando o que se chama de qualidade vocal. Desta maneira, com a análise da qualidade vocal, pode-se conhecer ou inferir sobre condições anátomo-fisiológicas das estruturas laríngeas. As alterações laríngeas das pregas vocais provocam alterações na estrutura da prega vocal que vão se manifestar na qualidade vocal com alterações que podem ser percebidas pelo ouvinte.

A avaliação perceptiva é a primeira ferramenta do médico para o diagnóstico de alteração da laringe, ou seja é uso de avaliação subjetiva, que é fundamental e insubstituível por ser a mais acessível, treinável pelo médico e por não precisar de aparatos.

Porém não devemos descartar a importância das avaliações objetivas, pois estas se completam com as subjetivas tornando as informações diagnósticas mais consistentes.

A comparação de dados se faz necessário para estabelecer em que casos colaboram para um diagnóstico comum ou não. Podemos ter alterações perceptivas que não concordem com a análise acústica e o contrário também.

Novos conceitos e teorias no processamento da voz, estão sendo descobertos. Existem vários pacotes de programas de processamento da voz que já foram bem avaliados

tanto comercialmente como academicamente, por exemplo o MAD (MATLAB *Auditory Demonstrations*), Wrigley, Cooke and Brown, 1999 e o Colea, *Cool Edit* 2000.[1] [2]

Alguns programas são de acesso livre, como por exemplo, o Colea. Entretanto as possíveis ferramentas de análise destes programas são, freqüentemente, muito básicas, limitadas e a análise Cepstral, em geral, não está presente nas suas opções.[3]

Estes programas podem não ser satisfatórios para todas as necessidades de pesquisas com análise de voz, e por isto estudos que aprimoram estas ferramentas de análise acústica são necessários para que se consiga um método computacional mais adequado para o conhecimento das características objetivas da voz.

Também para com este propósito existem experimentos com uso de um microfone colocado na ponta de um equipamento endoscópico flexível que permite a aproximação deste da glote, com o propósito de captar somente o som da vibração das pregas vocais. Este dispositivo é chamado de microfone glotal, mas os resultados não demonstraram grandes vantagens quando realizada a análise do som vocalizado ao nível da saída dos lábios e do microfone glotal.[4]

Um modelo de processamento digital de sinais que determine um método de suporte ao diagnóstico de alterações laríngeas, será muito útil como método auxiliar, apesar do diagnóstico ser realizado quase que exclusivamente pelo exame de videolaringoscopia.

Algumas alterações laríngeas às vezes são muito parecidas no aspecto, apesar de apresentarem origens e alterações fisiopatológicas diferentes.

Existem dificuldades comprovadas na prática médica otorrinolaringológica e também na realização de técnicas computacionais de processamento de sinais em diferenciar as alterações da laringe, pois em determinados casos elas não são eficientes o suficiente para diferenciá-las.

A análise Cepstral do sinal de voz permite trabalhar com o sinal da glote separadamente da região de ressonância, facilitando o estudo dos achados das alterações laríngeas. Os resultados serão analisados e classificados de maneira que se consiga um método de reconhecimento de padrões de alterações laríngeas, esperando obter-se um método eficiente que, no futuro, possa treinar um sistema automático de identificação de alterações laríngeas, como por exemplo, os modelos de Markov tipo HMM (*Hidden Markov Models*) ou as Redes Neurais Artificiais.

Obtendo-se bons resultados será possível determinar mais rapidamente e mais precisamente o diagnóstico da alteração laríngea, de tal maneira que se possa empregar o

tratamento mais adequado, proporcionando uma melhora mais rápida do paciente, já que as alterações laríngicas citadas apresentam tratamentos distintos.

## 1.2 Objetivos

1- Desenvolver de técnica de processamento de sinais da voz, para identificação de alteração laríngea.

2- Criar protocolo de diagnóstico como método de reconhecimento das alterações da laringe a partir da avaliação dos gráficos do sinal fornecido pelo processamento, que servirá de base para o treinamento de algoritmo de reconhecimento de padrões.

## 1.3 Estrutura da dissertação

Esta dissertação foi estruturada em 6 capítulos. Este Capítulo (1) tem por objetivo permitir ao leitor uma visão ampla da dissertação e das características das alterações laríngicas das pregas vocais e ao mesmo tempo procura focalizar sua atenção no objetivo principal deste trabalho que é uso do processamento digital de sinais para o diagnóstico das alterações laríngicas.

Também, esta seção apresenta uma breve visualização dos demais capítulos deste documento, mostrando nos parágrafos a seguir a descrição sucinta desses capítulos.

O Capítulo 2 é dedicado a fundamentação teórica dos tópicos envolvidos no projeto, onde serão descritos: o mecanismo de produção da voz e o seu modelo correspondente, as alterações laríngicas, as etapas do processamento de sinais, o método de Análise Cepstral e o seu uso com reconhecimento de padrões de alterações laríngicas.

No Capítulo 3 é realizada a descrição da proposta do trabalho no uso das técnicas de análise para a identificação das alterações das pregas vocais.

No Capítulo 4 são apresentados os métodos empregados na formulação da pesquisa, o *corpus* de dados usado na investigação, a descrição das técnicas desenvolvidas e os equipamentos utilizados no processo de análise do sinal de voz.

O Capítulo 5 apresenta os resultados da análise dos gráficos e tabelas obtidos e como foi desenvolvido o sistema de identificação das alterações laríngicas das pregas vocais. O Capítulo 6 apresenta as conclusões obtidas com base nos resultados do Capítulo 5, demonstrando a eficiência do protocolo de diagnóstico desenvolvido, bem como em que

condições ele deve ser utilizado. Serão também apresentadas condições e sugestões para trabalhos futuros.

Em seguida são apresentadas as Referências Bibliográficas.

No Anexo A, temos o Parecer do Comitê de Ética em Pesquisa, no Anexo B o Consentimento Pós-informado, conforme as normas da Pró-reitoria de Pesquisa e Pós-graduação da Pontifícia universidade Católica do Rio Grande do sul. No Anexo C está a Declaração de Aceitação do artigo e no Anexo D, o artigo.

## **2 Fundamentação teórica**

A revisão da anatomia e da fisiologia do aparelho fonador permite um melhor entendimento da repercussão das alterações laríngeas das pregas vocais sobre a acústica da voz.

### **2.1 Anatomia e fisiologia do aparelho fonador**

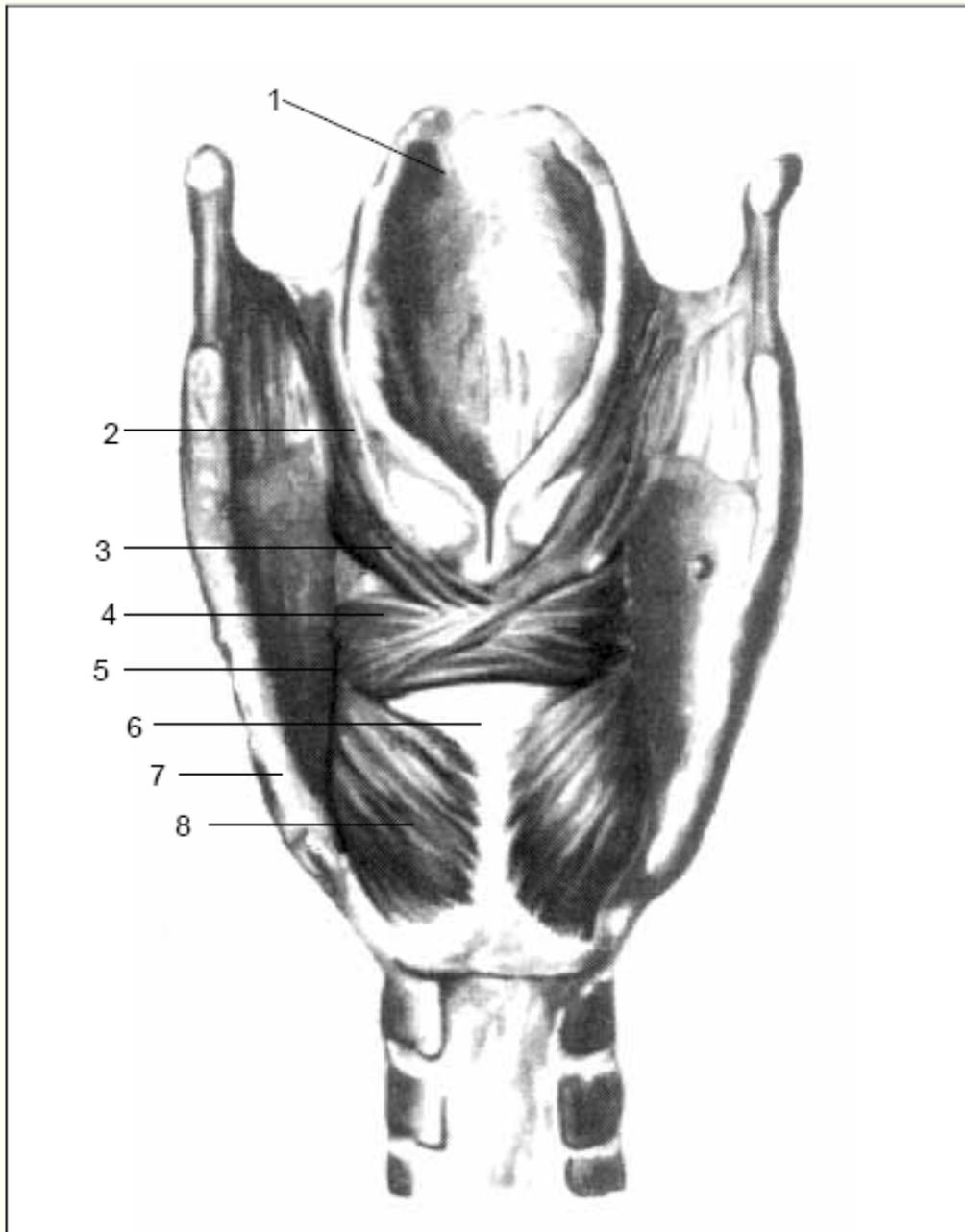
O aparelho fonador apresenta várias estruturas importantes que vão participar na formação sonora. A laringe é o principal órgão do aparelho fonador por que abriga as pregas vocais, que são as estruturas vibrantes responsáveis pela produção sonora. Serão destacados os aspectos considerados mais importantes das estruturas para a melhor compreensão das alterações vocais de interesse no trabalho.

#### **2.1.1 Cartilagens e musculatura da laringe**

##### **Cartilagens da laringe**

A laringe é uma estrutura que apresenta um esqueleto de cartilagens hialinas (figuras 1,2 e 3), sendo que as mais importantes relacionadas com a fonação são a tireóidea, cricóidea e aritenóideas.

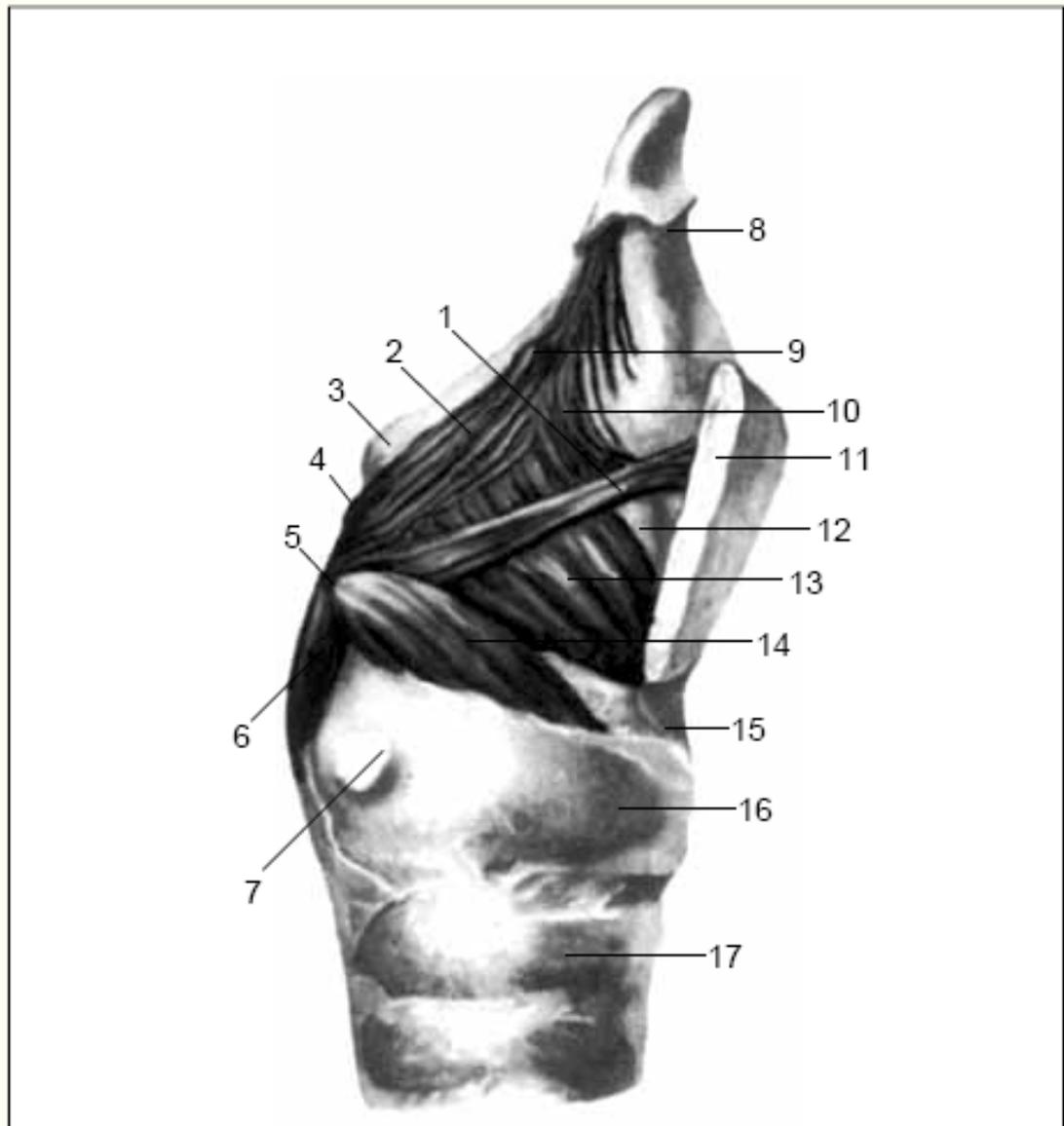
A cartilagem tireóidea é a maior e a cricóidea é a segunda maior cartilagem laríngea. A cartilagem tireóidea articula-se com a cricóidea sobre sua superfície póstero-lateral, permitindo assim, movimento de rotação e um pequeno movimento ântero-posterior.



Músculos posteriores da laringe: músculo cricoaritenóideo posterior – dilatador; músculos aritenóideos transverso e oblíquo – constritores da rima vocal. 1 = Epiglote; 2 = músculo ariepiglótico; 3 = músculo aritenóideo oblíquo; 4 = músculo aritenóideo transverso; 5 = processo muscular cartilaginoso aritenóideo; 6 = lâmina cartilaginosa cricóidea; 7 = cornu inferior da cartilagem tireóidea; 8 = músculo cricoaritenóideo posterior.

Figura 1- Estruturas anatômicas da laringe.

