

Universidade de São Paulo
Escola de Engenharia de São Carlos

Regiane Denise Solgon Bassi

Identificação Inteligente de Patologias no Trato Vocal

São Carlos - SP

2013

Universidade de São Paulo
Escola de Engenharia de São Carlos

Regiane Denise Solgon Bassi

Identificação Inteligente de Patologias no
Trato Vocal

Trabalho apresentado ao Programa de Pós-graduação em Engenharia Elétrica da Escola de Engenharia de São Carlos da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Área de concentração: Sistemas Dinâmicos.
Orientador: Prof. Dr. Ivan Nunes da Silva

Trata-se da versão corrigida da dissertação. A versão original se encontra disponível na EESC/USP que aloja o Programa de Pós-Graduação de Engenharia Elétrica.

São Carlos - SP

2013

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

B321i Bassi, Regiane Denise Solgon
Identificação inteligente de patologias no trato vocal / Regiane Denise Solgon Bassi; orientador Ivan Nunes da Silva. São Carlos, 2013.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em Sistemas Dinâmicos -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2013.

1. Processamento de sinais. 2. Patologias da laringe. 3. Distância euclidiana. 4. Redes neurais RBF.
I. Título.

FOLHA DE JULGAMENTO

Candidata: Licenciada **REGIANE DENISE SOLGON BASSI**.

Título da dissertação: "Identificação inteligente de patologias no trato vocal".

Data da defesa: 30/01/2014

Comissão Julgadora:

Resultado:

Prof. Associado **Ivan Nunes da Silva (Orientador)**
(Escola de Engenharia de São Carlos/EESC)

APROVADA

Prof. Dr. **Rodrigo Capobianco Guido**
(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP- São José do Rio Preto)

Aprovada

Profa. Titular **Agma Juci Machado Traina**
(Instituto de Matemática e Computação/ICMC)

APROVADA

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica e Presidente da Comissão de Pós-Graduação:
Prof. Titular **Denis Vinicius Coury**

*“Aos meus dois grandes mestres, Rodrigo Guido e Ivan Silva,
à minha família e aos meus amigos.”*

Agradecimentos

Tenho muito a agradecer, pois nunca estive sozinha nessa caminhada. Inicialmente, agradeço a Deus e a Nossa Senhora, pela iluminação muito necessária no desenvolvimento deste trabalho. Também, por ter me capacitado a vencer quando a vontade era desistir.

Agradeço ao meu orientador prof. Dr. Ivan Nunes da Silva pela confiança, pois sem a sua permissão, este trabalho não teria sido possível.

Ao prof. Dr. Rodrigo Capobianco Guido, um ser humano incrível, agradeço pela oportunidade de trabalhar ao seu lado, pelo companheirismo, pelos conselhos, pela sua amizade, por ter sido, sem dúvidas, o maior incentivador nos momentos difíceis.

A todas as minhas amigas que direta ou indiretamente sempre me incentivaram e se colocaram à minha disposição, em especial a uma grande amiga, Kátia Cristina Silva Paulo, incansável companheira de disciplinas, que sempre esteve ao meu lado, me apoiando, ajudando e algumas vezes até brigando, quando foi necessário.

Aos meus pais, por me ensinarem desde muito cedo, o trabalho, a perseverança e a importância de sonhar alto e aos meus irmãos e cunhado, por sempre estarem ao meu lado me incentivando, mesmo quando todas as circunstâncias eram contra, eles acreditaram que eu conseguiria. Da roça ao mestrado, um caminho penoso, mas uma prova de que o impossível não existe.

Ao meu marido, por estar sempre por perto e entender as minhas ausências, quando era necessário produzir.

Ao grupo, do *SpeechLab/IFSC/USP*, Sylvio, Leonardo, João Paulo, Lucimar, Everthon, Luciene, Fabrício e todos os outros que de alguma forma contribuíram para este projeto.

A todos do colégio Ressurreição - Rio Preto, em especial à Rita Raposo, minha amiga e coordenadora, que sempre me entendeu quando precisei faltar para fazer as disciplinas e à Edna Chierigato, secretária, por entender os meus atrasos na entrega das provas e das notas e à professora de Geografia, a Bárbara Gil, que sempre reza por mim.

Também a todos os meus amigos da Unorp, em especial o André Delorme e o Geraldo Zafalon que me ajudaram sempre que precisei e também ao pró reitor acadêmico, prof. Dr. José Luiz Falotico Corrêa, um grande amigo e admirador de todo este meu percurso.

Muito obrigada a todos!

*“Eu espero que eu sempre possua firmeza e
virtude suficientes para manter o que
eu considero o mais invejável de
todos os títulos, o caráter de
um homem honesto.”*

George Washington

Resumo

SOLGON, R.D.B *Identificação Inteligente de Patologias no Trato Vocal*. 2013. 75 p. Dissertação (Mestrado em Ciências) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2013.

Com base em exames como a videolaringoscopia, que é considerado um procedimento médico invasivo e desconfortável, diagnósticos têm sido realizados visando detectar patologias na laringe. Geralmente, esse tipo de exame é realizado somente com solicitação médica e quando alterações na fala já são marcantes, ou há sensação de dor. Nessa fase, muitas vezes a doença está em grau avançado, dificultando o seu tratamento. Com o objetivo de realizar um pré-diagnóstico computacional de tais patologias, este trabalho apresenta uma técnica não invasiva na qual são testados e comparados três classificadores: a Distância Euclidiana, a Rede Neural RBF com o *kernel* Gaussiano e a Rede Neural RBF com o *kernel* Gaussiano modificado. Testes realizados com uma base de dados de vozes normais e aquelas afetadas por diversas patologias demonstram a eficácia da técnica proposta, que pode, inclusive, ser implementada em tempo-real.

Palavras-chave: Processamento de sinais, Patologias da laringe, Distância Euclidiana, Redes Neurais RBF.

Abstract

SOLGON, R.D.B. *Intelligent Detection of Pathologies in the Vocal Tract*. 2013. 75 p. Dissertação (Mestrado em Ciências) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2013.

Based on examinations such as laryngoscopy, which is considered an invasive and uncomfortable procedure, diagnosis have been performed aiming at the detection of larynx pathologies. Usually, this type of test is carried out upon medical request and when the speech changes are notable or are causing pain. At this point, the disease is possibly at an advanced degree, complicating its treatment. In order to perform a computational pre-diagnosis of such conditions, this work proposes a noninvasive technique in which three classifiers are tested and compared: the Euclidean distance, the RBF Neural Network with the Gaussian kernel and RBF Neural Network with a modified Gaussian kernel. Tests carried out with a database of normal voices and those affected by various pathologies demonstrate the effectiveness of the technique that may even be implemented to work in real time.

Key-words: Signal processing, Larynx pathologies, Euclidian Distance, RBF Neural Networks.

Lista de Figuras

2.1	Esquema de funcionamento do sistema respiratório (adaptado do site [28]).	28
2.2	[Esquerda]: Esquema da laringe, (adaptado do site [13]); [Direita]: Localização da laringe, (adaptado do site [14])	29
2.3	Esquema articulatorio: Cavidade nasal (adaptado do site [27]).	30
2.4	Desenho esquemático do sistema fonte: filtro da produção de voz (adaptado de SCALAS-SARA [4]-modificado).	31
2.5	Pregas vocais (adaptado de RITCHINGS [9]).	32
2.6	Calos das pregas vocais. (adaptado de JO [10]).	33
2.7	Pólipos (adaptado de FROHLICH [7] e de KENT[12], respectivamente).	34
2.8	Cistos (adaptado de [17] e de DELLER[16], respectivamente.)	35
2.9	Edemas (inchaços) de Reinke (adaptado de [18] e de [19], respectivamente.)	35
2.10	Modelos de filtros ideais.	40
2.11	Curva de Resposta em Frequência para um Filtro Passa Faixa (adaptado de [32]).	41
2.12	A arquitetura de uma rede RBF, onde os W_n são os pesos.	44
2.13	A função de ativação com base radial.	45
2.14	Gráfico da função Gaussiana.	46
3.1	Esquema do processamento de um sinal de voz no presente trabalho.	47
3.2	Arquitetura básica para treinamento do sistema de detecção e classificação de patologias do trato vocal.	53
3.3	Arquitetura para testes e uso do sistema de detecção de patologias.	54

Lista de Tabelas

2.1	Principais patologias da laringe (adaptado de [20]).	36
2.2	Escala <i>Bark</i>	43

Lista de Abreviaturas

DFT	<i>Discrete Fourier Transform</i>
FFT	<i>Fast Fourier Transform</i>
RBF	<i>Radial Basis Function</i>
PDS	Processamento Digital de Sinais
RNA	Redes Neurais Artificiais
WAV	<i>Waveform Audio Format</i>