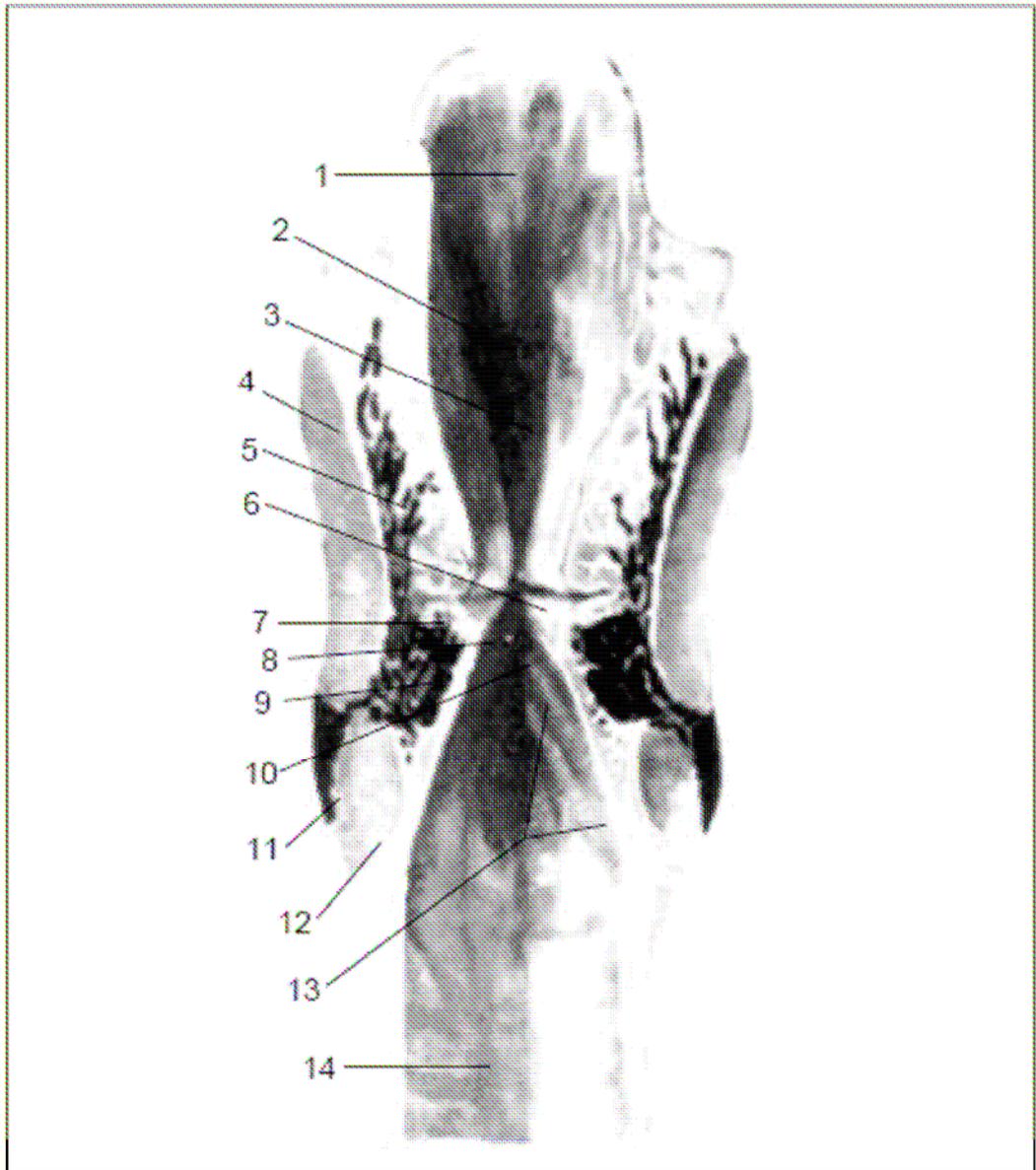
Fonte: Otacílio L.F. e Carlos A. H. de C.; Tratado de Otorrinolaringologia (1994, p 982)



Músculos intrínsecos da laringe. Vista lateral. 1= Fascículo muscular acessório; 2 = músculo ariepiglótico; 3 = cartilagem corniculada; 4 = músculo aritenóideo oblíquo; 5 = processo muscular (cartilagem aritenóidea); 6 = músculo cricotireóideo posterior; 7 = face articular tireóidea; 8 = epiglote; 9 = músculo estilofaríngeo (ressecado); 10 = músculo tireoepiglótico; 11 = cartilagem tireóidea da lâmina direita (ressecada); 12 = sáculo laríngeo; 13 = músculo tireoaritenóideo; 14 = músculo cricoaritenóideo lateral; 15 = ligamento cricoaritenóideo; 16 = arco da cartilagem cricóidea; 17 = traquéia.

Figura 2- Estruturas anatômicas da laringe.

Fonte: Otacílio L.F. e Carlos A. H. de C.; Tratado de Otorrinolaringologia (1994, p 983)



### Corte frontal da laringe

1-Epiglote; 2-vestíbulo laríngeo; 3-tubérculo da epiglote; 4- lâmina cartilaginosa tireóidea; 5- músculo tireoaritnoideo; 6-prega vestibular; 7-ventrículo laríngeo; 8- prega vocal; 9-músculo vocal; 10-rima glótica; 11- músculo cricotireioideo; 12-arco cartilaginoso cricotireoideo; 13- região infraglótica (cone elástico); 14-traquéia

Figura 3- Estruturas anatômicas da laringe.

Fonte: Otacílio L.F. e Carlos A. H. de C.; Tratado de Otorrinolaringologia (1994, p 985)

Há duas cartilagens aritenóideas, cada uma posicionada sobre um lado da linha média, sobre a superfície supra-posterior da cartilagem cricóidea. A base horizontal da cartilagem aritenóidea, que se articula com a cricóidea, é côncava, onde estas cartilagens se articulam.

Cada cartilagem aritenóidea apresenta uma região chamada de processo vocal, sendo este o ponto de fixação posterior da prega vocal. Um outro ponto relevante é o processo muscular onde se fixam vários músculos.

Também existe uma cartilagem em forma de folha chamada de epliglote que se fixa sobre a superfície medial da cartilagem tireóidea. Essa cartilagem auxilia no direcionamento de alimentos ao esôfago durante a deglutição contribuindo para evitar a entrada de alimentos na via respiratória inferior. Na fonação, ela se movimenta durante a produção de diferentes vogais e consoantes, podendo dificultar a visualização das pregas vocais quando são empregadas técnicas de exames com introdução de equipamentos diagnósticos pela boca da pessoa examinada, como por exemplo o uso da ótica rígida na realização de videolaringoscopia diagnóstica

Outras cartilagens menores são as corniculadas e as cuneiformes.

## **Musculatura da laringe**

Os músculos da laringe são divididos em extrínsecos e os intrínsecos. São oito músculos extrínsecos separados em dois grupos: supra-hióides e infra-hióides. Todos os músculos extrínsecos têm suas fixações laríngeas a partir do osso hióide. As fixações fora da laringe incluem várias estruturas, tais como a mandíbula, mastóide e estruturas do tórax.

O grupo supra-hióideo pode proporcionar movimentos como tracionar a laringe para frente.

No grupo infra-hióideo os músculos podem produzir ações como “puxar” a laringe para baixo e alterar o ângulo da cartilagem tireóidea com relação à cartilagem cricóidea.

São cinco músculos intrínsecos da laringe, cujas fixações encontram-se associadas a estruturas dentro da laringe.

O músculo aritenóideo tem duas partes e posiciona-se entre as duas cartilagens aritenóideas. A sua atividade pode promover a adução (aproximação das pregas vocais) fundamental na fonação e também a aproximação das aritenóides, fechando a via aérea posterior, fundamental para a sobrevivência, por exemplo, na deglutição.

O músculo cricoaritenóideo lateral é pareado, e pode puxar o processo muscular da cartilagem aritenóidea anteriormente, deslocando-a medialmente e aduzindo as próprias pregas vocais. Os músculos aritenóideo e o cricoaritenóideo lateral atuam juntos para aduzir as pregas vocais e o cricoaritenóideo posterior é o único que abduz (separação das pregas vocais)

as pregas vocais, pois um movimento de curta duração de sua atividade, rapidamente abduz as pregas vocais o suficiente para cessar sua vibração.

O músculo cricotiróideo possui formato de leque e localiza-se entre as cartilagens cricóide e tireóide. Sua função consiste em diminuir e aumentar o espaço entre a tiróide e a cricóide; diminuir a massa das pregas vocais; e aumentar sua tensão, sendo que estas atividades têm como principal objetivo alterar a frequência vocal.

O músculo tiroaritenóideo compõe o corpo das pregas vocais, tem a forma de um grosso feixe, e sua ação é encurtar e aduzir as pregas vocais, diminuindo a distância entre as cartilagens aritenóideas e tireóidea, tornando-se um feixe mais largo e reduzindo a frequência da voz gerada. Estes músculos também podem ser visualizados na Figura 1,2 e 3.

### **2.1.3 Fisiologia da laringe**

As funções mais importantes são: a respiratória, a deglutitória e a fonatória.

#### **Função Respiratória**

Esta função ocupa a maior parte do tempo desse órgão. A abertura da laringe permite a entrada e a saída do ar dos pulmões. Durante a inspiração a laringe é ligeiramente rebaixada, passando a ser deslocada ligeiramente para cima durante a expiração; esse deslocamento é proporcional à intensidade da respiração. A traquéia e o diafragma participam da descida da laringe, contribuindo, junto com o músculo cricoaritenóideo posterior, para o afastamento das aritenóideas e das pregas vocais. Os músculos infra-hióideos contribuem para a descida da cartilagem tireóidea e, conseqüentemente, para a abertura da laringe.

#### **Função Deglutitória**

Primeiramente a deglutição ocorre na boca e é voluntária, sendo seguida de uma fase involuntária faríngea, onde ocorre o fechamento máximo da laringe.

Durante a deglutição, na fase faríngea, a luz laríngea se fecha de baixo para cima protegendo as vias aéreas, sendo essa proteção uma função vital.

## Função Fonatória

A fonação é uma habilidade adquirida pela espécie humana.[9] Com a evolução, o ser humano aprendeu a coordenar órgãos que, fisiológica e anatomicamente apresentavam a respiração, alimentação e proteção das vias aéreas inferiores como funções primordiais. A voz não é produzida por um único órgão, e sim pela ação integrada e harmoniosa de vários sistemas, que agem ativamente no processo de produção da voz.

Podemos destacar a participação do sistema respiratório dentre eles, que se adaptou para se torne ativo na vocalização. Ao emitir a voz, as pregas vocais se aproximam da linha média, controlando a saída do ar que advém dos pulmões. O ar, por sua vez, ao passar pela laringe, coloca em vibração as pregas vocais, que estão próximas entre si. As pregas vocais, neste instante, se aproximam e se afastam numa seqüência rápida, realizando a função fonatória, Figura 4.

A freqüência de vibração é determinada pela elasticidade ou tensão em que vão se encontrar as pregas vocais durante a produção do som, onde teremos a alteração da espessura e alongamento ativo das mesmas pelas estruturas da laringe, melhor demonstradas na secção da anatomia da laringe.

Um som de freqüência alta significa uma vibração rápida das pregas vocais e a percepção acústica do ouvinte será de um som mais agudo. A intensidade da produção da voz é determinada pela intensidade de vibração das pregas vocais, associada ao nível de pressão subglótica.[9]

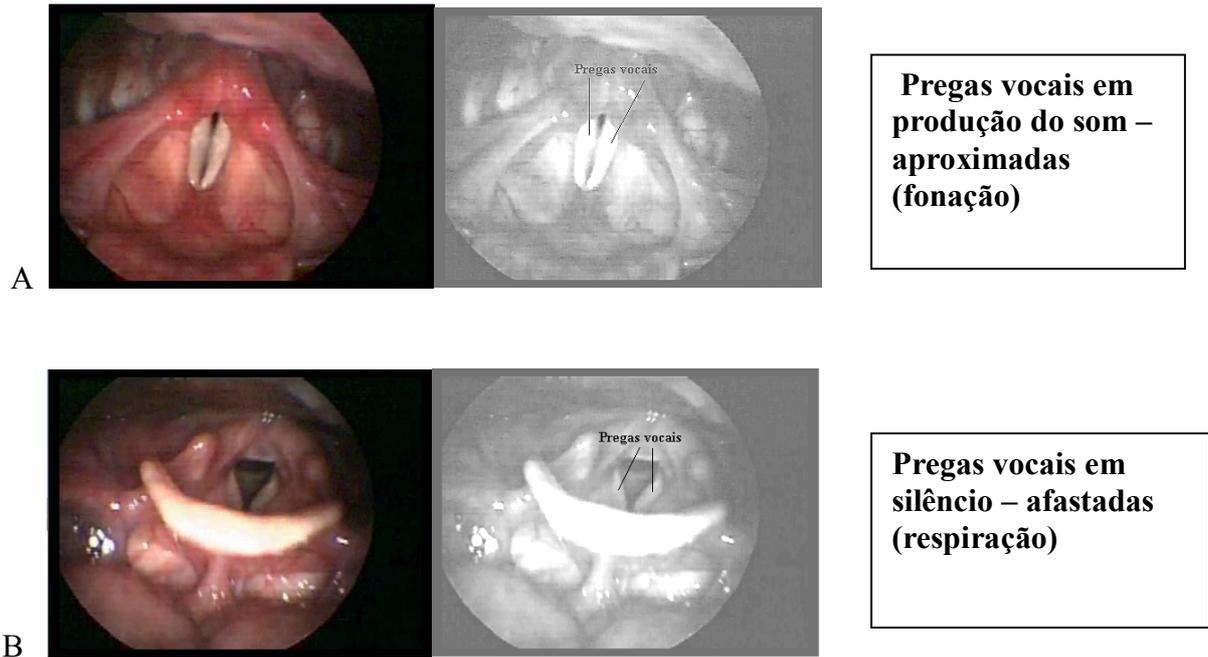


Figura 4 - Laringe com pregas vocais em posição aduzidas (A) e com pregas vocais em posição abduzidas (B).

#### 2.1.4 Anatomia das pregas vocais

O conhecimento da composição estrutural das pregas vocais é fundamental para se entender como os desvios na estrutura normal, resultantes das alterações laríngeas, causam alterações nas características vibratórias. Considerando-se o funcionamento vocal, o conceito mais importante sobre a estrutura da prega vocal é que esta se comporta como um vibrador com várias camadas epiteliais. De um modo geral, pode-se dizer que a prega vocal tem o comportamento de uma lâmina vibrante composta de mucosa e músculo. A prega vocal apresenta várias camadas envolvidas nos fenômenos vibratórios e na Figura 5 são vistas como estão dispostas.

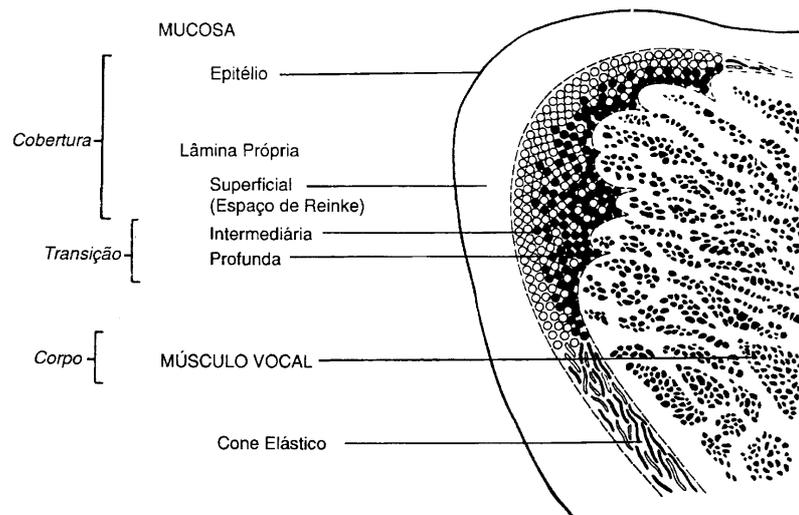


Figura 5 - Desenho esquemático de um corte frontal, representando as estruturas de uma prega vocal humana adulta.

Fonte: MINORU HIRANO e DIANE M. B.; Exame videoestroboscópico da laringe; Porto Alegre: Artes Médicas (1997, p36)

Na prega vocal humana, a área na borda da prega vocal e em torno dela é a que apresenta um movimento acentuado durante a fonação. Do mesmo modo, é nesta área que uma lesão da fará a maior diferença perceptiva na produção da voz.

Do ponto de vista histológico, esta área consiste de cinco camadas:

- 1- Epitélio da mucosa, constituído de células escamosas.  
Do ponto de vista mecânico: pode ser considerado um escudo fino cujo objetivo é manter o formato da prega vocal.
- 2- Camada superficial da lâmina própria da mucosa: consiste basicamente de fibras soltas e tecido matricial. É referida como espaço de Reinke.  
Do ponto de vista mecânico: esta camada é bastante flexível, podendo ser considerada como algo semelhante a uma massa mole de gelatina e é a que vibra mais intensamente na fonação.
- 3- Camada intermediária da lâmina própria da mucosa, composta basicamente de fibras elásticas.  
Do ponto de vista mecânico: esta camada pode ser comparada a um feixe de elásticos macios.

4- Camada profunda da lâmina própria é composta principalmente de fibras de colágeno (tecido elástico).

Do ponto de vista mecânico: esta camada é semelhante a um feixe de fios de algodão.

5- Músculo vocal, são fibras musculares dispostas de modo aproximadamente paralelo à borda da prega vocal, que constitui o principal corpo da prega vocal.

Do ponto de vista mecânico: este músculo que se contrai é como um feixe de tiras de borracha bastante rígida, em que a rigidez pode ser alterada de acordo com o grau de contração muscular.[22]

As estruturas descritas podem ser agrupadas em três seções; a cobertura, (consistindo do epitélio e da camada superficial da lâmina própria); a transição (compostas pelas camadas intermediária e profunda da lâmina própria); e o corpo (composto do músculo vocal), Figura 5.

### **2.1.5 Vibração das pregas vocais**

A vibração das pregas vocais pode ser alterada, gerando novos modos de vibração em decorrência de mudanças na simetria da sua massa, ocorrendo variações da elasticidade e tensão nas próprias pregas vocais. Também podemos ter alterações vibratórias pela presença de lesões chamadas de efeito de massa como por exemplo os nódulos, cistos e pólipos e edema de Reinke. O sulco vocal, entretanto é uma lesão em que há perda de massa da prega vocal, mas também provoca importantes alterações vibratórias. Outras alterações, como paralisia dos músculos respiratórios, podem causar pressão subglótica insuficiente, alterando as forças aerodinâmicas que agem sobre as pregas vocais o que, também, alterará as vibrações.

A conceituação de vozes normais e alteradas é muito variada, visto que, implica uma certa subjetividade. A normalidade vocal implica na dependência de vários fatores, como normalidade anatômica e funcional, boa coordenação entre respiração e fonação, boa saúde geral, ausência de problemas psicológicos relevantes e emissão agradável.[9]

A voz deve ser produzida sem esforço e com conforto, identificando o sexo e a faixa etária a que pertence. Encontra-se na literatura uma variedade de classificações destinadas a conceituar as alterações vocais (disfonias).[9]

Os sintomas acústicos percebidos durante o processo patológico da laringe são mudanças na frequência fundamental, na intensidade da voz ou em sua qualidade.

Esses sintomas são indicativos de alterações orgânicas e/ou funcionais e a natureza desses sintomas varia para cada paciente e também com o estágio da alteração laríngea.