Sumário

1	Introdução	23
1.1	Motivação	23
1.2	Objetivos e Contribuições	24
1.3	Organização do Trabalho	24
2	Revisão Bibliográfica	27
2.1	Fisiologia da Voz	27
2.1.1	O Subsistema Respiratório	27
2.1.2	O Subsistema da Laringe	28
2.1.3	O Subsistema Articulatorio	29
2.1.4	A produção da voz	30
2.2	Patologias na Laringe	32
2.2.1	Calos / Nódulos / Pólipos / Cistos	33
2.2.2	Edema de Reinke	35
2.2.3	Procedimentos Tradicionais de Diagnóstico Médico	36
2.3	Fundamentação Teórica	37
2.3.1	Elementos Estatísticos	37
2.3.2	Matriz de Confusão	38
2.3.3	Energia	39
2.3.4	Filtragem Digital	39
2.3.5	Filtros Digitais	40
2.3.6	Filtragem via Convolução	42

2.3.7	Bandas Críticas, Escala Bark e Amostragem	42
2.3.8	Redes Neurais RBF (<i>Radial Basis Functions</i>)	44
2.3.9	A Função Gaussiana	46
3	A Abordagem Proposta	47
3.1	Estrutura do Sistema	47
3.2	Classificador I: Técnica Baseada na Distância Euclidiana	48
3.3	Classificador II: Rede Neural RBF e <i>Kernel</i> Gaussiano	49
3.4	Classificador III: Rede Neural RBF e <i>Kernel</i> Gaussiano Modificado	51
3.5	Implementação	54
4	Testes e Resultados	55
4.1	Métodos e Materiais Utilizados	55
4.2	Procedimentos de Testes e Respective Resultados com o Uso do Classificador I - Distâncias Euclidianas	55
4.3	Procedimentos de Testes e Respective Resultados com o Uso do Classificador II - Rede Neural RBF e <i>Kernel</i> Gaussiano	57
4.4	Procedimentos de Testes e Respective Resultados com o Uso do Classificador III - Rede Neural RBF e <i>Kernel</i> Gaussiano Modificado	60
5	Conclusões e Trabalhos Futuros	63
	Referências	65
	Apêndice I - Código fonte do sistema proposto.	69
	Apêndice II - Publicações durante o mestrado.	77

Capítulo 1

Introdução

Neste capítulo serão abordados a motivação, as contribuições, os objetivos e a organização deste trabalho.

1.1 Motivação

Baseado no trabalho de GUIDO [1], que propõe a classificação de padrões por meio de Redes Neurais Artificiais, assim como nos trabalhos de BARBON JR [2], que abordou a Lógica Paraconsistente (*Discriminative Paraconsistent Machine - DPM*) e de SOUZA [3], que trabalhou com as (*Support Vector Machine - SVMs*), recentemente desenvolvidos junto ao laboratório *SpeechLab* do IFSC/USP, constata-se que a voz carrega características importantíssimas do aparelho fonador, fornecendo, assim, toda a informação necessária para a sua análise e a detecção de patologias na laringe do respectivo locutor.

De acordo com SCALASSARA [4], a presença de nódulos ou edemas, originados principalmente por esforço demasiado ou uso de tabaco, modifica a dinâmica do sistema de produção de voz, gerando pequenas variações em parâmetros acústicos do trato vocal, quando observados ao longo do tempo de fonação. De acordo com BARBON JR [2], BAKEN [5], CHILDERS [6] e FROHLICH [7], a busca de novas técnicas que visam a melhoria da qualidade de vida dos seres humanos é um dos principais estímulos à pesquisa, não só na criação de procedimentos, mas também no aprimoramento de abordagens. Assim, surge a necessidade, conforme proposto

em DAJER [8], de uma técnica de análise que permita a visualização da dinâmica do sistema vocal e que seja de baixo custo e de caráter não invasivo. Tais fatos motivaram, inicialmente, o desenvolvimento do presente trabalho. Em vista do desconforto causado pelas técnicas citadas, justifica-se este projeto, já que o mesmo fornece um meio não invasivo para o pré-diagnóstico das doenças.

1.2 Objetivos e Contribuições

As patologias mais comuns que ocorrem na laringe, tais como os edemas, causados geralmente pelo uso abusivo do cigarro, os nódulos, gerados pelo esforço demais da voz, entre outras, interferem na dinâmica do sistema de produção de voz, gerando variações em parâmetros acústicos do trato vocal. Com base nesse fato, o principal objetivo deste trabalho é o de elaborar uma técnica que contribua para o pré-diagnóstico não-invasivo de patologias na laringe, visando prover um maior apoio às decisões médicas, como proposto em RITCHINGS [9] e em JO [10].

Parâmetros acústicos baseados nas concentrações médias e nos níveis de variabilidade das energias de determinadas sub-bandas de frequências serão explorados no presente trabalho. Tais elementos servirão de base para alimentar um classificador do tipo Função de Base Radial (RBF - *Radial Basis Function*), segundo MARTINEZ [11], para o qual um núcleo (*kernel*) particular será experimentado, contrastando-o com o tradicional núcleo Gaussiano e também com a análise feita por meio da distância Euclidiana, geralmente adotados em experimentos do presente tipo. Tal modelamento constitui a contribuição que o presente trabalho oferece para impulsionar as pesquisas na área.

1.3 Organização do Trabalho

Este trabalho está organizado da forma como segue. No capítulo 2 é apresentada uma revisão da literatura, abrangendo as principais características do sistema de fala humana e das suas patologias, das redes RBFs, além de outros conceitos importantes. No capítulo 3, é descrito, com detalhes, o algoritmo proposto. Os resultados obtidos com base no sistema desenvolvido

constituem o capítulo 4. Tendo em vista os resultados, e apoiado nos conceitos e características estudados durante a revisão de literatura, o capítulo 5 apresenta as conclusões das atividades. Por fim, encontram-se as referências bibliográficas e os apêndices.

Capítulo 2

Revisão Bibliográfica

Neste capítulo, é apresentada uma revisão da literatura, abordando basicamente os seguintes tópicos: visão geral do sistema gerador da fala humana e das suas principais patologias, conceitos de processamento de sinais, assim como caracterização das redes neurais RBFs, que constituem os classificadores inteligentes adotados como ponto de partida para formulação do sistema proposto. Todos esses conceitos são importantes para a perfeita compreensão do sistema proposto neste trabalho.

2.1 Fisiologia da Voz

Segundo KENT [12], o aparelho da fala pode ser dividido em três subsistemas anatômicos principais: respiratório, laringal e articulatório, os quais são descritos a seguir.

2.1.1 O Subsistema Respiratório

O aparelho respiratório é imprescindível para o ato da fala, pois é como uma bomba de ar que fornece a energia aerodinâmica aos subsistemas laríngeo e articulatório, ou seja, é a corrente de ar indispensável à produção dos sons na voz humana. Os sons são formados no momento da expiração por meio da regulação da pressão de ar subglotal. O subsistema respiratório é formado pelos pulmões, traqueia, diafragma e brônquios, como ilustrado na Figura 2.1.