As alterações da laringe que alteram a qualidade vocal são muitas, mas neste trabalho foram avaliadas as seguintes: os nódulos, pólipos, cistos, sulco vocal e edema de Reinke.

Estatisticamente, este grupo inclui as alterações laríngeas mais freqüentemente diagnosticadas no consultório otorrinolaringológico.[23]

## **2.2 Alterações das pregas vocais**

As alterações laríngeas são diagnosticadas na sua grande maioria das vezes pelo exame de videolaringoscopia, porém o seu diagnóstico de certeza é obtido com a análise de microscopia patológica através de uma biópsia para extração de fragmento da lesão. As alterações incluídas no trabalho são descritas a seguir. Serão comentadas suas principais características, como: origem, aspecto, alteração vibratória produzida nas pregas vocais, sensação acústica percebida e sua incidência.[22]

### **2.2.1 Nódulo de prega vocal**

Origem: os nódulos das pregas vocais (Figura 6) resultam de abuso vocal crônico.

Aspecto: a lesão é normalmente bilateral e simétrica.

A lesão localiza-se na camada superficial da lâmina própria. As pregas vocais não fecham anterior e posteriormente aos nódulos. A estrutura de cada prega vocal é ligeiramente heterogênea devido à existência do nódulo. A borda da prega vocal é não-linear, porque o nódulo produz um pequeno abaulamento.

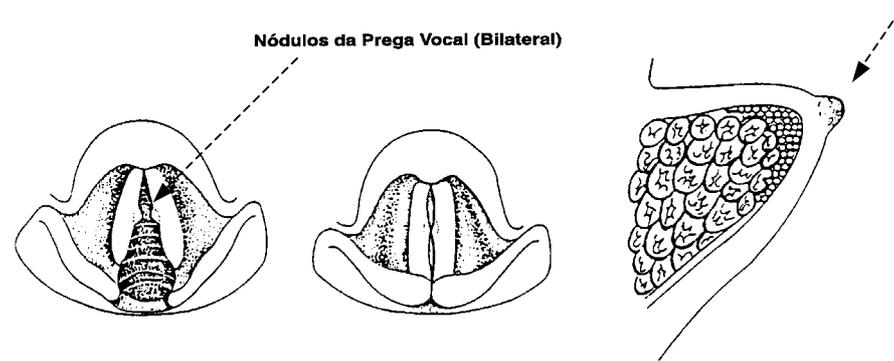


Figura 6 - Nódulo de prega vocal.

Fonte: Modificado de MINORU HIRANO e DIANE M. B.; Exame videostroboscópico da laringe; Porto Alegre; Artes Médicas (1997, p146)

Alterações vibratórias: a natureza da rigidez, da cobertura varia, dependendo da característica histológica — ou seja, a rigidez aumenta nos nódulos fibrosos, ao passo que diminui nos nódulos edematosos. A massa da cobertura aumenta ligeiramente. Durante a vibração, o fechamento glótico é incompleto, surgindo uma fenda vocal. A onda mucosa é ausente no nódulo, quando este é fibroso e firme. Se o nódulo é edematoso e macio, a onda mucosa é muitas vezes observada no mesmo.

Sensação acústica: rouquidão variável, soprosidade, aspereza, *pitch* grave (sensação de frequência, definição abordada com maiores detalhes na página 36) e pode ter diplofonia.

Incidência: entre lesões não neoplásicas é de 16%. [23]

### 2.2.2 Pólipo de prega vocal

Origem: os pólipos das pregas vocais (Figura 7) também podem ser o resultado de abuso vocal, bem como de uma série de outros fatores.

Aspecto: a lesão pode ser unilateral ou bilateral, mas uma lesão unilateral é mais freqüente. Quando bilateral, os pólipos geralmente são assimétricos. São localizados na camada superficial da lâmina própria. Em muitos casos, a glote não fecha anterior e posteriormente ao pólipo. As pregas vocais são assimétricas, e a prega vocal afetada apresenta estrutura heterogênea. A borda da prega vocal é não-linear, como resultado da configuração criada pelo pólipo.

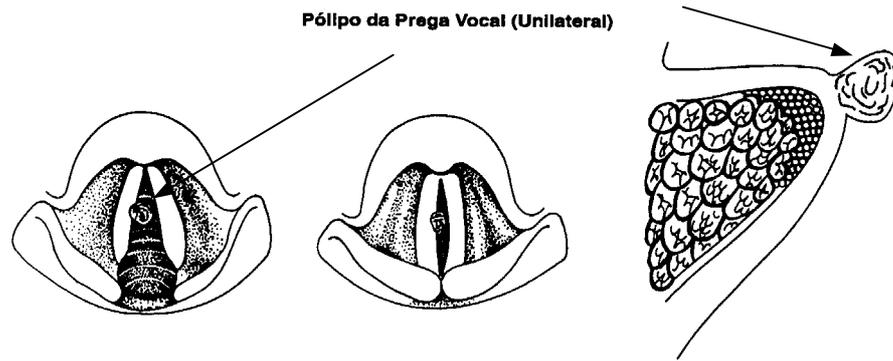


Figura 7 - Pólipo de prega vocal.

Fonte: Modificado de MINORU HIRANO e DIANE M. B.; Exame videoestroboscópico da laringe; Porto Alegre; Artes Médicas (1997, p 153)

Alterações vibratórias: a rigidez da cobertura varia: quando o pólipo é hemorrágico ou fibroso, a rigidez aumenta; se o pólipo é edematoso, a rigidez diminui. A massa da cobertura aumenta. O pólipo pode interferir nos movimentos vibratórios da prega vocal contralateral.

Durante a vibração, o fechamento glótico é incompleto, apresentando fendas anterior e posteriormente ao pólipo. Os movimentos vibratórios das pregas vocais são normalmente assimétricos. Os pólipos em geral se movem com um ligeiro atraso em relação à prega vocal, apresentando uma vibração casada, ou de pêndulo duplo. As vibrações sucessivas são muitas vezes aperiódicas. A amplitude da prega vocal afetada é pequena e a da prega vocal contralateral encontra-se, muitas vezes, ligeiramente reduzida, devido à interferência causada pelo pólipo. A onda mucosa está normalmente ausente quando o pólipo é hemorrágico ou fibroso, mas é de tamanho normal ou maior se o pólipo é edematoso e flexível.

Sensação acústica: *pitch* grave e intensidade de rouquidão dependendo do tamanho.

Incidência: entre lesões não neoplásicas é de 7%. [23]

### 2.2.3 Cisto de prega vocal

Origem: na maioria dos casos são cistos epidermóides ou cistos de retenção e quase sempre são unilaterais e podem ser congênitos ou adquiridos.

Aspecto: a patologia localiza-se na camada superficial da lâmina própria, (Figura 8). O cisto parece-se com um pequeno balão repleto de fluido. As pregas vocais, muitas vezes, não se fecham anterior e posteriormente ao cisto.

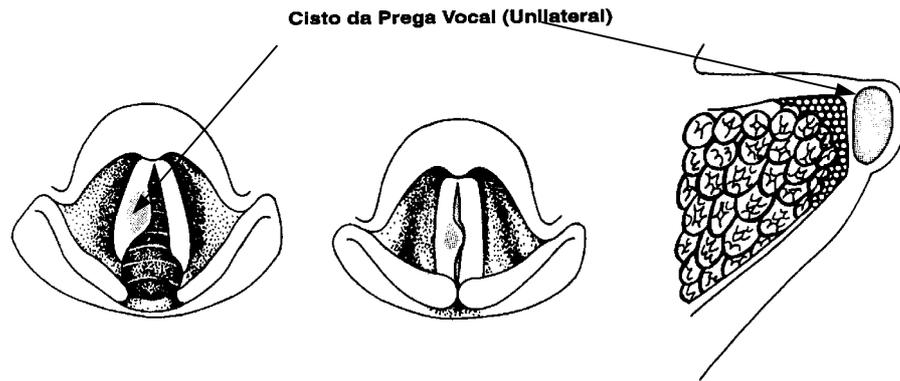


Figura 8 - Cisto de prega vocal.

Fonte: Modificado de MINORU HIRANO e DIANE M. B.; Exame videoestroboscópico da laringe; Porto Alegre; Artes Médicas (1997, p160)

Alterações vibratórias: as propriedades mecânicas são assimétricas entre as duas pregas vocais e heterogêneas na prega vocal afetada. A borda da prega vocal é não-linear devido ao inchaço e protuberância causada pelo cisto. A rigidez e a massa da cobertura encontram-se aumentadas. Geralmente, cistos são mais rígidos que nódulos e pólipos. O cisto age como um obstáculo para a prega vocal contralateral.

O fechamento glótico é, muitas vezes, incompleto durante a vibração. Pequenas fendas permanecem anterior e posteriormente ao cisto, durante o fechamento. Os movimentos vibratórios das duas pregas vocais são assimétricos. As vibrações sucessivas são ocasionalmente aperiódicas. A amplitude da excursão lateral é muito pequena no lado afetado e pode não se observar onda mucosa próxima ao cisto. Nos cistos epidérmicos, em geral, as limitações da vibração são maiores do que nos nódulos e pólipos.

Sensação acústica: *pitch* grave, diplofonia, bitonalidade, aspereza e rouquidão moderada.

Incidência: entre lesões não neoplásicas é de 24%. [23]

#### 2.2.4 Edema de Reinke

Origem: a etiologia desta patologia não é muito clara, mas geralmente se desenvolve em fumantes de meia idade e mais freqüente em mulheres.

Aspecto: a prega vocal fica edemaciada, às vezes muito inchada, normalmente bilateral, mas muitas vezes assimétrica (Figura 9) e localiza-se na camada superficial da lâmina própria (espaço de Reinke).

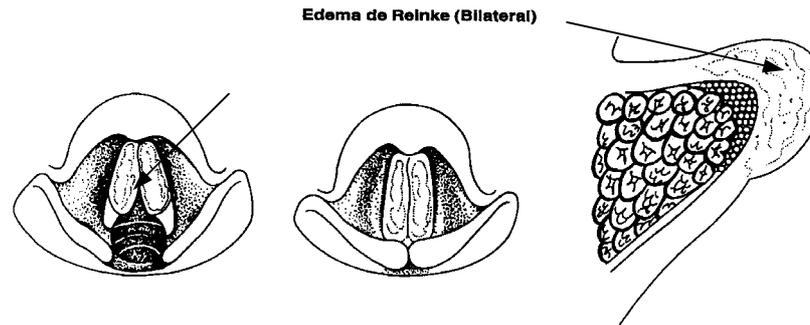


Figura 9 - Edema de Reinke.

Fonte: Modificado de MINORU HIRANO e DIANE M. B.; Exame videostroboscópico da laringe; Porto Alegre; Artes Médicas (1997, p157)

Alterações vibratórias: a rigidez da cobertura é diminuída, enquanto que sua massa e a profundidade da borda vibratória aumentam. O intumescimento edematoso interfere no movimento vibratório da prega vocal contralateral.

Em geral, durante a vibração, a glote fica completamente fechada. Os movimentos das pregas vocais bilaterais são assimétricos, e as vibrações sucessivas são muitas vezes aperiódicas. A amplitude da excursão horizontal é com frequência pequena, mas a onda mucosa é em geral acentuadamente grande.

Sensação acústica: de *pitch* muito grave e rouquidão intensa.

Incidência: entre lesões não neoplásicas é de 10%. [23]

### 2.3.5 Sulco de prega vocal

Origem: a expressão “sulco vocal” diz respeito a uma condição na qual um sulco, ao longo da borda de uma ou de ambas pregas vocais, causa distúrbios de voz. A etiologia desta alteração laríngea é desconhecida, mas acredita-se que seja congênita em alguns casos e secundária a processos inflamatórios crônicos repetidos ou de longa duração em outros.

Aspecto: a lesão é geralmente bilateral, e o sulco localiza-se na borda da prega vocal, normalmente percorrendo a extensão da prega. A lesão localiza-se na camada superficial da lâmina própria (Figura 10). O fechamento das pregas vocais em geral é incompetente, apresentando uma fenda fusiforme estreita durante a fonação, porque as bordas das pregas vocais podem ficar arqueadas.

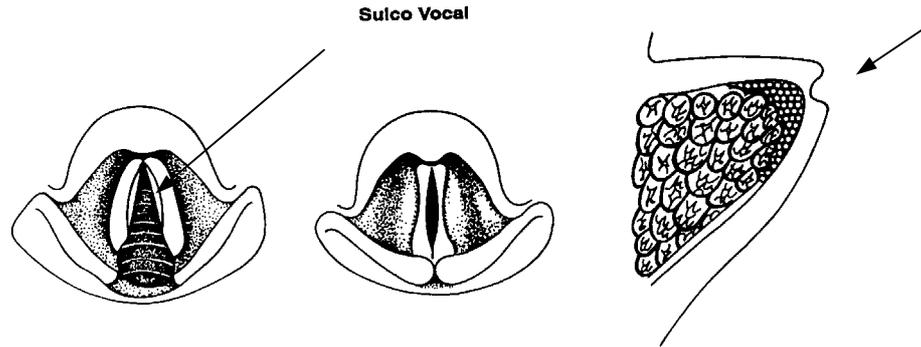


Figura 10 - Sulco vocal.

Fonte: Modificado de MINORU HIRANO e DIANE M. B.; Exame videostroboscópico da laringe; Porto Alegre; Artes Médicas (1997, p163)

Alterações vibratórias: a rigidez da cobertura encontra-se aumentada no sulco, e a massa da cobertura é ligeiramente diminuída. Durante os ciclos vibratórios, a excursão horizontal das pregas vocais é, em geral, reduzida. A onda mucosa é interrompida no sulco.

Sensação acústica: aspereza (a mais característica, devido à rigidez), *pitch* fica agudo, sopro, rouquidão e pode ter bitonalidade.

Incidência: entre lesões não neoplásicas é de 10%. [23]

A soma da incidência destas alterações laríngeas não neoplásicas é de 67%, representando assim a maior parte das alterações laríngeas que provocam sintomas de rouquidão.

### 2.3 Mecanismos de produção da voz

A voz é um sinal produzido como resultado de várias transformações que ocorrem em diferentes níveis: semântico, lingüístico, articulatório e acústico. As diferenças nessas transformações aparecem como mudanças nas propriedades acústicas do sinal de voz.

As características das vozes de locutores são um resultado da combinação das diferenças anatômicas inerentes ao trato vocal (características inerentes) e daquelas relacionadas ao movimento dinâmico do trato vocal, ou seja, a forma como a pessoa fala (características instruídas). Para gerar o som desejado, o falante exerce uma série de controles sobre o aparelho fonador, representado na Figura 11, produzindo a configuração articulatória e excitação apropriada. A Figura 11 também evidencia as características importantes do sistema vocal humano. O trato vocal é um nome genérico dado ao conjunto de cavidades e estruturas que participam diretamente na produção sonora, começa na abertura entre as pregas

vocais, ou glote e termina nos lábios. O trato vocal assim, consiste da faringe (a conexão entre o esôfago e a boca) e termina na boca, ou cavidade oral. O trato nasal começa na úvula e termina nas narinas. Quando a úvula é abaixada, o trato nasal é acusticamente acoplado ao trato vocal para produzir os sons nasais da voz. Verifica-se que o trato nasal, pode voluntariamente ser acionado ou não pelo locutor.

Após a filtragem, determinada pela conformação do aparelho fonador, o fluxo de ar injetado pelos pulmões é acoplado ao ambiente externo através dos orifícios dos lábios e/ou narinas.[5]

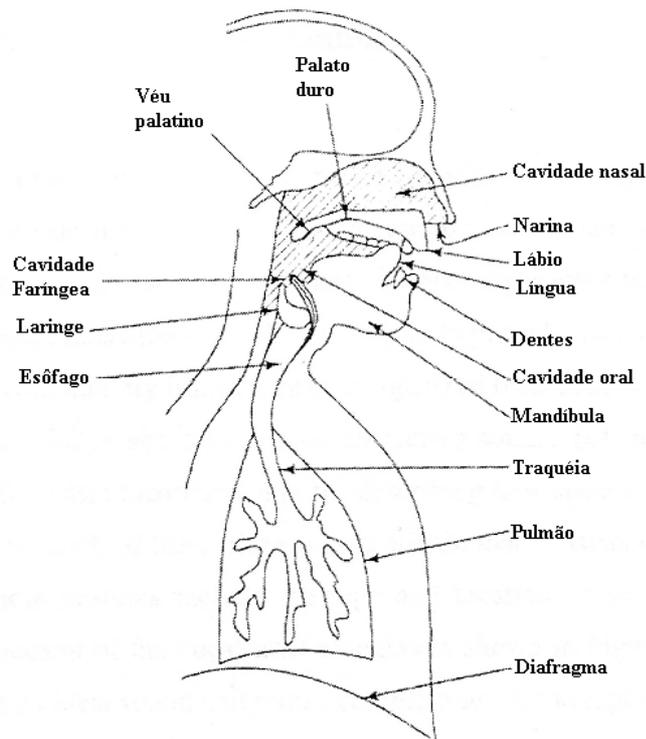


Figura 11 - Estruturas envolvidas na produção da voz humana.  
 Fonte: Modificado de DELLER, J. R. Jr.; HANSEN, J.H.L. and PROAKIS, J.G.; Discrete-Time Processing of Speech Signals, *Institit of Eletrical and Eletronics Engineers*, Press.(2000. p101)

Na Figura 12 é apresentado um modelo mecânico para a produção de voz. Nesse modelo os tratos oral e nasal são representados por tubos acusticamente acoplados.