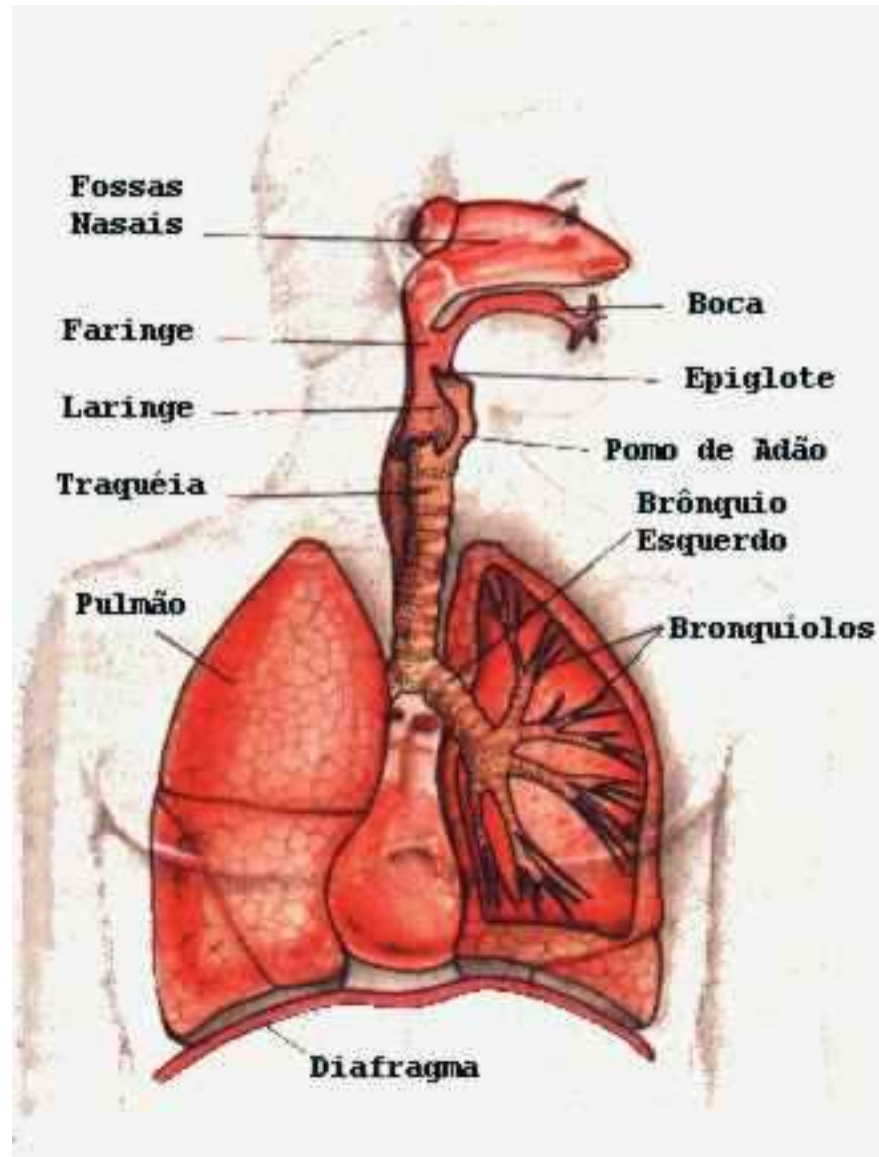


Figura 2.1 – Esquema de funcionamento do sistema respiratório (adaptado do site [28]).

2.1.2 O Subsistema da Laringe

A laringe é um tubo sustentado por peças de cartilagem articuladas e unidas por ligamentos, membranas e músculos. É uma câmara oca onde a voz é produzida. Encontra-se na parte superior da traqueia, em continuação à faringe. O pomo-de-adão, que aparece como uma saliência na parte anterior do pescoço, logo abaixo do queixo, é uma das peças cartilaginosas da laringe. A entrada da laringe é chamada glote. Acima dela existe uma espécie de “lingueta” de cartilagem, denominada epiglote, que funciona como uma válvula. A laringe é unida por meio de ligamentos ao osso hioide, situado na base da língua.

O revestimento interno da laringe apresenta pregas, denominadas pregas vocais; no ato da fala um conjunto de músculos as fazem encontrar-se. A laringe tem um par de pregas vocais, formadas por tecido conjuntivo elástico. A vibração que o ar proveniente dos pulmões provoca neste par de cordas forma os sons glotais. Os sons produzidos na laringe são modificados pela ação da faringe, da boca, da língua e do nariz, o que nos permite articular os diversos fonemas. Na Figura 2.2 é possível observar os elementos descritos.

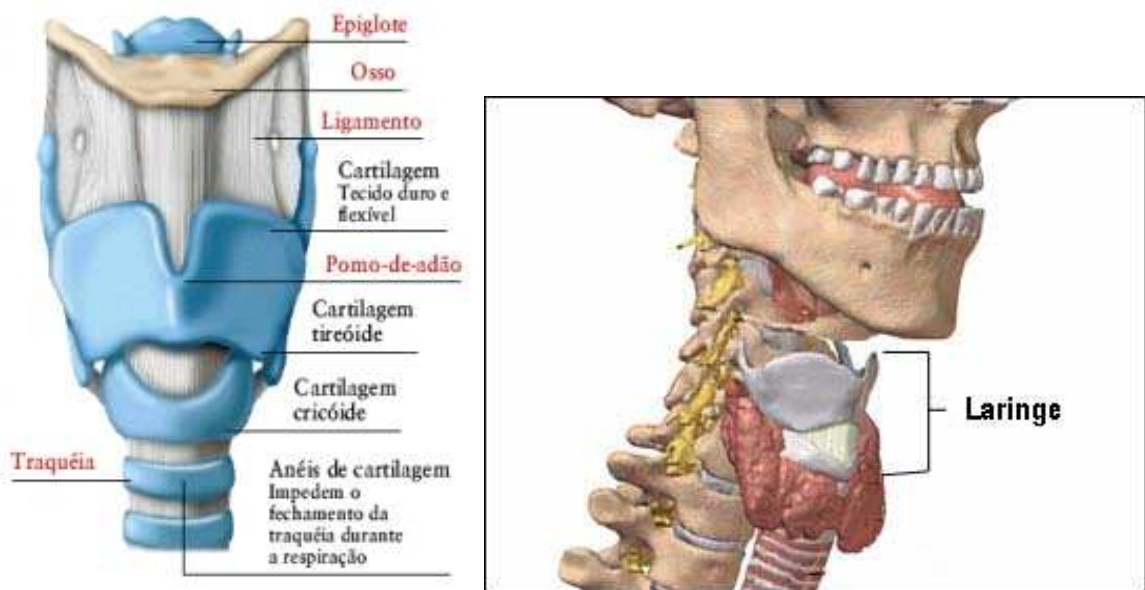


Figura 2.2 – [Esquerda]: Esquema da laringe, (adaptado do site [13]); [Direita]: Localização da laringe, (adaptado do site [14])

2.1.3 O Subsistema Articulatório

O subsistema articulatório modifica os sons glotais, alterando o seu conteúdo harmônico. O sistema vocal é formado pela língua, lábios, mandíbula, palato mole e palato duro, sendo que o seu movimento determina as propriedades ressonantes que formam a voz. A cavidade faríngeo-bucal pode estar associada à cavidade nasal quando o palato mole não está acoplado à parede posterior da faringe. Nesse caso, é permitida a passagem do ar, ao mesmo tempo, pela boca e pelas fossas nasais, e é a articulação do palato mole que determina se um som é oral ou nasal, segundo KENT [12]. Na Figura 2.3, é possível observar alguns dos elementos descritos.

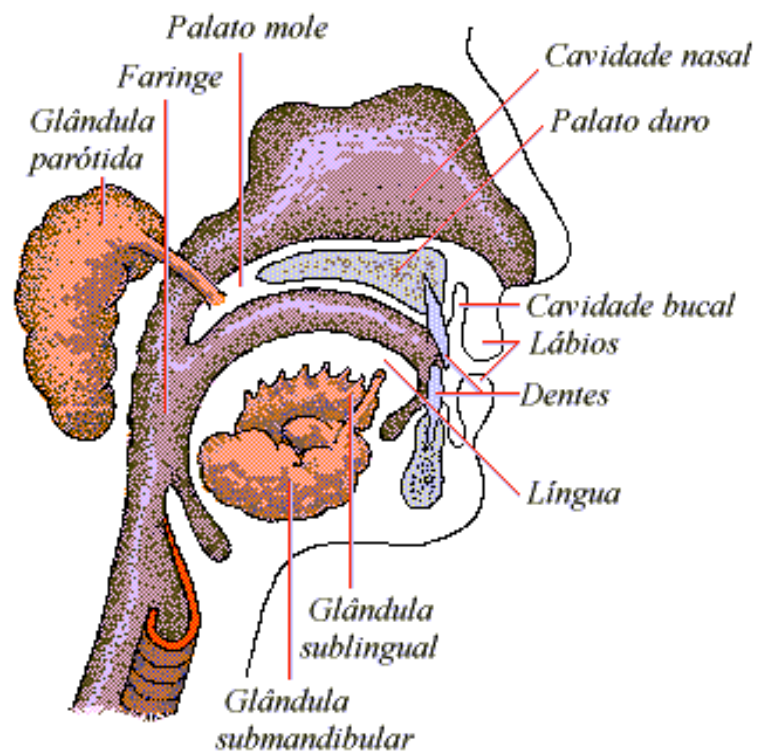


Figura 2.3 – Esquema articulatorio: Cavidade nasal (adaptado do site [27]).

2.1.4 A produção da voz

A voz é produzida quando o ar expiratório (vindo dos pulmões) passa pelas pregas vocais e, por nosso comando neural que produzem ajustes musculares, faz pressões de diferentes graus na região abaixo das pregas vocais, fazendo-as vibrarem. Esse mecanismo se assemelha ao balão (“bexiga de festas”), quando, apertando sua “boca”, provoca-se um ruído agudo, fruto da vibração da borracha.