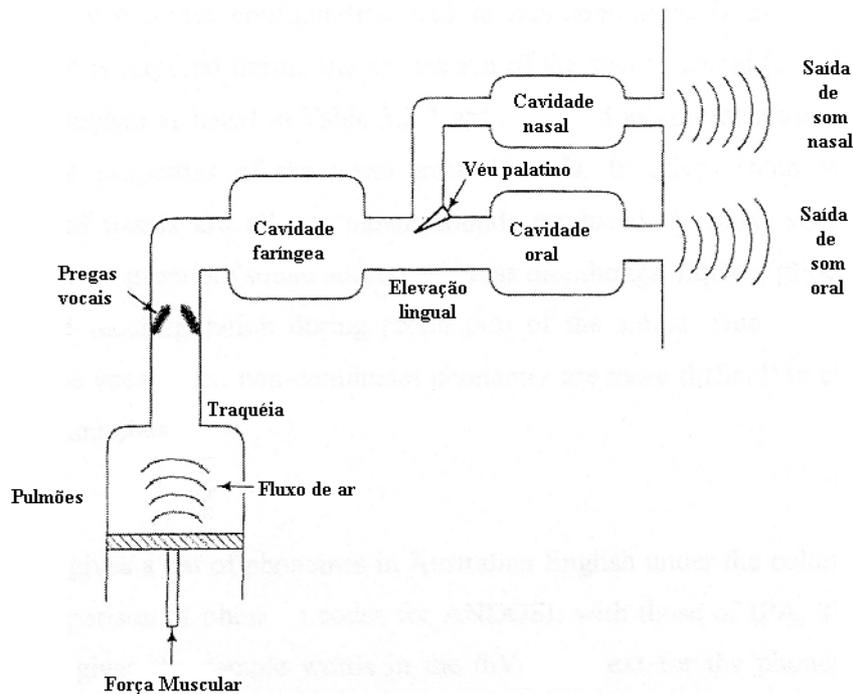


Figura 12 - Diagrama em blocos do mecanismo de produção de voz humana  
 Fonte: Modificado de DELLER, J. R. Jr.; HANSEN, J.H.L. and PROAKIS, J.G.;  
 Discrete-Time Processing of Speech Signals, *Institot of Eletrical and Eletronics  
 Engineers*, Press.(2000. p102)

O diagrama completo inclui o sistema subglotal composto dos pulmões, brônquios e traquéia. O sistema subglotal funciona como uma fonte de energia para produção da voz. A voz é a onda acústica radiada do sistema quando o ar é expelido dos pulmões.[11]

O trato vocal e o trato nasal podem ser vistos como tubos de seção transversal não uniformes. O som se propaga através desses tubos e o espectro de frequência é modelado pela seletividade de frequência do tubo. Esse efeito é muito similar aos efeitos de ressonância que podem ser observados em instrumentos de sopro. No contexto da produção da voz, as frequências de ressonância do tubo do trato vocal são chamadas de frequências formantes ou simplesmente formantes. As frequências formantes dependem sobretudo da forma e dimensões do trato vocal. Cada forma é caracterizada por um conjunto de frequências formantes. Sons diferentes são formados em função das variações da forma assumida pelo trato vocal. Assim, as propriedades espectrais do sinal de voz variam com o tempo e com a forma do trato vocal.[11]

Se o ouvinte decodificar de forma correta a seqüência de sons emitida, a cadeia de comunicação se completará, fechando o ciclo, que compreende desde a concepção da idéia até sua completa assimilação pelo interlocutor.

Diante do exposto, ao se realizar a análise acústica, deve-se obter informações que permitam a melhor compreensão das formas de excitação do aparelho fonador, bem como os métodos mais adequados para avaliar os parâmetros temporais do sinal de voz que vão ser a base para a implementação de modelo para um sistema de reconhecimento de alterações laríngeas das pregas vocais. Um método que pode colaborar com isto é o sistema de processamento de sinais através da análise Cepstral que permite a separação das regiões de ressonância da excitação e que será descrito com mais detalhes no decorrer do trabalho.

### **2.3.1 Bases da análise acústica da voz**

O sinal de voz tem características espectrais que são variantes no tempo (ou não estacionárias), visto que este sistema físico varia com o tempo. Podemos dividir o sinal de voz em segmentos que possuem propriedades acústicas semelhantes para curtos intervalos de tempo em duas categorias básicas:

- 1- vogais que quase não apresentam restrição à passagem do ar através do trato vocal e
- 2- consoantes que apresentam uma maior restrição à passagem do ar e são, em geral, mais “fracas” em amplitude e podem ser semelhantes a uma fonte de ruído. Algumas das diferenças entre vogais e consoantes são evidentes visualizando a forma de onda do sinal, por exemplo, na Figura 13 tem-se a onda do sinal acústico de uma vogal.

A observação das formas das ondas ajuda verificar o que estas podem revelar sobre os aspectos acústicos da voz.

A voz humana apresenta muitas características básicas do sinal, tais como: periodicidade, intensidade, duração, etc. Uma das mais importantes características da voz, é que a voz não é constituída por sons discretos bem definidos.

Existem variações evidentes na forma de onda do sinal da voz que são uma consequência direta dos movimentos do sistema articulatorio da voz, o qual raramente permanece inalterado por um considerável período de tempo.[8][11]

Para a análise do sinal de voz devemos considerar estas propriedades acústicas.

### 2.3.2 Classificação das Formas de Excitação dos Sons da Voz

O sinal de voz pode apresentar segmentos que se repetem quase periodicamente e segmentos basicamente aleatórios, sem nenhuma periodicidade.

Os sons da voz podem ser classificados em 3 classes distintas de acordo com o modo de excitação. As classes são as seguintes: sons vocálicos, sons não vocálicos e sons explosivos.[8][11]

### 2.3.3 Sons Vocálicos

O fluxo de ar vindo dos pulmões é controlado pela abertura e fechamento das pregas vocais, que são ligamentos semelhantes a dois lábios que podem ser tensionados e (ou) aproximados sob o controle do falante. A abertura entre as dobras é denominada glote.

Estando a glote completamente fechada, o fluxo de ar vindo dos pulmões é interrompido e a pressão subglótica aumenta até que as pregas vocais sejam separadas, liberando o ar pressionado, gerando um pulso de ar de curta duração. Com o escoamento do ar, a pressão glótica é reduzida, possibilitando uma nova aproximação das pregas vocais.

O processo se repete de forma quase periódica. Dessa forma, são obtidas ondas de pressão, quase periódicas, excitando o trato vocal, que atuando como um ressonador, modifica o sinal de excitação, produzindo frequências de ressonância, denominadas de formantes, que caracterizarão os diferentes sons vocálicos. [7]

Quanto mais rápida a repetição, mais alta a frequência e mais aguda será a voz, como nas vozes femininas e infantis; quanto mais lentamente essas repetições se reproduzem, mais grave será a voz, como no caso das vozes masculinas.[9]

As vogais, cujo grau de nasalização é determinado pelo abaixamento da úvula, são exemplos típicos de sons vocálicos. A Figura 13 mostra a forma de onda do sinal de voz para a vogal /a/.

Algumas consoantes, como /l/ e /m/, também são produzidas com a excitação glotal.

A frequência média dos pulsos é denominada frequência fundamental de excitação,

$F_0$ , e o período fundamental (ou período de *pitch*),  $P_0$ , é dado por:

$$P_0 = \frac{1}{F_0} \quad (2.3.3.1)$$

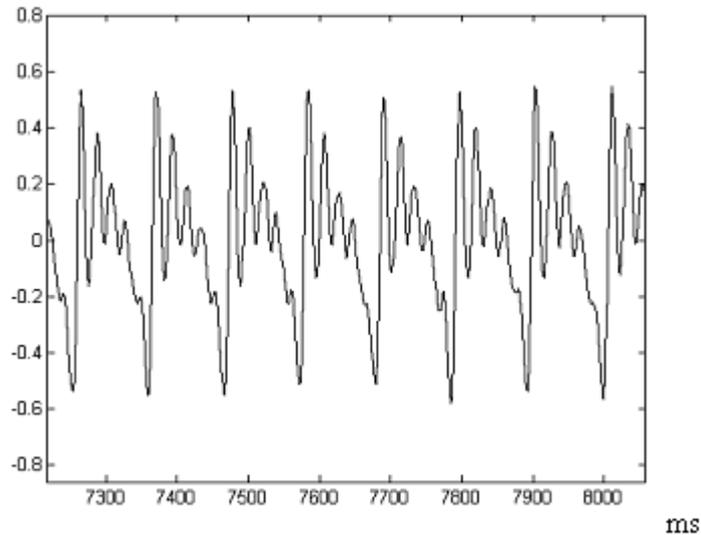


Figura 13- Segmento de vogal /a/

Em processamento de voz os termos *pitch* e frequência fundamental são utilizados como sinônimos, embora o conceito de *pitch* seja mais abrangente. A rigor o *pitch* de um determinado estímulo sonoro (não necessariamente um sinal de voz), corresponde à frequência, em Hz, de um tom senoidal que está “afinado” com o estímulo, segundo a percepção auditiva de um determinado ouvinte. Como, na percepção de voz, o *pitch* dos sons vocálicos geralmente corresponde ao valor da frequência fundamental, para as pessoas com audição normal, os dois termos passaram a ser empregados indistintamente.[8]

A frequência fundamental dos sons vocálicos está entre 80-120 Hz (para homens) e 350 Hz (para crianças), sendo 240 Hz um valor típico para mulheres.[9]

### 2.3.4 Sons Não Vocálicos

São gerados por uma constrição em alguma parte do trato vocal (usualmente próximo ao final da boca), assim o ar adquire velocidade suficientemente alta para produzir turbulência gerando um ruído de espectro largo (semelhante ao ruído branco) para excitar o trato vocal.

Na produção desses sons a glote permanece aberta, não havendo vibração das pregas vocais. Por exemplo, na produção do fonema /s/ (Figura 14), os lábios e dentes são ligeiramente pressionados, deixando assim uma passagem estreita para o ar, produzindo um fluxo de ar turbulento nas imediações da constrição, o qual excita as cavidades do trato vocal.

O som produzido dessa forma tem características ruidosas com concentração relativa de energia nas mais altas componentes de frequência do espectro de sinais de voz.[5][6][7][8]

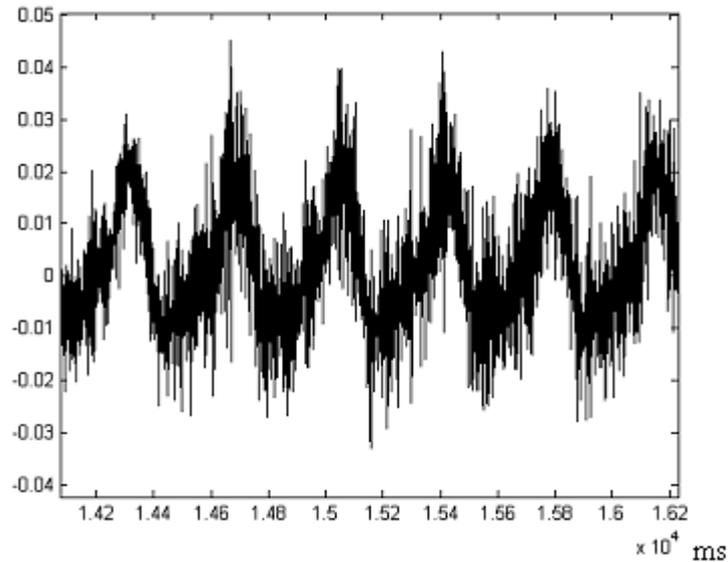


Figura 14 - Forma de onda do fonema /s/.

### 2.3.5 Sons plosivos

Na geração dos sons explosivos, o ar é totalmente dirigido à boca, estando esta completamente fechada. Com o aumento da pressão, a oclusão é interrompida bruscamente, gerando um pulso que excita o aparelho fonador. Com a excitação ocorre um movimento rápido dos articuladores em direção à configuração do próximo som.

Exemplos de sons plosivos são os fonemas /p/, /t/, /k/, dentre outros.[5][6][7] A Figura 15 mostra a forma de onda do explosivo /p/.

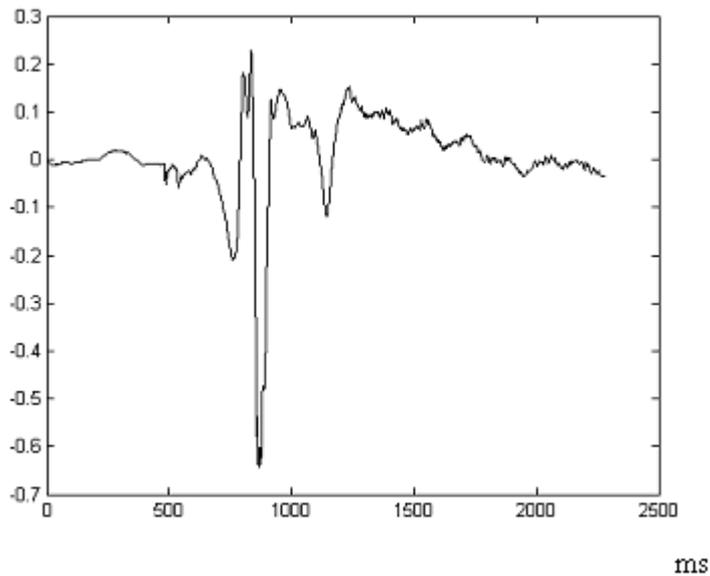


Figura 15 - Forma de onda do fonema /p/ na fonação de “pa”.

### 2.3.6 Sons com excitação mista

Os sons fricativos vocálicos, como /j/, /v/ e /z/, são produzidos combinando-se vibração das pregas vocais e excitação turbulenta. Nos períodos em que a pressão glótica atinge um máximo, o escoamento através da obstrução torna-se turbulento, gerando o caráter fricativo do som; quando a pressão glótica cai abaixo de um dado valor, termina o escoamento turbulento do ar e as ondas de pressão apresentam comportamento mais suave.[5][6][7][8]

A Figura 16 mostra o fonema fricativo sonoro /z/.

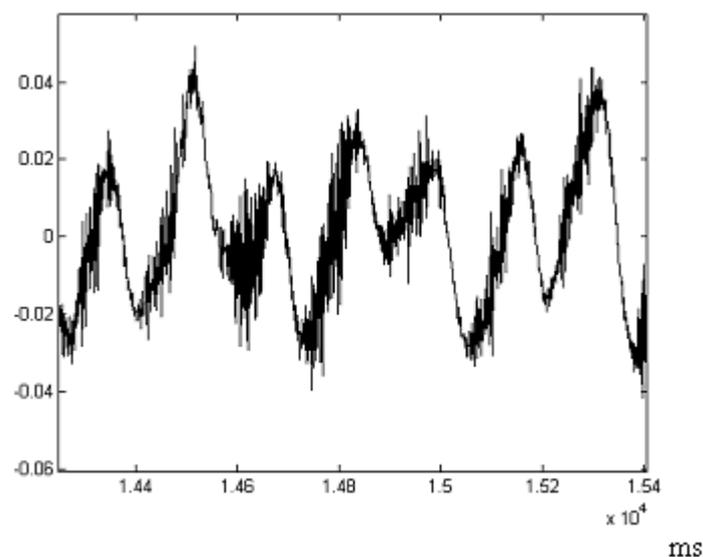


Figura 16 - Forma de onda do fonema /z/.

Os sons oclusivos (ou plosivos) vocálicos, como /d/ e /b/, são produzidos de forma semelhante aos correspondentes não vocálicos, /p/ e /t/, porém há vibração das pregas vocais durante a fase de fechamento da cavidade oral. A Figura 17 mostra a forma de onda do fonema explosivo sonoro /b/.

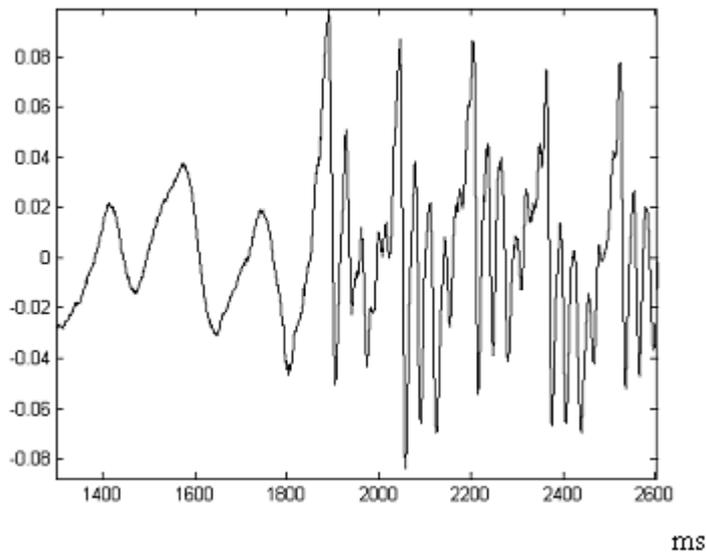


Figura 17- Forma de onda do fonema /b/ na fonação de “ba”.

Portanto, as vogais são as produções sonoras mais periódicas da vocalização humana e por isto são as mais utilizadas em sistemas de identificação de alterações laríngicas das pregas vocais, onde esta quase periódica produção facilita os processos de análise computacionais.