O ar expiratório, que faz as pregas vocais vibrarem, vai sendo modificado e os sons vão sendo articulados (vogais e consoantes). Depois, emitidos pela boca, criam a onda sonora que vai atingir a cóclea do ouvinte. Então, a voz é ouvida. As pregas vocais vibram muito rapidamente. Nos homens, esse número de ciclos vibratórios fica em torno de 125 vezes em um segundo. Na mulher, que tem voz, geralmente, mais aguda, o número aumenta para 250 vezes por segundo. As pregas vocais do homem têm mais massa e são menos esticadas que as da mulher (como no violão, as cordas mais esticadas são mais agudas e vibram mais que as cordas mais graves).

Do ponto de vista acústico, a produção da voz ocorre na laringe, quando da passagem do ar expirado por entre as pregas vocais, modificando-se pelas estruturas do trato vocal, o que confere uma personalidade às vozes. Na passagem de ar, as pregas vocais controlam seu fluxo, fazendo com que esse sinal de excitação seja periódico, vibrando em determinada frequência. Nesse caso, a voz é classificada como vozeada (*voiced speech*) e esse período é chamado de período de *pitch* ou aperiódica, similar a um sinal ruidoso, ou como não-vozeada (*unvoiced speech*). Na Figura 2.4 é possível observar o sistema descrito.

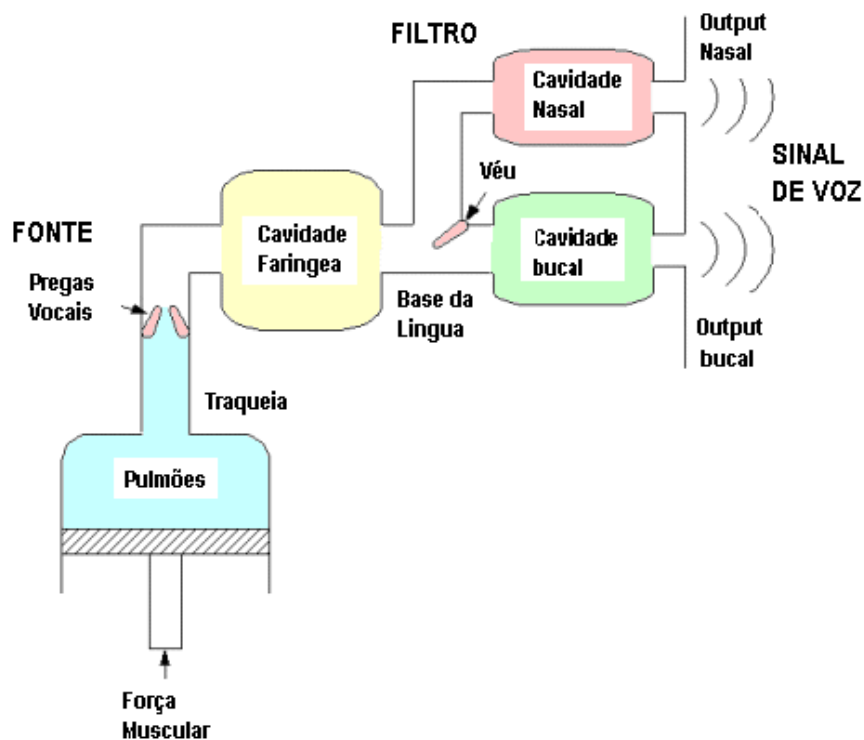


Figura 2.4 – Desenho esquemático do sistema fonte: filtro da produção de voz (adaptado de SCALAS-SARA [4]-modificado).

A taxa de vibração das pregas vocais costuma variar de 60 até 300 *Hz*, variação que depende principalmente do sexo e da idade, entre outras particularidades do falante, segundo BARBON JR [2]. Fisiologicamente falando, quanto mais longas e espessas as pregas vocais, mais lentas as suas vibrações, e quanto mais curtas e finas, maior a frequência. O tempo que as pregas vocais gastam para abrir e fechar duas vezes consecutivas é chamado de período fundamental (T_0), enquanto que a taxa de vibração é chamada de frequência fundamental ($F_0 = \frac{1}{T_0}$). O período fundamental é dependente do tamanho e da tensão das pregas vocais do falante em um

determinado instante, como proposto em DELLER [16]. Na Figura 2.5, podem-se observar as pregas vocais de um indivíduo.

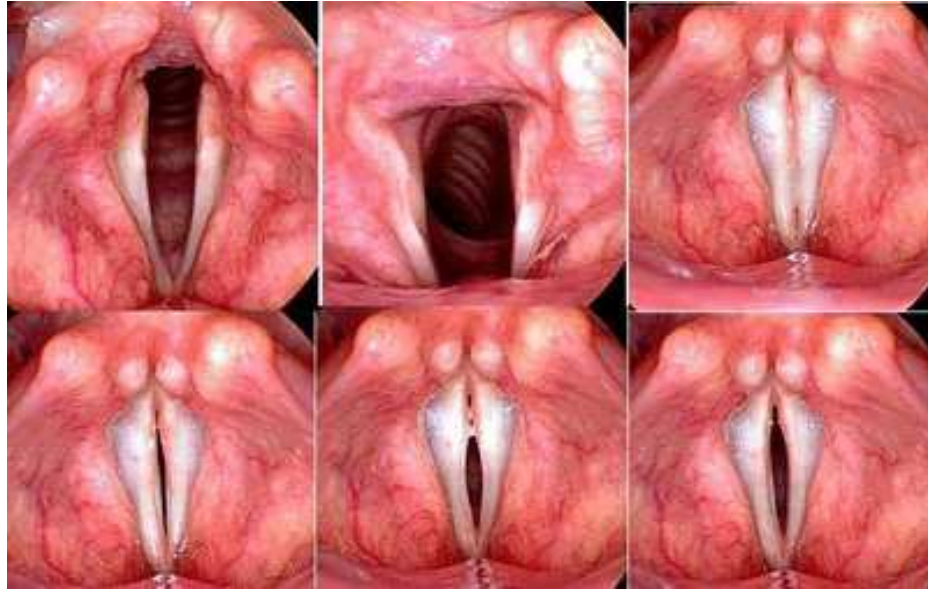


Figura 2.5 – Pregas vocais (adaptado de RITCHINGS [9]).

2.2 Patologias na Laringe

Segundo PARRAGA [15], a maioria das doenças na laringe provoca mudanças na voz do paciente. O sinal mais comum que pode indicar uma alteração na laringe é a rouquidão. Uma rouquidão passageira pode ser resultante de excessos ou de uma gripe ocasional. Quando, entretanto, a rouquidão se instala e passa a ser característica da voz, indicando uma alteração mais séria como, por exemplo, uma disfonia, é recomendável investigar o que está ocorrendo. As patologias mais comuns, segundo CUNHA COSTA [29], que afetam a voz são os nódulos vocais, os pólipos, cistos e os edemas (inchaço) de Reinke. Outras patologias de laringe que podem levar a uma disfonia são a úlcera de contato, o granuloma e a leucoplasia [30]. De qualquer forma, uma alteração de voz, como a rouquidão, também pode ser um dos primeiros sintomas de um câncer de laringe [31].

2.2.1 Calos / Nódulos / Pólipos / Cistos

As lesões mais frequentes das pregas vocais são os nódulos vocais, também conhecidos como “calos” das pregas vocais, visíveis na Figura 2.6. Habitualmente, surgem simetricamente nas duas pregas, esbranquiçadas e rígidas. Trata-se de um espessamento em determinado ponto delas, causando, na maioria das vezes, irregularidade nas suas laterais. Suas principais consequências são a disfonia (rouquidão) prolongada, dor e sensação de fadiga na região da garganta, além da voz que tende a sumir depois de algum tempo de fonação.



Figura 2.6 – Calos das pregas vocais. (adaptado de JO [10]).

São mais comuns em mulheres adultas, entre 25 a 40 anos, muitas vezes com atividades que exijam grande quantidade de fala e também em crianças - mais em meninos do que em meninas - por volta dos 7 aos 9 anos. O diagnóstico normalmente é feito pela videolaringoscopia, e o tratamento, inicialmente, exige repouso vocal e/ou reabilitação (terapia vocal), procurando-se melhorar a técnica vocal, reduzindo ou anulando o esforço e a tensão das pregas vocais. A cirurgia é realizada em poucos casos selecionados e somente após verificar-se que a terapia vocal não trouxe o resultado esperado.

Os pólipos, como pode-se notar na Figura 2.7, são lesões de massa geralmente unilaterais, de características gelatinosas e que variam de tamanho e coloração.

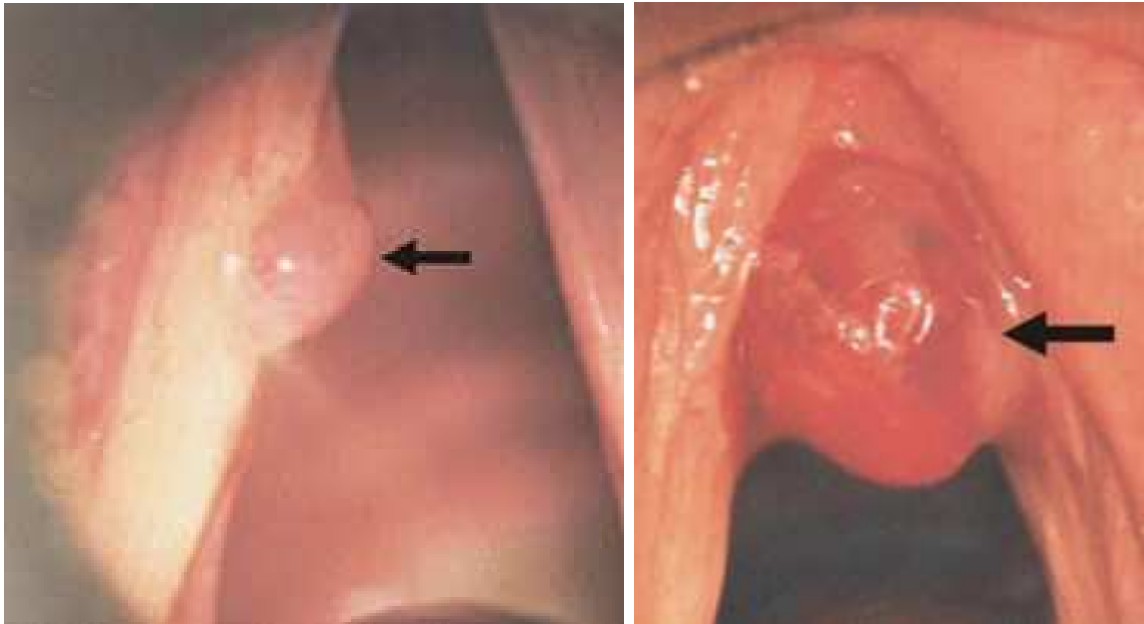


Figura 2.7 – Pólipos (adaptado de FROHLICH [7] e de KENT[12], respectivamente).

Essas inflamações são decorrentes de traumas em camadas mais profundas da lâmina própria da laringe, normalmente causadas por um único incidente como, por exemplo, um grito em uma partida de futebol. De caráter benigno, suas principais causas são o esforço vocal, oriundo dos atos de sussurrar, de falar muito alto, de pigarrear, ou são provenientes de alergias, infecções agudas e também do tabagismo. Ocorrem com maior frequência em homens adultos com idade entre 35 e 45 anos, sendo raros em crianças. Os pacientes com pólipos apresentam uma rouquidão característica, com sopro variável, e suas queixas geralmente são de uma sensação de corpo estranho na garganta. O tratamento dos pólipos nas pregas vocais é quase sempre cirúrgico. A possibilidade de recuperação espontânea ou com fonoterapia, como no caso dos calos, é muito reduzida. Após a remoção cirúrgica, se não for observada a normalização da voz, em duas semanas, fica indicada a reabilitação vocal, geralmente de curta duração, segundo o INCA [31].

Os cistos são lesões benignas, esbranquiçadas, amareladas ou avermelhadas, uni- ou bilaterais, como pode ser visto na Figura 2.8. Podem ocorrer em qualquer idade e são causados

por fatores embriogénéticos. O indivíduo pode queixar-se de cansaço na voz e de que sua voz está diferente, alterada, rouca ou áspera.



Figura 2.8 – Cistos (adaptado de [17] e de DELLER[16], respectivamente.)

2.2.2 Edema de Reinke

A camada superficial da lâmina própria é também conhecida como espaço de Reinke, onde se desenvolvem os edemas (inchaços) de Reinke. São lesões bilaterais alargadas, de coloração rosada, com fenda irregular e pouca abertura respiratória, como pode ser observado na figura 2.9.

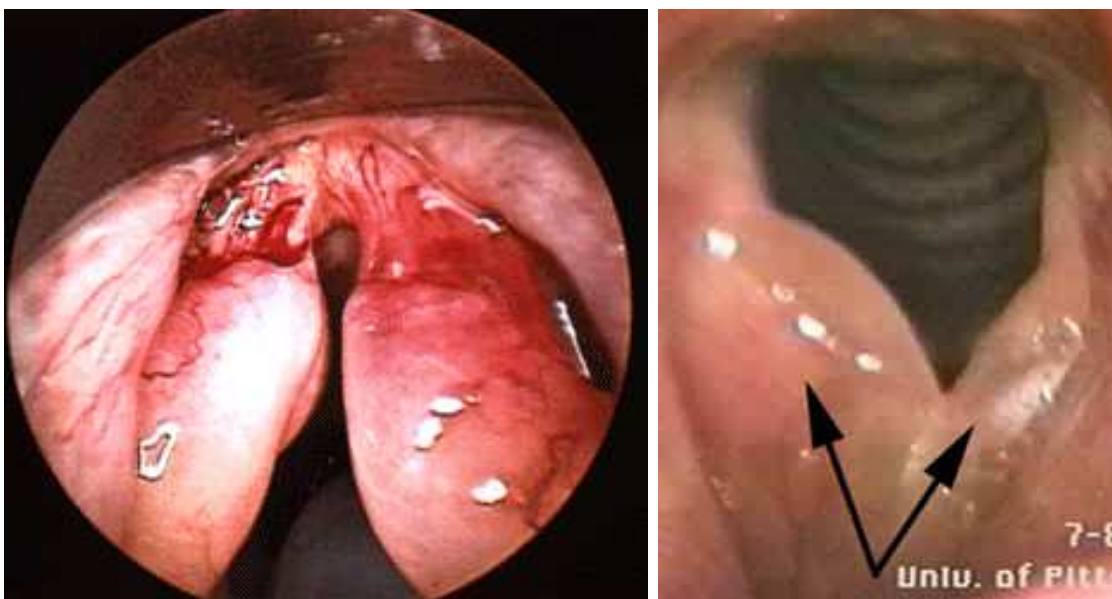


Figura 2.9 – Edemas (inchaços) de Reinke (adaptado de [18] e de [19], respectivamente.)

Normalmente, é bem localizado e agudo. O tratamento é medicamentoso ou baseado em repouso vocal. É encontrado em pessoas expostas a fatores irritantes externos, especialmente o tabagismo, sendo que o mais importante fator está associado ao uso excessivo e abusivo da voz. Se discretos, os edemas podem ser tratados com medicamentos e até fonoterapia, caso seja assegurada a eliminação de seu fator causal; se volumosos, necessitam de remoção cirúrgica, seguida de reabilitação fonoaudiológica.

Na tabela 2.1, consta um resumo das principais patologias da laringe.

Tabela 2.1 – Principais patologias da laringe (adaptado de [20]).

Patologias	Localização	Características Vocais
<i>Edema de Reinke</i>	Ao longo de toda a prega vocal	voz grave, fluída, rouca ou crepitante, no início agradável; virilizada.
<i>Nódulo</i>	Metade da porção vibratória; junção do $\frac{1}{3}$ anterior e do $\frac{1}{3}$ médio da prega vocal	voz rouco-soprosa, grau discreto a extremo, ataques bruscos, ressonância baixa, alguns casos sem sinais vocais.
<i>Úlcera de contato</i>	Porção cartilaginosa da prega vocal	voz grave, crepitante, disfonia silente, odinofonia, odinofagia e otalgia.
<i>Calos</i>	Borda livre da prega vocal	rouquidão, afonia, ardor na garganta, pigarro constante, cansaço ao falar.
<i>Pólipo</i>	Borda livre da prega vocal	voz grave, rouca, rouco-soprosa, frequência dicrótica, incoordenação fadiga vocal.

2.2.3 Procedimentos Tradicionais de Diagnóstico Médico

A avaliação otorrinolaringológica é comumente utilizada para o diagnóstico das doenças que podem afetar a voz. Tudo começa na entrevista detalhada, com informação sobre a qualidade da queixa vocal, duração dos sintomas, evolução, tratamentos prévios, investigação de causas associadas, identificação de fatores irritantes às vias aéreas, tipo de atividade profissional, entre outras. Após a avaliação otorrinolaringológica geral, a investigação armada, isto é,

por meio de exames, é o próximo passo.