## 2.4 Modelos de produção do sinal da voz

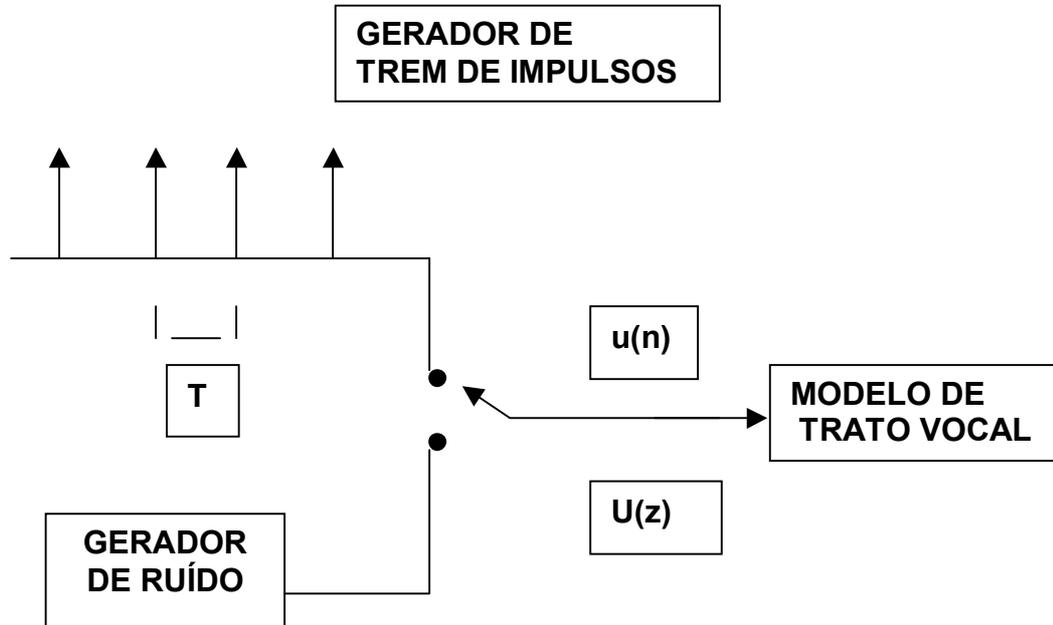
Existem vários modelos de produção da voz sendo, porém, comum a todos, a separação da excitação do trato vocal.

### Modelagem do trato vocal

Pode-se definir um modelo matemático de produção da voz, baseado no processo fisiológico, e que será utilizado posteriormente para a análise do sinal da voz.

A fonte sonora é aproximadamente modelada por um gerador com dois tipos de excitação:

O trem de impulsos, responsável pelos sons vocálicos, e o ruído branco, responsável pelos sons não vocálicos. Estas excitações são comutadas em todo momento, em função dos fonemas produzidos. A Figura 18 apresenta o diagrama de blocos de modelo de excitações:



**T= período dos pulsos,  $u(n)$ =pulsos e  $U(z)$ =excitação**

Figura 18 - Diagrama de blocos de modelo de excitações.

Fonte: RUBEM D. R. FAGUNDES; Reconhecimento de Voz, Linguagem Contínua, usando Modelos de Markov, Dissertação de Mestrado, S.P.(1993, p17).

A articulação será modelada segundo um filtro que reproduza as características acústicas do trato vocal, em algum intervalo de tempo, visto que o trato vocal comporta-se como um tubo acústico que altera todo o momento sua conformação, em virtude dos diferentes sons que são produzidos, pela ressonância do som no seu interior, Figura 19.

[6][7][8]

O modelo assumido para o trato vocal terá as seguintes características [6][7][8]:

- 1- O modelo proposto consiste de um tubo de M secções interconectadas, de igual comprimento, e, de área individual uniforme.
- 2- A secção transversal do tubo é muito pequena, quando comparada com o comprimento da onda do som, de modo que, em cada secção, a propagação pode ser tratada como a de uma onda plana. Segue-se ainda que os tubos são rígidos, de forma que perdas internas por vibração das paredes, viscosidade e calor são negligenciadas.

3- O modelo é linear e desacoplado da glote.

Para definir um modelo mais apropriado para a articulação deve-se também levar em consideração o efeito da glote no modelo de filtro considerado.

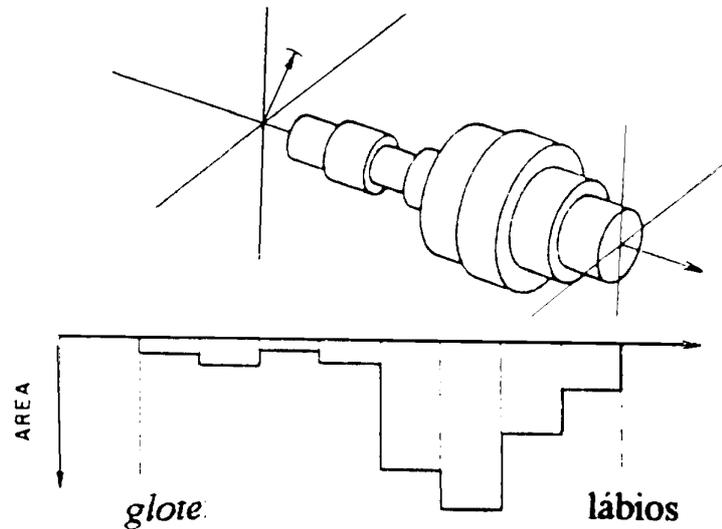


Figura 19- Modelo de tubos acústicos concatenados  
 Fonte: RUBEM D. R. FAGUNDES; Reconhecimento de Voz,  
 Linguagem Contínua, usando Modelos de Markov, Dissertação de  
 Mestrado, S.P.(1993, p18).

Deste modo, chamando de  $H(z)$  o filtro em  $z$  que modela a articulação, de  $T(z)$  o filtro que modela o próprio trato vocal e  $G(z)$ , o filtro relativo a ação da glote, a expressão de  $H(z)$  terá a forma:

$$H(z) = G(z) \cdot T(z) \quad (2.4.1.1)$$

e, a radiação do sinal de voz pelos lábios é caracterizada pelo filtro  $L(z)$ , [7][8][10]:

$$L(z) \cong 1 - z^{-1} \quad (2.4.1.2)$$

Assim o sinal de voz  $S(z)$  é formulado por:

$$S(z) = U(z) \cdot H(z) \cdot L(z) \quad (2.4.1.3)$$

Onde:

$U(z)$  é expressão em  $S(z)$  da excitação,

$H(z)$  é a formulação relativa a articulação e,

$L(z)$  é a parcela referente a radiação pelos lábios. [7][8][10]

O modelo de ação da glote pode ser expresso por:

$$G(z) = \frac{1}{L^2(z)} \quad (2.4.1.4)$$

Assim, lembrando a expressão de  $H(z)$ , chega-se a expressão final do sinal de voz  $S(z)$ :

$$S(z) = \frac{U(z) \cdot T(z)}{L(z)} \quad (2.4.1.5)$$

A expressão  $S(z)$  pode ser apresentada na forma:

$$S(z) = U(z) \cdot T(z) \frac{1}{L(z)} \quad (2.4.1.6)$$

Assim, com o propósito de retirar esta contribuição do efeito da irradiação, e permanecer somente com o trato vocal, que em última análise contém os parâmetros essenciais do sinal de voz, deve-se pré-enfatizar o sinal, isto é, utilizar um filtro de pré-ênfase, da forma  $P(z) = 1 - \alpha \cdot z^{-1}$ , então:

$$\hat{S}(z) = S(z) \cdot P(z) \quad (2.4.1.7)$$

lembrando que pela expressão  $L(z) \cong 1 - z^{-1}$ , então, para  $\alpha \approx 1$  (os valores típicos para  $\alpha$  são 0,95 ou 0,97),  $L(z) \cong P(z)$ , e:

$$\hat{S}(z) = U(z) \cdot T(z) \frac{1}{L(z)} P(z) \quad (2.4.1.8)$$

$$\hat{S}(z) = U(z) \cdot T(z) \quad (2.4.1.9)$$

deste modo,  $\hat{S}(z)$  possui apenas os termos referentes a excitação  $U(z)$  e o trato vocal  $T(z)$ . [7][8][10]

$$\boxed{S(z) = U(z) \frac{1}{A(z)}} \quad \text{Modelo de Síntese} \quad (2.4.1.10)$$

Lembrando as características propostas para o trato vocal, o termo  $A(z)$  fica:

$$A(z) = \sum_{i=0}^M \alpha_i z^{-i} \quad (2.4.1.11)$$

e  $\alpha_0 = 1$

onde  $A(z)$  é um “*all-zero filter*”, isto é, um filtro somente com zeros, e que será chamado de filtro inverso. O filtro  $1/A(z)$  é um “*all-pole filter*”, ou seja um filtro somente com pólos, os quais representam o comportamento espectral do sinal de voz.

“M” representa o número de secções do tubo acústico utilizado como modelo do trato vocal. O filtro  $1/A(z)$  modela o comportamento de ressonância acústica do trato vocal.

A expressão  $S(z)$ , é denominada de modelo de síntese, desde que  $U(z)$  é aplicada ao filtro  $1/A(z)$  e a saída é o sinal de voz, correspondendo, deste modo, a um modelo de síntese de sinal de voz.

Outro tipo de modelo a ser considerado é o da forma:

$$\boxed{U(z) = S(z)A(z)} \quad \text{Modelo de Análise} \quad (2.4.1.12)$$

No modelo de análise, o sinal  $S(z)$  é agora o sinal de entrada para o filtro  $A(z)$ , cujos coeficientes são obtidos da análise do sinal de voz, e a saída é a própria excitação  $U(z)$ , definida no modelo de produção de voz proposto.

Deste modo, o modelo de produção de voz final, proposto, tem o diagrama de blocos mostrado na Figura 20.

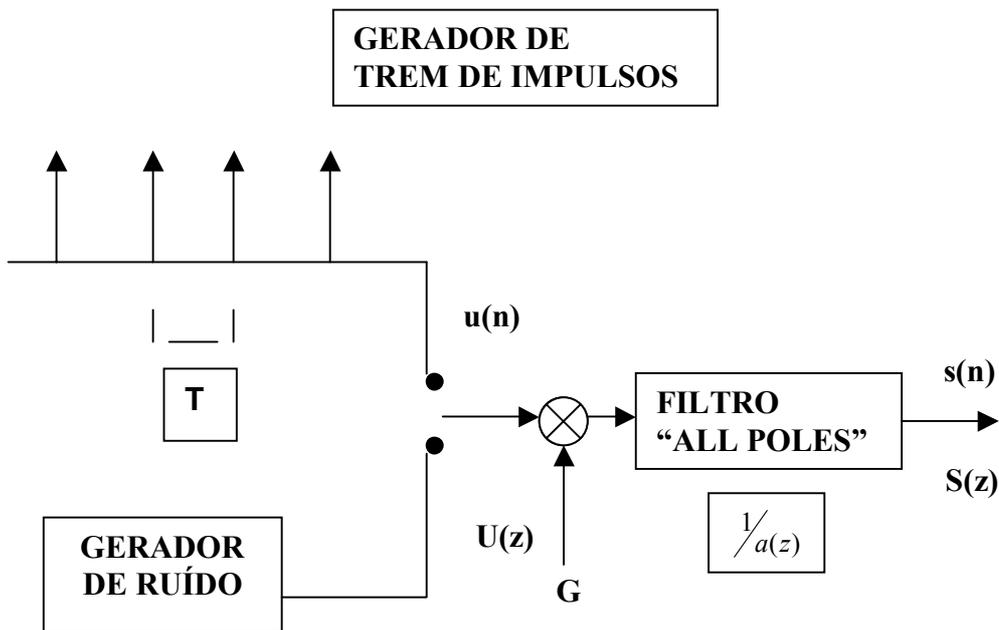


Figura 20 – Diagrama de blocos do modelo de produção da voz.  
 Fonte: RUBEM D. R. FAGUNDES; Reconhecimento de Voz,  
 Linguagem Contínua, usando Modelos de Markov, Dissertação de  
 Mestrado, S.P.(1993, p21).

O termo  $G$  é relativo a intensidade do sinal de voz produzido. Para fins de análise do sinal de voz é executada uma normalização ganho, resultando  $G=1$ .

A expressão de  $S(z) = U(z) \frac{1}{A(z)}$  pode ser expressa por:

$$s(n) = G.u(n) - \sum_{i=0}^M \alpha_i s(n-i) \quad (2.4.1.13)$$

e representa o modelo matemático do sinal de voz no tempo discreto.

Cabe salientar que esta abordagem é coerente para a análise do sinal de voz de tempo curto, isto é, considerando os intervalos de tempo de análise do sinal de voz suficientemente pequenos para que os parâmetros dos modelos sejam invariantes. [5][7][8][10]

## 2.5 Processamento dos sinais de voz

### 2.5.1 Pré-Ênfase

Experimentos mostram que as características do aparelho fonador mudam lentamente na produção do sinal de voz. As mudanças ocorrem para períodos em torno de 10 a 30 ms. Pode-se, então, modelar o aparelho fonador humano por um sistema linear, lentamente variante com o tempo, que pode ser excitado por um trem de pulsos quase-periódicos (sons sonoros) ou por ruído branco (sons surdos). Filtrando o sinal de voz proveniente do microfone com um filtro  $L(z)$  do tipo:

$$L(z) = 1 - \alpha_p \cdot z^{-1} \quad (2.4.1)$$

obtém-se o modelo do sistema glotal, com redução dos efeitos da radiação dos lábios e da variação da área da glote. A esse processo de tratamento do sinal de voz, dá-se o nome de pré-ênfase. O parâmetro  $\alpha_p$  é denominado fator de pré-ênfase. Valores típicos de  $\alpha_p$  são próximos de 1,0. Assim, a pré-ênfase é realizada por meio da fórmula usual, usando por exemplo um  $\alpha_p = 0,95$ :

$$Sp(n) = s(n) - 0,95 \cdot s(n-1) \quad (2.4.2)$$

Além disso, as componentes de alta frequência do sinal de voz são caracterizadas por apresentarem baixas amplitudes e por isso são facilmente afetadas pelo ruído. Apesar do sinal de voz ter a energia mais concentrada nas baixas frequências, as frequências mais altas são responsáveis pela geração dos sons surdos (fricativos). Assim sendo, após a aquisição do sinal, também pode-se usar a pré-ênfase com o propósito de tornar mais plano o espectro desse sinal. [5][6][7]

### 2.5.2 Janelamento

O janelamento é uma segmentação do sinal para a análise em curtos intervalos. Esta segmentação consiste em particionar o sinal de voz em segmentos, selecionados em janelas ou quadros de duração perfeitamente definida, como mostra a Figura 21.

Os segmentos são escolhidos dentro dos limites de estacionariedade do sinal. A segmentação é levada a uma superposição de 50%, visando reduzir os efeitos da descontinuidade entre segmentos.

Se o sinal é, simplesmente, particionado em blocos consecutivos, então está sendo aplicada, implicitamente, uma janela retangular de comprimento igual ao comprimento do bloco. Entretanto, como resultado no domínio da frequência, surgem fugas espectrais alterando o espectro do sinal. Para evitar este efeito indesejável no domínio de frequência, utilizam-se janelas de Hamming ou Hanning, definidas no domínio do tempo, que proporcionam, no domínio da frequência, um lóbulo principal de amplitude bastante superior a dos lóbulos secundários, diminuindo, portanto, o efeito destes. O efeito do janelamento de Hamming é a manutenção das características espectrais do centro do quadro e a eliminação das transições abruptas das extremidades Figura 22.

A janela de Hanning assemelha-se a de Hamming porém, proporciona um reforço menor nas amostras do centro e uma suavização maior nas amostras da extremidade.

A janela de Hamming, porém, apresenta uma característica nem sempre desejável, que corresponde à atribuição de um peso muito baixo às amostras da extremidade.

Entretanto, estas amostras podem representar eventos importantes de curta duração do sinal de voz e multiplicá-los por um peso baixo representa pouca atenção no processamento subsequente realizado em segmentos. Para assegurar que a tais eventos seja dado o peso necessário, segmentos adjacentes são sobrepostos de modo que um evento seja “coberto” por outros segmentos. Para o contexto da produção da voz, a característica do janelamento de Hamming se mostra, portanto, mais eficiente, quando comparada aos outros tipos de janela (Retangular e Hanning), com uma boa aproximação da janela ideal.[6][7][8] A Figura 21 mostra o resultado de algumas janelas.

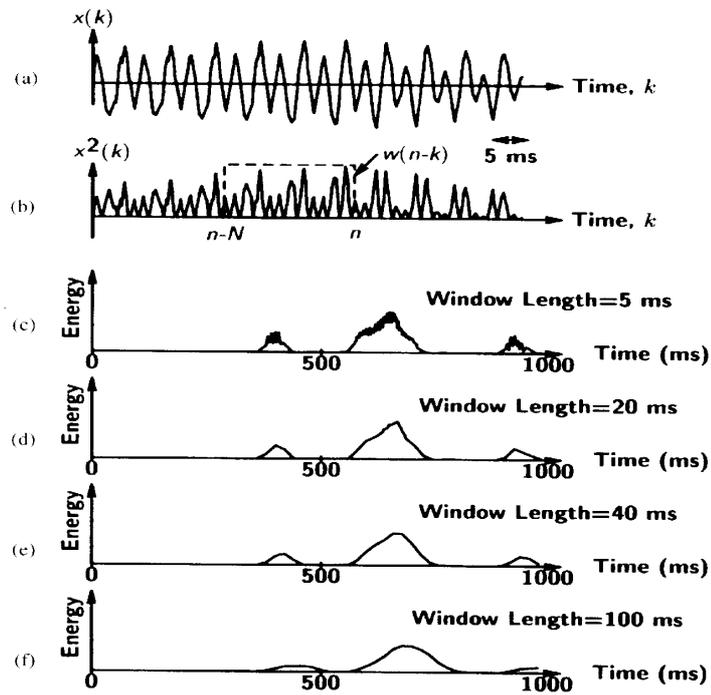


Figura 21 - Exemplo de sinal de voz segmentado com diferentes comprimentos de janela.