O exame mais comum é o de videolaringoscopia (exame com um instrumento de fibra ótica), que com ótica telescópica permite até mesmo a participação do paciente no próprio exame, facilitando assim a sua compreensão diagnóstica. A videoestroboscopia, que consiste na iluminação estroboscópica da laringe, útil para visualização dos movimentos, permite a visualização do comportamento vibratório das pregas vocais, como aborda COLTON [21]. Existe ainda a eletromiografia, que é a observação indireta do estado funcional da laringe, além da videofluoroscopia, que constitui uma técnica radiográfica em razão da qual o paciente ingere uma determinada quantidade de substância rádio-opaca para avaliar a deglutição, segundo MARTINEZ [11]. Tais exames não deixam de ser relativamente desconfortáveis, por causarem, entre outros inconvenientes, náuseas, ansiedade e até dores aos pacientes. Segundo PARRAGA [15], essas técnicas visuais resultam em uma avaliação qualitativa, de resultados difíceis de serem quantificados, e que necessitam do conhecimento e da experiência do avaliador.

2.3 Fundamentação Teórica

2.3.1 Elementos Estatísticos

No presente trabalho, são utilizados os conceitos de média e desvio-padrão, tradicionais da análise estatística. A média dos valores de um vetor $x[\cdot]$ de N pontos, isto é,

$$\mu(x[\cdot]) = \frac{1}{N} \sum_{i=1}^N x_i, \quad , \quad (2.1)$$

possibilita obter uma ideia da tendência dos valores contidos no vetor. A dispersão de tais valores, que é conhecida como variância estatística e é definida por

$$\sigma^2(x[\cdot]) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad , \quad (2.2)$$

tem o objetivo de analisar o grau de variabilidade de um conjunto de valores. Por meio dela, pode-se perceber desempenhos iguais, muito próximos ou muito distantes; ela determina de forma mais específica as possíveis variações no intuito de não comprometer os resultados da análise.

A raiz quadrada positiva da variância, simbolizada por σ , é conhecida como desvio-padrão. Outro conceito oriundo da variância é a quasi-variância, isto é,

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2 \quad , \quad (2.3)$$

que constitui um estimador não-viciado, isto é, não-tendencioso. Neste trabalho, a modificação introduzida no *kernel* Gaussiano, descrita no capítulo seguinte, baseia-se na quasi-variância.

2.3.2 Matriz de Confusão

Dentro da área de reconhecimento de padrões, uma matriz de confusão, segundo [35], é um *layout* de tabela específico que permite a visualização do desempenho de um algoritmo, normalmente utilizada numa aprendizagem supervisionada, (em uma aprendizagem não supervisionada, geralmente é chamada matriz de correspondência). Cada coluna da matriz representa as instâncias de uma classe prevista, enquanto que cada linha representa as instâncias de uma classe real. O nome deriva do fato de que ela torna mais fácil para ver se o sistema está confundindo duas classes (ou seja, comumente classificar um como outro). Por exemplo, se um sistema de classificação foi treinado para distinguir entre gatos, cães e coelhos, uma matriz de confusão irá resumir os resultados de testar o algoritmo para uma inspeção mais aprofundada. Supondo que uma amostra de 27 animais: 8 gatos, 6 cães, e 13 coelhos, a matriz de confusão resultante pode parecer com a matriz abaixo:

	Gato	Cachorro	Coelho
Gato	5	3	0
Cachorro	2	3	1
Coelho	0	2	11

Nesta matriz de confusão, dos 8 gatos, o sistema preveu que 3 eram cães, e dos 6 cães, que 1 era coelho e 2 eram gatos. Podemos ver a partir da matriz, que o sistema em questão tem dificuldade em distinguir entre cães e gatos, mas pode fazer a distinção entre coelhos e outros tipos de animais muito bem. Todos os palpites corretos estão localizados na diagonal da matriz, por isso é fácil de inspecionar visualmente, já que será representado por qualquer valor diferente de zero fora da diagonal. Por esse motivo, o presente trabalho apresenta os resultados em matrizes de confusão.

2.3.3 Energia

A energia de um sinal $x[\cdot]$ com N amostras discretas é um escalar definido, segundo FONSECA [22], como

$$E = \sum_{i=1}^N x_i^2 \quad . \quad (2.4)$$

Neste trabalho, as energias de várias sub-bandas de frequências de um sinal serão utilizadas para alimentar o sistema de classificação adotado.

2.3.4 Filtragem Digital

O grande motivo de expressar um sinal temporal discreto, tal como é um sinal de voz, no domínio frequencial, é a possibilidade de realizar uma inspeção espectral e até mesmo fazer a filtragem daquelas frequências que não são desejadas. Segundo HAYKIN [24], o processo mais custoso de filtragem de sinais digitais consiste em calcular a Transformada Discreta de Fourier (DFT) do sinal, multiplicá-la por outro sinal discreto que atua como filtro e, por fim, calcular a DFT inversa deste último sinal.

O grande problema dessa abordagem é que existe a necessidade de dois cálculos de DFT. Se o sinal possuir muitas amostras, a computação se torna um processo lento e complexo. Uma alternativa para melhorar esse procedimento é o uso da Transformada Rápida de Fourier (*Fast Fourier Transform* - FFT). Entretanto, existe ainda um outro processo equivalente, usado neste trabalho, mais simples e que produz exatamente o mesmo resultado, conhecido como convolução.

2.3.5 Filtros Digitais

Filtros são elementos especialmente projetados para fornecer sinais de saída com uma amplitude dependente da frequência do sinal aplicado em suas entradas. A curva de resposta em frequência é uma representação gráfica da amplitude do sinal na saída de um filtro, em função da frequência do sinal aplicado em sua entrada. De acordo com a resposta em frequência, os filtros podem ser classificados como passa-baixa, passa-alta, passa-faixa e rejeita-faixa, como mostrado na Figura 2.10.

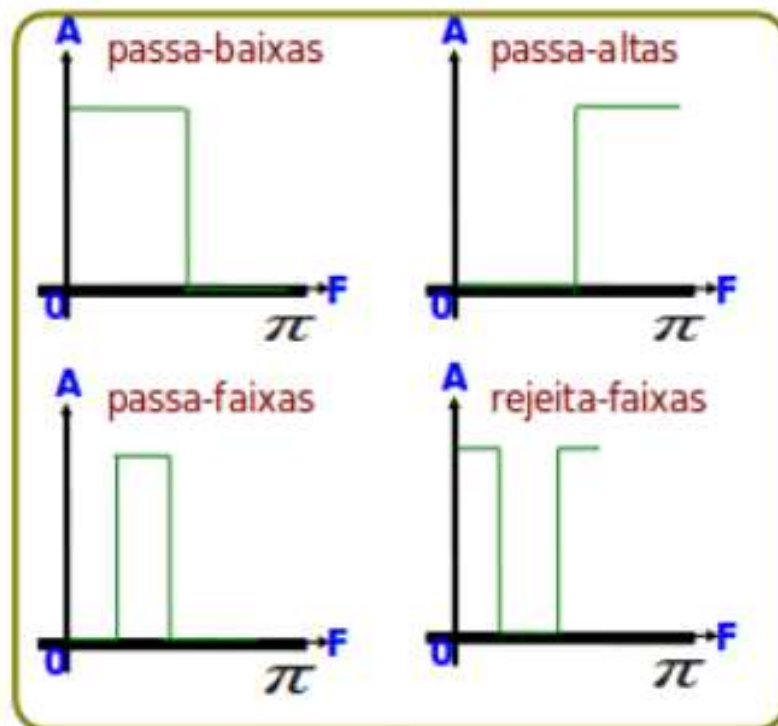


Figura 2.10 – Modelos de filtros ideais.

Na Figura 2.11 consta um exemplo da resposta em frequências de um filtro passa-faixas, que é o modelo de filtro utilizado no presente trabalho. Para sinais de frequência média, ou seja, acima da frequência de corte inferior e abaixo da frequência de corte superior do filtro, o ganho é unitário, portanto o módulo do sinal de saída é igual ao de entrada.

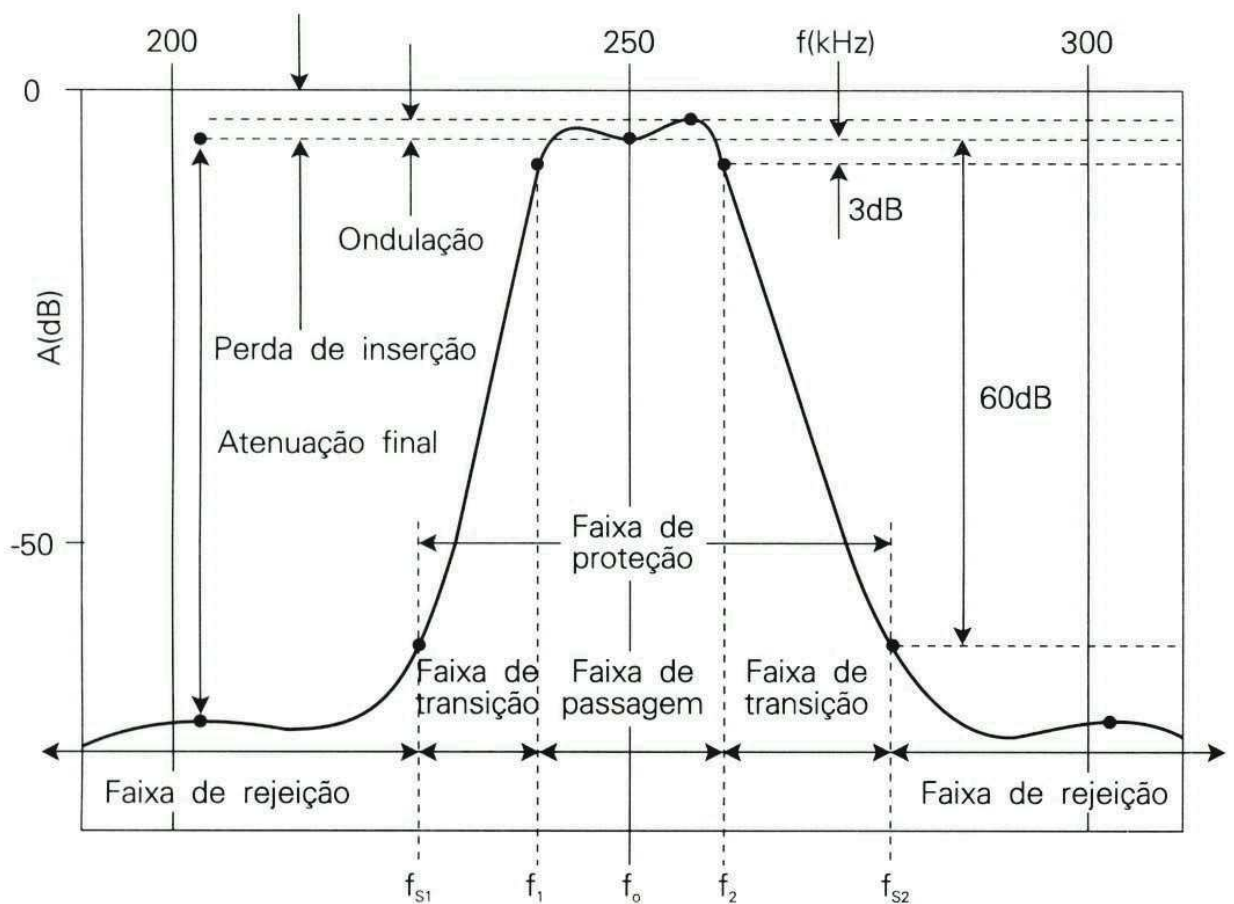


Figura 2.11 – Curva de Resposta em Frequência para um Filtro Passa Faixa (adaptado de [32]).

Em situações ideais e segundo MUSSOI [23], para sinais abaixo da frequência de corte inferior ou acima da frequência de corte superior, o ganho do filtro é nulo, ou seja, o módulo do sinal de saída é totalmente atenuado. Na prática, entretanto, não se obtêm respostas em frequência ideais.

2.3.6 Filtragem via Convolução

De acordo com HAYKIN [24], o processo de filtragem via convolução é realizado inteiramente no domínio do tempo e constitui o procedimento utilizado atualmente, pelo fato de a *Transformada de Fourier* ser altamente custosa, a partir do ponto de vista computacional. A convolução entre dois sinais discretos $x[\cdot]$ e $t[\cdot]$, que será a abordagem de filtragem digital utilizada daqui para frente, é dada por:

$$y[\cdot] = x[\cdot] * t[\cdot] = \sum_k t_k x_{n-k} \quad , \quad (2.5)$$

onde N é a quantidade de pontos.

2.3.7 Bandas Críticas, Escala Bark e Amostragem

A escala *Bark* está associada à largura das bandas críticas da audição humana. A largura de cada banda crítica difere de acordo com a faixa de frequências: abaixo de 500 *Hz* as bandas são constantes e com larguras iguais a 100 *Hz*. Acima de 500 *Hz* a largura da próxima banda crítica é 20% maior que a da anterior. Foi então criada uma unidade psico-acústica especial: o *Bark*. Um *Bark* (z) corresponde à largura de uma banda crítica. A faixa de frequência audível corresponde a, aproximadamente, 24 *Barks*, ou seja, é sub-dividida em 25 intervalos.

Segundo SOUZA [3], quando dois ou mais sons, de frequências diferentes mas pertencentes à mesma banda crítica, estão sendo captados simultaneamente, o de maior amplitude mascara o(s) outro(s). Portanto, o de menor amplitude não é percebido. Na Tabela 2.2, são apresentadas as bandas críticas, na escala de frequências em *Hertz* (*Hz*) e na escala *Bark* (z), que é uma escala adotada para linearizar o intervalo entre as bandas. A conversão para Barks é dada pela equação

$$z = \frac{26.81f}{1960 + f} - 0.53, \quad (2.6)$$

na qual f representa a frequência em *Hertz*.

Tabela 2.2 – Escala *Bark*

Escala Bark	Bandas Críticas
0	0 - 100
1	100 - 200
2	200 - 300
3	300 - 400
4	400 - 510
5	510 - 630
6	630 - 770
7	770 - 920
8	920 - 1080
9	1080 - 1270
10	1270 - 1480
11	1480 - 1720
12	1720 - 2000
13	2000 - 2320
14	2320 - 2700
15	2700 - 3150
16	3150 - 3700
17	3700 - 4400
18	4400 - 5300
19	5300 - 6400
20	6400 - 7700
21	7700 - 9500
22	9500 - 12000
23	12000 - 15500
24	15500 - 22050

A Tabela 2.2 apresentada contém as 25 bandas de energias, porém neste trabalho foram utilizadas apenas as primeiras 21 bandas por serem as que contêm as frequências disponíveis, de acordo com a taxa de amostragem dos sinais de voz utilizados. Esses foram digitalizados com taxa de 22050 amostras por segundo, o que significa que somente existem frequências até 11025 Hz, de acordo com o Teorema da Amostragem, como abordado em HAYKIN [24]. Além disso, os sinais foram quantizados com 16 bits, o que equivale a dizer que existem $2^{16} = 65536$ possibilidades distintas para representar cada valor de amplitude digitalizado.

2.3.8 Redes Neurais RBF (*Radial Basis Functions*)

De acordo com SILVA[25], as redes neurais artificiais (RNAs) são modelos computacionais inspirados no sistema nervoso dos seres vivos. Possuem a capacidade de aquisição e manutenção do conhecimento, baseada em informações, podendo ser definidas como conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligados por um grande número de interconexões. Constituem uma das mais antigas técnicas de inteligência artificial em uso. A primeira publicação relacionada à neurocomputação data de 1943, por meio de um artigo elaborado por *Walter Pitts* (matemático) e *McCulloch* (neurofisiologista).

As redes do tipo RBF surgiram como uma variante das redes neurais artificiais, no final dos anos 80. No entanto, suas raízes estão relacionadas a técnicas para reconhecimento de padrões muito mais antigas, tais como as funções potenciais, *clustering*, aproximação funcional, *splines* de interpolação, entre outras. São definidas como redes *feed-forward* de camadas múltiplas, cujo treinamento é efetivado de forma supervisionada e têm tipicamente três camadas: uma camada de entrada passiva, onde o fluxo de informação se inicia; percorrendo então uma camada oculta com uma função de ativação não-linear, e finalizando na camada neural de saída, contendo um ou mais neurônios com funções de ativação linear.

Em uma aplicação de classificação de padrões, os dados representam as entradas, enquanto que cada saída corresponde a uma classe, de acordo com BORS [26] e conforme exemplificado na Figura 2.12.

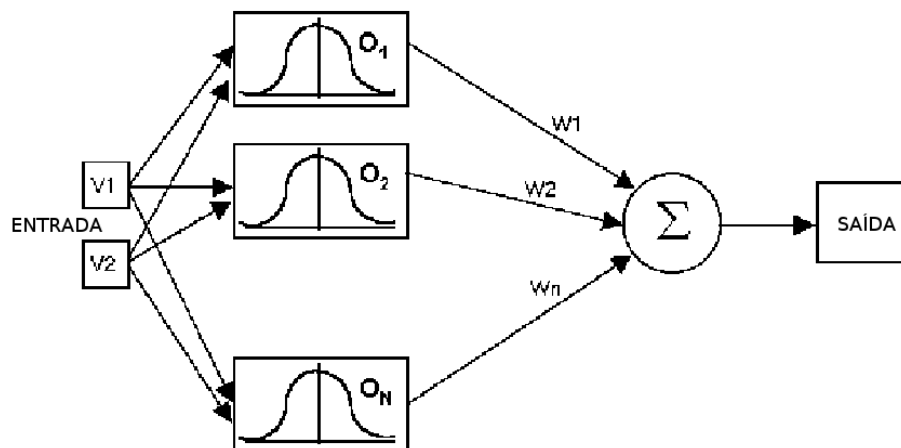


Figura 2.12 – A arquitetura de uma rede RBF, onde os W_n são os pesos.