Fonte: DOUGLAS O' SHAUGHNESSY; Speech Communications: Human and Machine; *Institit of Eletrical and Eletronics Engineers.Inc.* N.Y.; second edition (2000, p181).

Os tipos de janela freqüentemente utilizados são: Janela Retangular, Janela de Hamming e Janela de Hanning, cujas características são mostradas a seguir:

### Janela Retangular

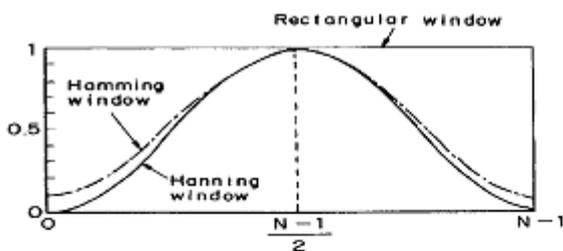
$$J(n) = \begin{cases} 1 & , 0 \leq n \leq N_A - 1 \\ 0 & , \text{caso contrário} \end{cases}$$

### Janela de Hamming

$$J(n) = \begin{cases} 0,54 - 0,46 \cdot \cos[2\pi n / (N_A - 1)] & , 0 \leq n \leq N_A - 1 \\ 0, & \text{caso contrário} \end{cases}$$

### Janela de Hanning

$$J(n) = \begin{cases} 2\alpha \cos[\pi n / N_A] + b & , 0 \leq n \leq N_A - 1 \\ 0, & \text{caso contrário} \end{cases}$$



(a)

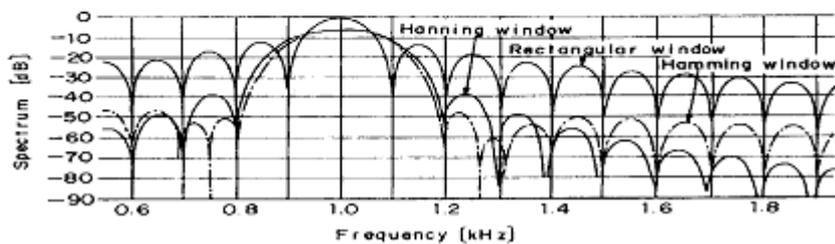


Figura 22 - Função da janela (a) e o espectro de algumas janelas.

Fonte: FURUI, S, Digital Speech Processing, Synthesis, and Recognition, Marcel Dekker, Inc.(2001, p59)

## 2.6 Análise Homomórfica

O processamento homomórfico é uma técnica de análise da voz, que faz parte das técnicas não-lineares de processamento de sinal, sendo baseado na generalização do princípio de superposição que define os sistemas lineares.[7]

Sistemas que atendam à generalização do princípio de superposição são denominados sistemas homomórficos, uma vez podem ser representados por meio de um mapeamento linear (homomórfico, ou que preserva a forma) da entrada para a saída.

A partir do modelo utilizado para a geração de um sinal, o problema de análise do sinal de voz, em segmentos curtos no tempo, pode ser encarado como uma questão de separação das componentes de uma convolução da função excitação (pulsos quase periódicos ou ruído branco) com a função de resposta impulsiva do trato vocal, ou seja, um problema de deconvolução proveniente do modelo de produção da voz (Figura 20).

A partir destas considerações, a deconvolução homomórfica pode ser aplicada para separar estes sinais. O processo consiste em tornar o resultado da transformada de Fourier do sinal de voz e calcular a transformada inversa de Fourier do logaritmo deste sinal. O logaritmo é aplicado com o objetivo de transformar o produto entre a função de excitação e o trato, em uma soma:

$$\log(S) = \log(EV) = \log(E) + \log(v) \quad (2.6.1)$$

onde: S= transformada de Fourier do sinal de voz

E= parcela correspondente a função de excitação

V= parcela correspondente a função de transferência do trato vocal

$$c(n) = \mathfrak{Z}^{-1} \log(S) = \mathfrak{Z}^{-1} \log(E) + \mathfrak{Z}^{-1} \log(v) \quad (2.6.2)$$

O resultado desta transformação, em função das partes real e imaginária da transformada de Fourier, é o chamado Cepstrum complexo, abordado no próximo item.

### 2.6.1 Análise Cepstral

A análise cepstral é uma técnica que nos permite separar as características do filtro do trato vocal da seqüência de excitação no modelo até agora usado em que se assume que o sinal de voz, onde  $s(t)$  é composto por um sinal de excitação  $e(t)$  aplicado ao filtro do trato vocal, com uma resposta impulsional  $v(t)$ , de um ponto de vista no domínio temporal como apresentado na Figura 23.

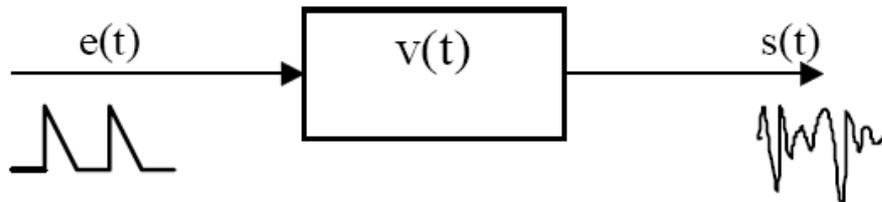


Figura 23 - Modelo simplificado da produção da voz no domínio do tempo.

$s(t)$  é a convolução de  $e(t)$  com  $v(t)$

$$s(t) = e(t) \otimes v(t) \quad (2.6.1.1)$$

que no domínio da freqüência fica :

$$S(w) = E(w)V(w) \quad (2.6.1.2)$$

em que  $S(w)$ ,  $E(w)$  e  $V(w)$  são as transformadas de Fourier das funções contínuas no tempo  $s(t)$ ,  $e(t)$  e  $v(t)$  ou as transformadas discretas de Fourier das seqüências de amostras temporais  $s(n)$ ,  $e(n)$  e  $v(n)$ .

O sinal de excitação  $E(w)$  e o filtro do trato vocal aparecem combinados multiplicativamente, o que torna difícil a sua separação. Contudo, tomando o logaritmo de  $S(w)$  a excitação e a função do trato vocal tornam-se aditivas.

$$\log[S(w)] = \log[E(w)] + \log[V(w)] \quad (2.6.1.3)$$

A propriedade aditiva do espectro logarítmico continua-se a verificar quando lhe for aplicada a Transformada de Fourier, sendo o resultado dessa operação chamado de função Cepstral ou Cepstro.[7] Sumariamente, o processo de estimação do Cepstro apresenta-se no diagrama de fluxo da Figura 24.

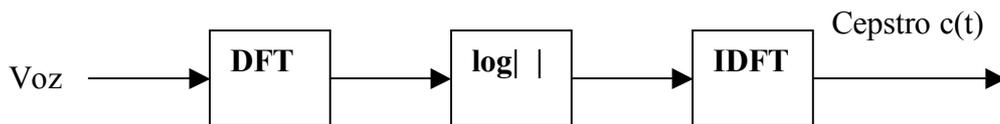


Figura 24 - Diagrama de fluxo da estimação do Cepstro.

A forte componente periódica no espectro logarítmico de um sinal de voz num intervalo de frequências equivalente ao inverso do período fundamental  $T$ , aparece no Cepstro como um pico.

A variável independente, no eixo horizontal, da função cepstral tem dimensões temporais e o nome de quefrências. Na voz pode ser feita uma clara distinção entre a componente de excitação e a contribuição do trato vocal que aparece como um aglomerado de componentes aos baixos valores de quefrência afastado da componente do período fundamental que aparece com valores mais altos de quefrência.

Na Figura 25 apresenta-se o Cepstro de um segmento de voz. Vê-se que o pico correspondente ao período fundamental próximo da quefrência de 10 ms, correspondendo à frequência fundamental, separado das componentes do trato vocal às de baixas quefrências. Nesta Figura são apresentadas apenas as componentes do Cepstro superiores a sensivelmente 1 ms, pois as componentes de mais baixas quefrências têm valores comparativamente muito superiores aos restantes e a sua apresentação não deixaria claro o pico correspondente à frequência fundamental.[5][6][7][8]

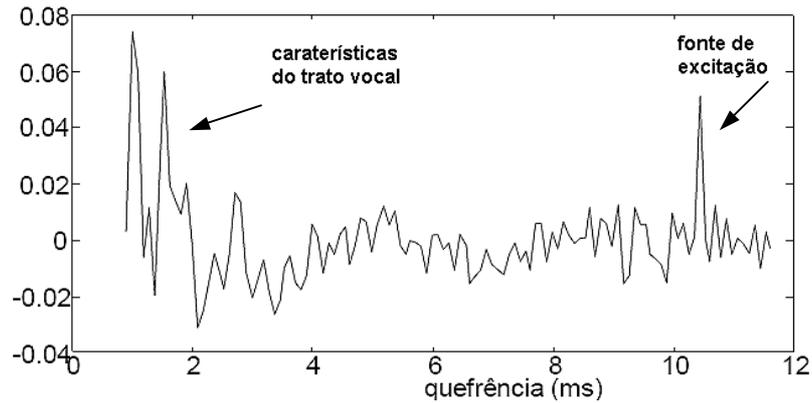


Figura 25 - Cepstro de um segmento de fala da vogal /a/.

Também é salientado que existem o Cepstrum complexo e Cepstrum real. A diferença básica entre estas duas formas consiste em que o Cepstrum real, ou simplesmente cepstrum, descarta a informação de fase. Embora o Cepstrum complexo seja mais preciso, apresentando informações que são perdidas no cepstrum real, sua aplicação em termos computacionais é mais onerosa. Suas aplicações vão depender de sua finalidade, por exemplo em aplicações onde sejam necessário informações sobre a fase, como alguns codificadores de voz.

A obtenção do cepstrum real é feita através da mesma seqüência do Cepstrum complexo, porém o logaritmo é aplicado apenas à magnitude da transformada de Fourier, correspondente ao produto entre a excitação e a resposta ao impulso do trato vocal. Obtém-se assim um sinal composto da soma de duas componentes, o cepstrum da excitação e da resposta do trato vocal.

Uma vez que o cepstrum retém toda a informação contida no espectro de tempo curto, o mesmo pode ser encarado como uma ferramenta alternativa de representação para o sinal da voz. O cepstrum real é o que foi utilizado neste trabalho, fornecendo os gráficos para a identificação das regiões de excitação e ressonantal separadamente.

### 2.6.2 *Liftering*

A função de transferência do trato vocal e a função de excitação da voz aparecem em partes separadas da escala de quefrências, pelo que podem ser separadas as duas funções, ou removida uma delas por um processo de “lifteragem” (Figura 26). O Cepstro é constituído por um conjunto de valores cepstrais discretos, que são o conjunto de valores de saída do processo final de IDFT (Inverse Discrete Fourier Transform). Então podemos aplicar um *lifter* com uma função de janela retangular. Contudo, a transformada de Fourier da janela retangular

indica que tal processo gera lobos laterais indesejados às componentes requeridas no domínio das frequências logarítmicas, pelo que é requerida uma função de uma janela mais gradual.

Assim aplicando esta função de liftering ao sinal do Cepstro e tomando a transformada de Fourier Discreta Inversa (IDFT) do sinal resultante, obtém-se uma versão alisada do espectro logarítmico do filtro do trato vocal.[7]

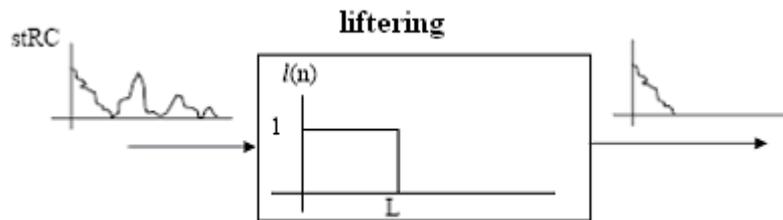


Figura 26 - Exemplo de uma função de *liftering*.

Fonte: M.W. Mak, Sit C.Hung, SPANA – Development of Multimedia Tool for Learning Speech Analysis.(2004, p16)

## **3 Proposta de trabalho**

### **3.1 Técnica de Processamento de Sinais na identificação de alterações laríngeas**

A principal proposta deste trabalho é utilizar a análise Cepstral como método de caracterizar as alterações laríngeas, obtendo-se informações que sirvam de base de dados para desenvolvimento de método de diagnóstico diferencial das alterações laríngeas. A seguir as etapas do sistema proposto:

#### **1- Pré-ênfase, se devidamente justificável**

A pré-ênfase permite a filtração da radiação dos lábios diminuindo assim a influência desta região sobre os resultados.

#### **2- Janelamento (Windowing)**

Leva a efeito reduzir os efeitos da descontinuidade entre segmentos do sinal, evitando que ocorram fugas espectrais que poderiam alterar o espectro do sinal. Para evitar este efeito indesejável pode-se utilizar Hamming ou Hanning. Serão testados diferentes tipos de janelas assim como comprimentos da janela, que melhor proporcionem a obtenção dos resultados.

#### **3- Análise Cepstral**

Esta é uma técnica de processamento de sinais que permite a separação do sinal da glote (excitação) do sinal da ressonância (cavidades de ressonância). Isto permitirá estudar separadamente o efeito das alterações laríngeas das pregas vocais sobre estas regiões.

#### **4- Análise dos gráficos cepstrais de cada alteração laríngea**

Os resultados serão gráficos cepstrais das alterações laríngeas que trarão informações inovadoras sobre o efeito destas alterações laríngeas.

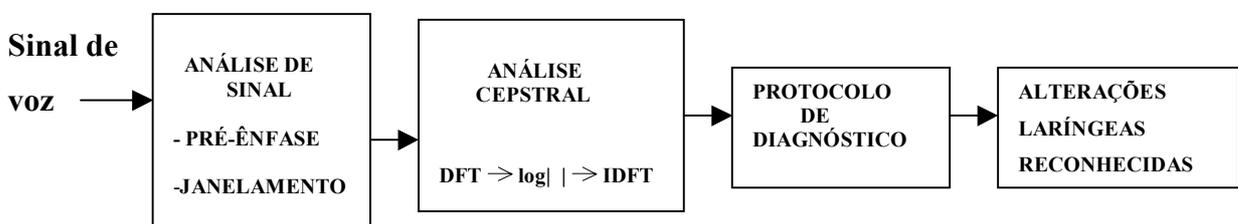
### 5- Identificação das alterações cepstrais de cada alteração laríngea

Através da análise dos resultados será possível encontrar características em comum de cada alteração laríngea servindo de meio de identificação destas.

### 6- Uso das alterações para o reconhecimento da alteração laríngea

Desenvolver ferramentas de diagnóstico com os achados destas alterações laríngeas.

## 3.2 Diagrama de blocos proposto de um sistema de reconhecimento de padrões de voz



### 3.3 Análise das alterações laríngeas

A aplicação da técnica de processamento de sinais centrada na análise cepstral no sinal acústico irá gerar dados que possam ser classificados em grupos de características de cada alteração laríngea estudada. As alterações laríngeas serão separadas em grupos pelo seu diagnóstico com o exame de videolaringoscopia e também será determinado um grupo de normais femininos e masculinos para comparação e controle. As alterações laríngeas que participarão são as mais frequentemente diagnosticadas em consultório, sendo elas, portanto mais fáceis de obter amostras. Para a análise cepstral será usado o sinal da gravação da voz durante o exame endoscópico.

### **3.4 Características e diferenciação das alterações laríngicas**

Baseado, na capacidade do processamento de sinais da análise cepstral em separar o sinal da fonte e da ressonância, poderemos obter dados que identifiquem características de cada patologia, pois as alterações acústicas são diferentes entre as alterações laríngicas que deverá gerar achados diferenciados nas ondas cepstrais em cada alterações laríngica.

Algumas alterações laríngicas provavelmente provocarão alterações que ainda não foram descritas na literatura, o que trará novos modelos de padrões na análise acústica cepstral das alterações laríngicas da laringe.[8][15]

### **3.5 Técnica proposta de identificação das alterações laríngicas**

Os achados serão usados para classificar as alterações das alterações laríngicas de uma forma clara, como a criação de gráficos e tabelas dos dados, permitindo que os resultados sejam avaliados e comparados.

A visualização da alteração acústica é um método que permitirá demonstrar os fenômenos sonoros ocorridos quando a alteração laríngica modifica o mecanismo básico de produção da voz, facilitando assim seu estudo.

Ao obter-se os dados que caracterizem estas diferenças será possível criar uma rotina que seja capaz de identificar as alterações laríngicas com a aplicação do processamento do sinal, e se possível e justificável os dados serão usados no treinamento de processos de identificação automático de padrões.

## 4 Resultados

### 4.1 A interpretação dos gráficos obtidos

Inicialmente salienta-se que os gráficos cepstrais não são mencionados na bibliografia como fonte de interpretações detalhadas do sinal da voz, e principalmente nada descrito como método gráfico de diferenciação de alterações laríngeas da laringe. Por isto se desenvolveu uma rotina de análises dos achados que foram considerados importantes para caracterizar cada alteração laríngea.

Os gráficos cepstrais obtidos da análise da vocalização de cada alteração laríngea diagnosticada foram analisados com a ajuda das características fisiopatológicas já conhecidas e foram descritas as alterações percebidas que podem ser particularmente atribuídas as diferentes alterações laríngeas.

Foram considerados como padrões os gráficos cepstrais primeiramente obtidos dos pacientes normais, e que foram separados em grupo feminino e masculino. Eles apresentam características específicas. Também se percebe uma significativa diferença no traçado entre os casos femininos e os masculinos, principalmente no que se refere ao valor de quefrência do *pitch*.

Os gráficos obtidos fornecem muitas informações interessantes e inéditas que não são mencionados na revisão da literatura realizada, sendo necessário se criar critérios para a sua observação e análise. Sendo assim, os dados observados que foram considerados significativos para a caracterização das alterações laríngeas foram organizados em itens.

Assim foi possível a criação de tabelas que permitiram um agrupamento das principais características cepstográficas de cada alteração laríngea, o que facilitou, ao se observar um gráfico, identificar as características da alteração laríngea até classificá-la.

### 4.2 Aplicação da interpretação desenvolvida nos gráficos obtidos

Muitas características se repetem nas alterações laríngeas, proporcionando dados que servem como referências de uma alteração laríngea específica e tendem a ser iguais nas alterações do mesmo grupo, mantendo um padrão em cada grupo.

Foram observados detalhes nestas diferenças e que em certas alterações laríngeas são muito evidentes, porém em outras nem tanto.

## 5 Conclusões e Trabalhos Futuros

O método de diagnóstico das alterações laríngeas através da vocalização permitiu observar os diferentes achados dos casos em estudo, facilitando desenvolver um conjunto de parâmetros compactados e agrupados em tipos de alterações laríngeas e suas características.

Para a análise destes parâmetros foram criadas tabelas com valores encontrados em cada alteração laríngea, que além de simplificar a análise, permitiu se atingir o objetivo de diferenciar os casos em estudo. Para este objetivo uma tabela foi criada com os valores, com os quais pode-se diagnosticar de forma coerente as alterações laríngeas, tomando como base os achados cepstrais da vocalização.

Para testar e validar este procedimento de lingüística, casos escolhidos aleatoriamente, e que não participaram do estudo, foram aplicados e diagnosticados, somente com as informações da tabela. O resultado obtido, foi de 89% de acerto, quando os resultados dos gráficos são submetidos aos valores da tabela, validando significativamente o método proposto e contribuindo de forma efetiva para a elaboração de um protocolo de diagnóstico não-invasivo de alterações da laringe, baseado na técnica de análise cepstral.

Os gráficos do sulco vocal e do edema de Reinke são os que mais se diferenciam de todas outras alterações laríngeas. O edema de Reinke deixa a sensação de voz muito grave e rouca e por isto produz uma grande nos achados. Já alguns dos gráficos de cisto de prega e do sulco vocal podem ser muito semelhantes aos normais.

Os gráficos muito semelhantes quando analisados meticulosamente, na maioria das vezes, demonstram diferenças que foram descobertas. Também se destaca que algumas alterações laríngeas apresentam achados característicos e constantes que servem como método diagnóstico destas. Algumas alterações nem sempre são constantes, ou às vezes estão presentes em mais de uma alteração laríngea, e isto deve ser mais aprofundado em trabalhos futuros. Acredita-se que este método inédito, por ser não invasivo, de custo mais baixo e fácil execução e com resultados promissores, como demonstrados nestes dados prévios será uma ferramenta de diagnóstico muito útil. Por fim, a análise Cepstral forneceu informações sobre as alterações laríngeas da laringe que não são demonstrados com outros métodos, principalmente no que se refere à visualização de regiões alteradas específicas e a possibilidade de relacioná-las com as alterações anatomofisiológicas já conhecidas das alterações laríngeas.

A Tabela 1 demonstra uma tabela com o agrupamento dos achados e as características principais obtidas no estudo.

Tabela 1 - **Resumo das principais características de cada método.**

	Diagnóstico com videolaringoscopia (convencional)	Método cepstral de análise da vocalização (sistema proposto)
Invasivo	Sim, pouco invasivo	Não
Execução	somente profissional medico	pode ser executado facilmente
Custos	iguais de um procedimento consultório em	mais baixos que o convencional

Com os resultados da proposta de análise cepstral do sinal da voz nas alterações laríngeas, iniciados neste trabalho, permite-se prever que este método será uma ferramenta diagnóstica muito útil e promissora para este propósito, pois é um método não invasivo, de custo baixo e fácil execução. Para uma maior qualificação deste método deve-se realizar trabalhos futuros com a inclusão de mais amostras, com a inclusão das outras alterações laríngeas mais raras e realização de estudo científico estatisticamente adequado para a validação, como um estudo prospectivo duplo cego. Também é necessário obter modelos de alterações laríngeas com o diagnóstico mais preciso possível para o treinamento do sistema, através de diagnóstico de certeza das alterações da laringe.

O diagnóstico de certeza é obtido somente com uma biópsia da alteração laríngea, mas pode-se obter um diagnóstico confiável com o uso de equipamentos mais sofisticados de laringoscopia como a vídeoestroboscopia, por exemplo, ou com acompanhamento da alteração laríngea e sua evolução com o tratamento, obtendo o diagnóstico evolutivo, sem ser invasivo. Outro aspecto relevante é determinar as possibilidades de que o sexo da pessoa

testada possa influenciar o resultado, já que as alterações laríngeas tendem a alterar o pitch da voz, como observado na pesquisa.

E, ao ser obtido uma modelagem ideal das alterações laríngeas com este método inovador, ele poderá ser usado para treinamento mais preciso de um sistema de reconhecimento automático como, por exemplo, as redes neurais artificiais ou modelos de Markov [18], tornando-se um equipamento de diagnóstico.

## 6. Referências Bibliográficas

Conforme ABNT (NBR 6023)

- [1] Disponível em: <[http:// www.mathtools.net](http://www.mathtools.net)>. Acesso em: agosto de 2005.
- [2] Disponível em: <[http:// www.utdallas.edu/~loizou/speech/colea.htm](http://www.utdallas.edu/~loizou/speech/colea.htm)>. Acesso em: agosto de 2005
- [3] OCTAVIAN Y. Y. CHENG; Speech Processing Workstation, Department of Electrical & Electronic Engineering of the University of Auckland, Report, 2003
- [4] Érica E. Fukuyama; Análise acústica da voz captada na faringe próximo à fonte glótica através de microfone acoplado ao fibrolaringoscópio, Revista Brasileira Otorrinolaringologia, vol.67 no.6 São Paulo Nov. 2001
- [5] RABINER, R.L. and SCHAFER, R.W. Digital Processing of Speech Signal. Prentice Hall, Upper saddle River, New Jersey, 1978
- [6] RABINER, R.L.; GOLD,B. Theory and Application of Digital Processing. Prentice-Hall, 1975
- [7] FURUI, S, Digital Speech Processing, Synthesis, and Recognition.Marcel Dekker, Inc., 2001
- [8] DOUGLAS O' SHAUGHNESSY; Speech Communications: Human and Machine; Institute of Electrical and Electronics Engineers, Inc.NewYork, N.Y.; second edition 2000
- [9] RUSSO, I. e BEHLAU, M. Percepção da Fala: Análise Acústica. Editora Lovise, 1993
- [10] RUBEM D. R. FAGUNDES; Reconhecimento de Voz, Linguagem Contínua, usando Modelos de Markov, Dissertação de Mestrado, São Paulo, 1993
- [11] DELLER, J. R. Jr.; HANSEN,J.H.L. and PROAKIS, J.G.; Discrete-Time Processing of Speech Signals, *Institut of Eletrical and Eletronics Engineers*, Press, 2000
- [12] ANDREW C. LINDGREN, B.S.; Speech Recognition Using Features Extracted from Phase Space Reconstructions, Thesis, Computer Engineering of the Marquette University,Milwaukee,Wisconsin, May 2003.
- [13] VINAY K. INGLE and JONH PROAKIS; Digital Signal Processing using MATLAB, Brooks/cole, Canadá, 2000
- [14] MARTÍNEZ CESAR E., RUFINER HUGO L.; Acoustic analysis of speech for detection of laryngeal pathologies, Proceedings of the 22<sup>nd</sup> Annual EMBS International Conference, July 23-28,Chicago IL., 2000
- [15] MARCELO O. ROSA, JOSÉ C. PEREIRA, MARCOS GRELLER and ANDRÉ CARVALHO; Signal processing and statistical procedures to identify laryngeal pathologies, *Institut of Eletrical and Eletronics Engineers*, 423-426, EESC-USP, São Paulo, Brasil,1999
- [16] P. GOMEZ, J.I. GODINO,F. RODRIGUEZ, F. DIAZ, V. NIETO, A. ÁLVAREZ, V. RODELLAR; Evidence of Vocal Cord Pathology from the mucosal wave Cepstral contents, *Institut of Eletrical and Eletronics Engineers*, Universidade Politécnica de Madrid, Spain, 2004
- [17] JAY G. WILPON, LAWRENCE R. RABINER, CHIN-HUI LEE and E. R. GOLDMAN; Automatic recognition of keywords in unconstrained speech using Markov Models, *Institut of Eletrical and Eletronics Engineers*, Transactions on Acoustic, Speech and Signal Processing, Vol.38, N°11, 1870-1878, november1990
- [18] STEVAN HADJITODOROV, PETER MITEV; A Computer system for acoustic of pathological voices and laryngeal diseases screening, Technical note, Medical Engineering , & Physics 24,419-429, Sofia, Bulgaria, 2002
- [19] C. MANFREDI, M. D'ANIELLO, P. BRUSCAGLIONI, A. ISMAELLI; A

- comparative analysis of fundamental frequency estimation methods with application to pathological voices, *Medical Engineering & Physics* 22, 135-147, Firenze, Italy, 2000
- [20] WIESLAW WSZOLEK, RYSZARD TADEUSIEWICZ, ANDRZEJ IZWORSKI, TADEUSZ WSZOLEK; Automated understanding of selected voice tract pathologies based on the speech signal analysis, *Proceedings of the 23rd Annual EMBS International Conference*, October 25-28, Istanbul, Turkey, 2001
- [21] PETER MITEV, STEVAN HADJITODOROV; Fundamental frequency estimation of voice of patients with laryngeal disorders . *Information Sciences* 156, 3-19, Sofia, Bulgária, 2003
- [22] MINORU HIRANO, DIANE M. BLESS; Exame videoestroboscópico da laringe; Porto Alegre: Artes Médicas, 1997
- [23] ERICH C. M. DE MELO, LUPERCIO L. BRITO, OSIRIS C. BRASIL, MARA BEHLAU, DANIELLY M. DE MELO; Incidência de lesões laringicas não neoplásicas em pacientes com queixas vocais; *Revista Brasileira de Otorrinolaringologia*, vol.67, n.6, 788-94, nov/dez 2001

**ANEXO A - Parecer do Comitê de Ética em Pesquisa**

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
COMITÊ DE ÉTICA EM PESQUISA - CEP - PUCRS



Ofício 840/06-CEP

Porto Alegre, 31 de julho de 2006.

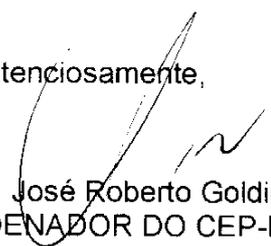
Senhor(a) Pesquisador(a):

O Comitê de Ética em Pesquisa da PUCRS apreciou e aprovou seu protocolo de pesquisa registro CEP 06/03218, intitulado: **“Técnica de processamento de sinais no diagnóstico diferencial de doenças laringeas”**.

Sua investigação está autorizada a partir da presente data.

Relatórios parciais e final da pesquisa devem ser encaminhados a este CEP.

Atenciosamente,

  
Prof. Dr. José Roberto Goldim  
COORDENADOR DO CEP-PUCRS

Ilmo(a) Sr(a)  
Mest. Iuberi Carson Zwetsch  
N/Universidade

**ANEXO B - Consentimento Pós-informado**

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
FACULDADE DE ENGENHARIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA  
MESTRADO EM ENGENHARIA BIOMÉDICA  
Av. Ipiranga 6681\_ CEP 90619-900\_ Fone: 33203540  
Porto Alegre\_ Rio Grande do Sul\_ Brasil

**TERMO DE CONSENTIMENTO PÓS-INFORMADO****NOME DA PESQUISA:**

“Técnica de Processamento de Sinais no Diagnóstico Diferencial de Alterações Laríngeas”

**PESQUISADOR :**

Iuberi Carson Zwetsch, médico otorrinolaringologista, CRM 17574

**ORIENTADOR RESPONSÁVEL:**

Professor Doutor Rubem Dutra Ribeiro Fagundes , Ph.D.

**CO-ORIENTADOR :**

Professora Doutora Thais Russomano , Ph.D.

**INFORMAÇÕES AOS PACIENTES:**

1. As pessoas que apresentarem problemas na voz, podem ter alterações nas pregas vocais, e estas realizarão um exame da pregas vocais e uma posterior análise da gravação da voz no computador para a realização do diagnóstico.
2. A pesquisa a ser realizada, tem a finalidade de investigar a eficácia de um programa de computador para voz, através da gravação da mesma, via microfone/computador e exame otorrinolaringológico denominado videolaringoscopia, que consiste em filmar a laringe/garganta. A filmagem será realizada, utilizando-se uma câmera pequena, que será introduzida dentro da boca, não provocando nenhuma dor.
3. Os procedimentos apresentam riscos mínimos, e com o conhecimento adquirido pelo

examinador nunca teve uma complicação, e através da pesquisa, se verificará a possibilidade de pré-diagnosticar problemas de voz com o uso de programa de voz computadorizado.

EU, \_\_\_\_\_,

RG número \_\_\_\_\_, abaixo assinado, tendo recebido as informações acima, e ciente dos meus direitos abaixo relacionados, concordo em participar.

1. A garantia de receber a resposta a qualquer pergunta ou esclarecimento de qualquer dúvida à cerca dos procedimentos, riscos, benefícios e outros relacionados com a pesquisa;
2. A liberdade de retirar meu consentimento a qualquer momento e deixar de participar no estudo sem que isso traga prejuízo à minha pessoa;
3. A segurança de que não serei identificado e que será mantido o caráter confidencial da informação relacionada com a minha privacidade;
4. O compromisso de me proporcionar informação atualizada durante o estudo, ainda que esta possa afetar minha vontade de continuar participando;
5. O compromisso de que serei devidamente acompanhado e assistido durante todo o período de minha participação na pesquisa;
6. Que se existirem gastos adicionais estes serão absorvidos pelo orçamento da pesquisa,

Tenho ciência do exposto acima e desejo participar desta pesquisa sob orientação do pesquisador que subscreve este documento.

Assinatura participante :

Assinatura do pesquisador :

Porto Alegre, \_\_\_\_\_ de \_\_\_\_\_ de

## ANEXO C – Declaração de aceitação do artigo



Pontifícia Universidade Católica do Rio Grande do Sul  
FACULDADE DE MEDICINA  
REVISTA SCIENTIA MEDICA

**DECLARAÇÃO**

Declaramos que o artigo intitulado "TÉCNICA DE PROCESSAMENTO DE SINAIS NO DIAGNÓSTICO DIFERENCIAL DE DOENÇAS LARÍNGEAS BENIGNAS", de autoria do senhor Iuberi Carson Zwetsch, foi recebido pela Comissão Editorial da Revista Scientia Medica e está em processo de avaliação para publicação.

Porto Alegre, 28 de setembro de 2006.

Prof. Luciano P. Diogo

Editor Executivo

Revista Scientia Medica

**PUCRS**

**Campus Central**  
Av. Ipiranga, 6690 – P. 60 – 3º andar – CEP 90610-000  
Porto Alegre – RS – Brasil  
Fone: (51) 3320-3304 – Fax (51) 3320-3040  
E-mail: revmed@pucrs.br

**ANEXO D – Artigo aceito para publicação****TÉCNICA DE PROCESSAMENTO DE SINAIS NO DIAGNÓSTICO  
DIFERENCIAL DE ALTERAÇÕES LARÍNGEAS BENIGNAS****IUBERI CARSON ZWETSCH<sup>1\*</sup>****RUBEM DUTRA RIBEIRO FAGUNDES<sup>2</sup>****THAIS RUSSOMANO<sup>3</sup>****DIOGO SCOLARI<sup>4</sup>**

1- Médico Otorrinolaringologista

2- Professor PhD da Faculdade de Engenharia Elétrica e do Mestrado em Engenharia Elétrica PUCRS; Coordenador do Laboratório de Sistemas, Sinais e Computação da Faculdade de Engenharia – SiSC/FENG-PUCRS

3- Professora PhD das Faculdades de Medicina e Ciências Aeronáutica e do Mestrado em Engenharia Elétrica/Biomédica da PUCRS; Coordenadora do Laboratório de Microgravidade/IPCT-PUCRS

4- Engenheiro Eletricista, Ênfase em Telecomunicações.

\*Autor Responsável: Prédio 30, PUCRS, Av Ipiranga 6681, Porto Alegre, RS, 90619-900

Fone: 91559131

Fax: 30264652

E-mail: iuberi.rs@brturbo.com.br

## RESUMO

Neste trabalho será apresentado um novo procedimento em Processamento Digital de Sinais (DSP) para avaliar alteração da voz, especificamente as alterações laríngeas benignas. Este método computacional é muito eficiente e não-invasivo, identificando com sucesso alteração laríngea pela análise homomórfica. A análise homomórfica, bem como a análise Cepstral, é a base teórica para a presente técnica, devido a sua propriedade matemática em isolar as frequências da excitação e do trato vocal, permitindo uma descrição precisa de comportamento da voz, a respeito das cavidades oral e nasal, assim como a laringe.

Este processamento do sinal de voz permitiu a obtenção de um novo protocolo diagnóstico, com precisão acima de 80 %, avaliados em casos de paciente reais.

Tal protocolo é original, com grandes benefícios para pesquisadores otorrinolaringologistas, assim como na prática diagnóstica otorrinolaringológica, por que ela é fácil e não-invasivo, sendo por isto muito adequado para as práticas clínicas diárias.

**Unitermos: alterações laríngeas, cepstrum, análise cepstral**

### **ABSTRACT**

In this work we will present a new Digital Signal Processing (DSP) procedure to evaluate voice disease, in a more specific way those ones involving benign larynx diseases. This computing method is very efficient and non-invasive, identifying successfully diseases by homomorphic analysis. Homomorphic analysis, well know as Cepstral analysis, has been the theoretical basis for this present technique, due to mathematic propriety of isolate excitation and vocal frequencies responses, allowing a precise description of voice behavior, concerning oral and nasal cavities, and larynx as well.

The use of this voice signal presentation, lead us to a new diagnostic protocol, as a direct result from this research, with accuracy in real patient cases above 80%.

Such protocol is original, with great benefits for otolaryngology research, as well as for the practice of otolaryngology diagnosis, because it is easy and non-invasive, therefore very suitable for day-by-day clinical practices.

Key words: larynx diseases, cepstrum, cepstral analysis

## Introdução

Este trabalho apresenta um modelo de análise da voz, com uso da técnica cepstral, como método de caracterização e diagnóstico de alterações laríngeas benignas. Atualmente o diagnóstico destas alterações laríngeas é realizado pelo exame de videolaringoscopia, que é um exame da laringe, onde se utiliza um sistema de ótica endoscópica para a visualização das pregas vocais. Esta ótica é conectada a uma microcâmara e o exame é gravado com o sinal de voz através de um microfone.

Certas alterações laríngeas, mesmo para médicos especialistas experientes, podem trazer dificuldade diagnóstica, pois às vezes são muito parecidas no aspecto, apesar de apresentarem origens e alterações fisiopatológicas diferentes. Estas dificuldades também são encontradas na realização de técnicas computacionais de processamento de sinais que, em determinados casos, não são eficientes o suficiente para a diferenciação das alterações laríngeas<sup>(1)</sup>.

No presente trabalho, aplicou-se a análise Cepstral em casos normais e nas seguintes alterações laríngeas: nódulo vocal, cisto vocal, pólipos vocais, edema de Reinke e sulco vocal.

Tais alterações laríngeas representam a grande maioria dos atendimentos de pacientes com alteração da voz, que não sejam alterações transitórias devido a infecções das vias aéreas superiores, onde temos uma inflamação das pregas vocais que, na grande parte, em alguns dias cedem com a melhora do quadro geral<sup>(2,3,4,5)</sup>.

Na maioria dos casos, as alterações laríngeas citadas produzem uma rouquidão com características típicas de cada uma, principalmente quando analisadas por ouvintes mais experientes, tais como médicos otorrinolaringologistas ou profissionais da área da fonoaudiologia<sup>(3,4)</sup>. Várias técnicas de análise de sinais da voz são estudadas para a identificação de alterações da laringe<sup>(1,6,7,8,9,10,11,12,13,14)</sup>.

A análise cepstral do sinal de voz para o estudo das alterações laríngeas da laringe é muito útil, permitindo trabalhar com o sinal da glote (excitação) separadamente das repercussões ressonantes do trato vocal, facilitando o entendimento das alterações que as alterações laríngeas causam nas pregas vocais.

A aplicação desta técnica de DSP no estudo do sinal acústico da vocalização de vozes alteradas permitirá detectar modificações nas ondas do sinal de voz que se relacionem com as alterações laríngeas e, conseqüentemente, a identificação de modelos para uma classificação, permitindo a obtenção de uma ferramenta de diagnóstico não-invasiva.

Determinando as características cepstrais das alterações laríngeas, tem-se uma valiosa fonte de informações para a obtenção de um método automático de reconhecimento de

alterações laríngicas da laringe, como por exemplo o uso das redes neurais artificiais ou modelos de Markov. Com isto pode-se determinar mais rapidamente e mais precisamente o diagnóstico de uma alteração laríngica, permitindo usar um tratamento mais adequado, o que possibilitará uma recuperação mais efetiva da alteração vocal do paciente, já que as alterações laríngicas citadas apresentam tratamentos distintos. Este é um método simples, rápido e de custos baixos quando comparado ao método diagnóstico convencional por videolaringoscopia, tornando-se por isto um método ideal para triagem de alterações vocais.

### Fisiopatologia das alterações das pregas vocais

A qualidade da voz depende do modo de fechamento e abertura da glote e da vibração das pregas vocais. Certas alterações laríngicas impedem que pregas vocais tenham uma vibração glotal harmônica. Os principais fatores que determinam a vibração vocal são:

- 1) posição da prega vocal, ou a extensão em que as pregas vocais são aduzidas ou abduzidas;
- 2) mioelasticidade, ou o grau de elasticidade das pregas vocais (determinado pela posição e grau de tensão decorrente da contração do músculo vocal);
- 3) nível de pressão do ar através das pregas vocais<sup>(15)</sup>.

As alterações laríngicas das pregas vocais também podem determinar que estas não vibrem em concordância, resultando em uma área vocal onde o trato vocal é excitado em duas frequências fundamentais diferentes<sup>(2,15)</sup>.

### Análise cepstral

A análise cepstral do sinal de voz permite trabalhar com o sinal da glote (excitação) e do trato vocal (ressonância) separadamente, pelas suas propriedades homomórficas, separando das características do filtro do trato vocal da seqüência de excitação, o que facilita o estudo das alterações que as alterações laríngicas causam nas pregas vocais. Dentro das propriedades matemáticas envolvidas no processo, salienta-se principalmente as transformadas de Fourier e funções logarítmicas que resultará em uma função chamada cepstral ou cepstro, responsável pela dissociação do sinal de voz.

O processo resumido de estimação do cepstro pode ser visto no diagrama de blocos apresentado na Figura 1:

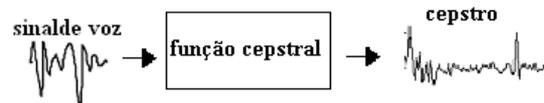


Fig.1: Diagrama de blocos da estimação do cepstro.

A concentração das componentes aparece no cepstro como picos. O eixo horizontal da função cepstral tem dimensões temporais e o nome de quefrências. Com isto, na voz se obtém uma clara distinção entre a componente de excitação e a contribuição do trato vocal, que aparece como um aglomerado de componentes aos baixos valores de quefrência afastado da componente do período fundamental que aparece com valores mais altos de quefrência.

Na Figura 2, vê-se o cepstro de um segmento de voz onde o pico correspondente ao período fundamental (excitação) está próximo da quefrência de 10 ms, separado das componentes do trato vocal às de baixas quefrências (trato vocal)<sup>(1,6,7,8,16,17)</sup>.

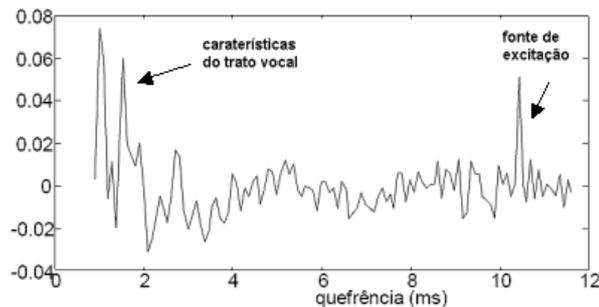


Fig.2: Cepstro de um segmento de fala.

## Liftering

A função de transferência do trato vocal e a função de excitação da voz aparecem em partes separadas da escala de quefrências, pelo que podem ser separadas as duas funções, ou removida uma delas por um processo de “lifteragem”, facilitando o estudo individualizado das alterações ocorridas na excitação e da parte ressonantal.

Assim, o estudo das alterações do sinal acústico de vozes alteradas, através da aplicação da técnica de análise cepstral na detecção de alterações no sinal relacionadas às alterações laríngeas e na identificação de modelos para a classificação das alterações laríngeas em estudo, permitirá obter uma ferramenta não-invasiva de diagnóstico.

## Objetivos

1- Desenvolver de técnica de processamento de sinais da voz, para identificação de alteração laríngea.

2- Criar protocolo de diagnóstico como método de reconhecimento das alterações da laringe a partir da avaliação dos gráficos do sinal fornecido pelo processamento.

## Resultados

Os achados cepstrais das vocalizações de cada alteração laríngea da tabela 1 foram analisados. Com a ajuda das características fisiopatológicas já conhecidas de cada tipo de alteração laríngea foram descritas as alterações percebidas, as quais podem ser particularmente atribuídas às diferentes alterações laríngeas.

Existem diferenças das alterações laríngeas no perfil cepstrográfico, quando comparadas com o normal e entre si. Estas diferenças se repetem na maioria dos casos analisados preliminarmente neste estudo.

Observando-se as diferentes alterações das alterações laríngeas em estudo, foi possível desenvolver um conjunto de parâmetros de análise, com os quais pôde-se diagnosticar de forma coerente as alterações laríngeas laríngeas, tomando como base os achados cepstrais de diferentes vocalizações. Para testar e validar este procedimento de linguística, casos escolhidos aleatoriamente, e que não participaram do estudo, foram aplicados e diagnosticados, somente com as informações da tabela. O resultado obtido, foi de 80% de acerto, quando os resultados dos gráficos são submetidos aos valores da tabela, validando significativamente o método proposto e contribuindo de forma efetiva para a elaboração de um protocolo de diagnóstico não-invasivo de alteração da laringe, baseado na técnica de análise cepstral.

## Discussão

Ao se observar as diferenças entre os resultados obtidos nas tabelas, percebe-se que existem diferenças entre as alterações laríngeas analisadas com o cepstro das diferentes vocalizações coletadas, sendo que algumas alterações laríngeas geram modificações significativas outras nem tanto.

Por exemplo, os achados do edema de Reinke são os que mais se diferenciam de todas as outras alterações laríngeas, pois esta alteração laríngea deixa o pitch muito agravado. O sulco

vocal é uma alteração laríngea que deixa a prega vocal rígida e este é um dos fatores que produzem alterações significativas na voz. Todas as alterações laríngeas tendem se diferenciar significativamente dos normais, colaborando com a idéia do trabalho em identificar as alterações laríngeas pelas alterações por elas produzidas no sinal acústico.

Também se destaca que algumas alterações laríngeas apresentam achados característicos e constantes que servem como método diagnóstico destas. Algumas alterações nem sempre são constantes ou, às vezes, estão presentes em mais de uma alteração laríngea, e isto deve ser mais aprofundado em trabalhos futuros. Acredita-se que este método inédito, por ser não invasivo, de custo mais baixo e fácil execução e com resultados promissores, como demonstrados nestes dados prévios será uma ferramenta de diagnóstico de triagem muito útil.

A Tabela 1 demonstra uma tabela com o agrupamento dos achados e as características principais obtidas no estudo.

Esta técnica de análise gráfica Cepstral fornece informações sobre as alterações laríngeas que não são demonstradas com outros métodos, principalmente no que se refere à visualização de regiões alteradas específicas e à possibilidade de relacioná-las com as alterações anatomofisiológicas já conhecidas para as alterações laríngeas, sendo isto uma inovação nesta área.

Tabela 1. Resumo das principais características de cada método.

	Diagnóstico com videolaringoscopia (convencional)	Método cepstral de análise da vocalização (sistema proposto)
Invasivo	Sim, pouco invasivo	Não
Execução	somente por profissional médico	pode ser executado facilmente com treinamento
Triagem	pouco prático para triagem	muito adequado para triagem
Custos	iguais de um procedimento em consultório	mais baixos que o convencional

## Conclusão

Com os resultados da proposta de análise cepstral do sinal da voz nas alterações laríngeas iniciados neste trabalho, permite-se prever que este método será uma ferramenta diagnóstica muito útil e promissora para este propósito, pois é um método não-invasivo, de custo mais baixo e fácil execução. Para uma maior qualificação deste método, deve-se realizar trabalhos futuros com a inclusão de mais amostras, de outras alterações laríngeas mais raras e de realização de estudo científico estatisticamente adequado para a validação, como um estudo prospectivo duplo cego.

Também é necessário desenvolver modelos de alterações laríngeas com o diagnóstico mais preciso possível para o treinamento do sistema, através de diagnóstico de certeza das alterações da laringe. O diagnóstico de certeza é obtido somente com uma biópsia da alteração laríngea, mas pode-se obter um diagnóstico confiável com o acompanhamento da mesma e sua evolução com o tratamento, obtendo o diagnóstico evolutivo, sem ser invasivo. Outro aspecto relevante é determinar as possibilidades de que o sexo da pessoa testada possa influenciar o resultado, já que as alterações laríngeas tendem a alterar o pitch da voz, como observado na pesquisa.

E, ao ser obtido uma modelagem ideal das alterações laríngeas com este método inovador, será possível usá-lo para treinamento mais preciso de um sistema de reconhecimento automático como, por exemplo, as redes neurais artificiais ou modelos de Markov<sup>(18)</sup>, tornando-se um equipamento de diagnóstico.

## Referências

1. Rosa MO, Pereira JC, GrellerM and CarvalhoA. “*Signal processing and statistical procedures to identify laryngeal pathologies*”, **Institut of Eletrical and Eletronics Engineers, Transactions on Biomedical Engineering** INSTITUT OF ELETRICAL AND ELETRONICS ENGINEERS/EESC-USP, 1999; 423-426.
2. Dedivitis RA.Barros, APB. Métodos de Avaliação e Diagnóstico de Laringe e Voz, 2ª ed Lovise, São Paulo, 2002.
3. Minoru H. Diane MB. Exame videoestroboscópico da laringe; Porto Alegre, Artes Médicas, 1997.

4. Erich CM, LupercioLB, Osiris CB, et al. Incidência de lesões laringicas não neoplásicas em pacientes com queixas vocais; *Revista Brasileira de Otorrinolaringologia*, vol.67, n.6, nov/dez 2001;788-94.
5. Khul I. *Manual prático de Laringologia*; Editora da universidade, Porto Alegre 1982.
6. Martinez CE, Rufiner HL. “*Acoustic analysis of speech for detection of laryngeal pathologies*”, *Proceedings of the 22<sup>nd</sup> Annual EMBS International Conference*, EMBS, Chicago, July 2000, 23-28.
7. Rabiner RL, Schafer R. *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
8. Rabiner R.L., Gold B. *Theory and Application of Digital Processing*, Prentice Hall, 1975.
9. Gomez P, Godino JI, Rodriguez F, et al. *Evidence of Vocal Cord Pathology from the mucosal wave Cepstral contents*, *Institut of Eletrical and Eletronics Engineers/Universidade Politécnica de Madrid*, Madrid, 2004;45-52.
10. Wilpon JG, Rabiner LR, Lee CH et al. “*Automatic recognition of keywords in unconstrained speech using Markov Models*”, *Institut of Eletrical and Eletronics Engineers, Transactions on Acoustic, Speech and Signal Processing*, Vol.38, N° 11, Nov.1990;1870-1878.
11. Hadjttodorov S, Mitev, P. “*A Computer system for acoustic of pathological voices and laryngeal diseases screening, Technical note*”, *Medical Engineering, & Physics*, Sofia, n. 24, 2002;419-429.
12. Manfredi C, D'aniello M, Brusciaglioni P, et al. “*A comparative analysis of fundamental frequency estimation methods with application to pathological voices*”, *Medical Engineering & Physics*, Firenze, n. 22, 2000, 135-147.
13. Wszolek W, Tadeusiewicz R, Izvorski A, et al. “*Automated understanding of selected voice tract pathologies based on the speech signal analysis*”, *Proceedings of the 23rd Annual EMBS International Conference*, EMBS, Istanbul, October, 2001;25-28.
14. Mitev P, Hadjttodorov S. “*Fundamental frequency estimation of voice of patients with laryngeal disorders*”, *Information Sciences*, Sofia, n. 156, 2003;3-19.
15. Hansen JH, Gavdida-Ceballos L, Kaiser RJF. “*A Nonlinear operator-Based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment*”. *Institut of Eletrical and Eletronics Engineers, Transactions on Biomedical Engineering*, Vol. 45, N°. 3, March 1998;937-940.
16. Furui S. *Digital Speech Processing, Synthesis and Recognition*, Marcel Dekker, Inc., 2001.
17. O' Shaughnessy D. *Speech Communications :Human and Machine*, *Institute of Electrical and Electronics Engineers*, New York, 2000.

18. Fagundes RDR. Reconhecimento de Voz, Linguagem Contínua , usando Modelos de Markov, Dissertação de Mestrado, Universidade de São Paulo, São Paulo, 1993.
19. Ingle VK, Proakis J. *Digital Signal Processing using MATLAB*, Brooks/Cole, Canadá, 2000.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)