Uma das principais particularidades da rede RBF e o motivo pelo qual este trabalho se baseia nela, está na estratégia de treinamento utilizada para os ajustes dos pesos de suas duas únicas camadas neurais. Um outro diferencial desta arquitetura de rede é o tipo de função de ativação assumida nos neurônios, da camada intermediária, sendo caracterizada por funções de base radial como as Gaussianas, como pode ser visto na Figura 2.13.

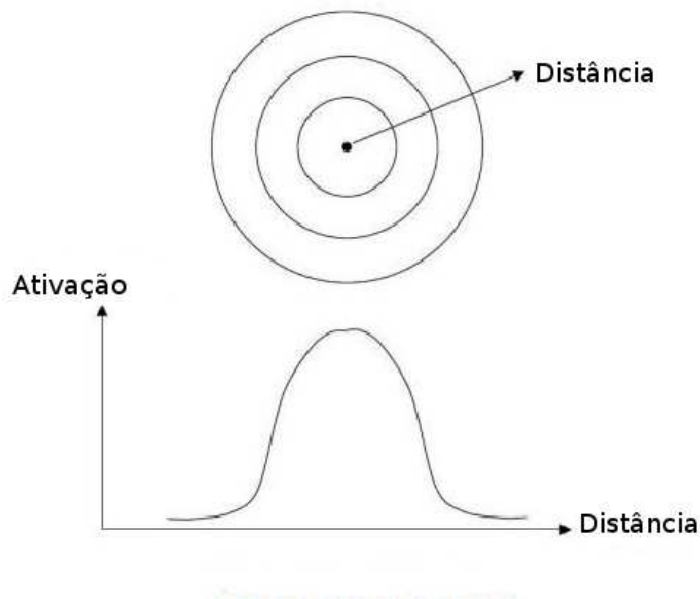


Figura 2.13 – A função de ativação com base radial.

Nas redes RBF, cada neurônio, ou nó, da saída produz um sinal que corresponde à soma ponderada das saídas da camada oculta. Por outro lado, a transformação realizada a partir da camada de entrada até a camada oculta é não-linear. O treinamento das redes RBF consiste em, no primeiro momento, ajustar, para cada caso de treinamento, as saídas da camada oculta, independentemente da classe à qual cada caso pertença. Em um segundo momento, as saídas da camada final são ajustadas conforme a classe a que o caso pertença. Assim sendo, a primeira etapa de treinamento é do tipo não-supervisionada e a segunda etapa é do tipo supervisionada.

Os procedimentos descritos acima implicam o fato de que um determinado problema, não linearmente separável, pode, de forma probabilística, ser transformado em um problema linearmente separável por meio de uma transformação não linear que mapeia o espaço original de

entrada para outro espaço de ordem maior. Particularmente, no presente trabalho, o espaço de entrada, de dimensão N , tem a sua dimensão elevada para R , sendo R o número total de casos de treinamento. Esse fato possibilita, normalmente, a obtenção dos pesos associados aos nós da camada final por meio da resolução de um sistema linear possível e determinado, facilitando o processo.

2.3.9 A Função Gaussiana

Podem ser usados diferentes tipos de *kernels*, isto é, funções matemáticas, como as funções não-lineares adotadas nas redes RBF, mas é mais comum o uso da função Gaussiana, que possui comportamento exibido na Figura 2.14.

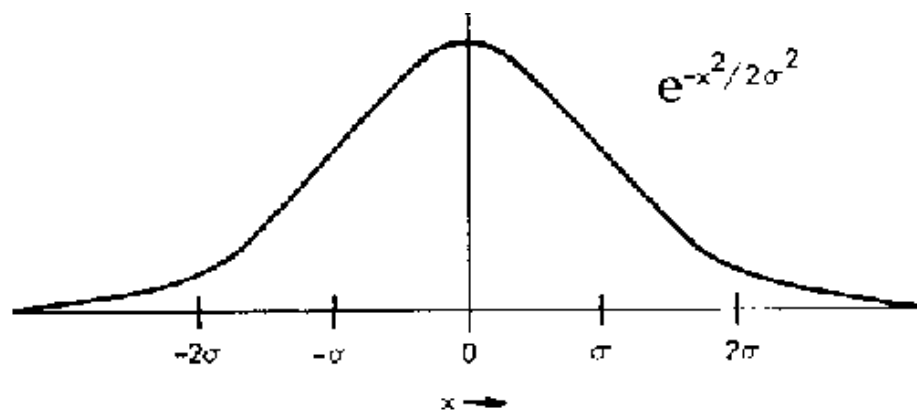


Figura 2.14 – Gráfico da função Gaussiana.

Nos trabalhos envolvendo detecção e classificação de patologias na laringe, citados anteriormente, o *kernel* Gaussiano tem se mostrado o mais propício. Esse fato fez com que ele fosse adotado no presente trabalho, enquanto que a quasi-variância, mencionada anteriormente, será utilizada para gerar uma modificação nesse *kernel*, criando uma função modificada, que será experimentada.

Capítulo 3

A Abordagem Proposta

Neste capítulo, será apresentado o sistema proposto, incluindo a sua arquitetura e os algoritmos associados para treinamento e teste, entre outros detalhes.

3.1 Estrutura do Sistema

Todo sistema de reconhecimento de padrões requer uma etapa anterior de extração de parâmetros, como pode ser visto na Figura 3.1, que visa reduzir a dimensão do sinal de entrada, a qual pode ser relativamente alta e variável, transformando-o em um conjunto reduzido e fixo de parâmetros, ou características, chamado de vetor de características. Esse vetor deve ser, posteriormente, entregue ao classificador que tem a função de determinar a qual classe o referido vetor, e conseqüentemente o sinal que o gerou, pertence. No presente trabalho, os sinais de voz, normais e patológicos, são convertidos em um vetor de características que, posteriormente, são entregues ao classificador.



Figura 3.1 – Esquema do processamento de um sinal de voz no presente trabalho.

Particularmente, o sistema proposto foi dividido em dois módulos distintos, cada qual baseado em um, classificador diferente, dentre três, isto é, Classificador I, Classificador II e Classificador III. No primeiro módulo foram desenvolvidos os detalhes relativos ao treinamento do sistema de classificação, enquanto que, no segundo, foram desenvolvidas as etapas relativas à aplicação, que se estabelecem baseado no sistema previamente treinado, sendo utilizado para classificar uma voz desconhecida como normal ou patológica.

3.2 Classificador I: Técnica Baseada na Distância Euclidiana

O treinamento e o teste do sistema proposto, com base no classificador por distância Euclidiana, ocorrem de acordo com os algoritmos 3.1 e 3.2 que seguem. Neste classificador I, o termo “treinamento” está sendo utilizado apenas por permitir uma analogia com os demais classificadores utilizados (II e III), já que o classificador não é treinado de fato, existindo apenas modelos de referência, isto é, *templates*.

Algoritmo 3.1 - Passos de Treinamento do Sistema Classificador I

- **INÍCIO**
- **PASSO T_1** : extrair os “dados brutos”, isto é, as amostras temporais digitalizadas, de todos os arquivos de voz utilizados no experimento, que estão armazenados no formato WAV, amostrados a 22050Hz, 16-bits;
- **PASSO T_2** : para cada sinal de voz, independentemente de ser oriundo de um indivíduo normal ou patológico, produzir o vetor de características representante dos dados brutos. Esse vetor deverá ser composto por $T = 42$ valores, assim dispostos:
 - **média** das energias dos J sinais resultantes das filtragens das J janelas do sinal sob análise, utilizando um filtro passa-faixas de ordem 1, com resposta ao impulso finita, projetado para a banda crítica **Bark 1**. Cada janela deverá possuir tamanho igual a 1024 amostras, sendo que a janela posterior se sobrepõe à anterior em 50%;
 - idem para a banda crítica **Bark 2**;
 - (. . .)
 - idem para a banda crítica **Bark 21**, que corresponde à maior banda possível no presente caso, tendo em vista a taxa de amostragem dos sinais de voz;
 - **variância** normalizada das energias dos J sinais resultantes das filtragens das J janelas do sinal sob análise, utilizando um filtro passa-faixas de ordem 1, com resposta ao impulso

finita, projetado para a banda crítica **Bark 1**. Cada janela deverá possuir tamanho igual a 1024 amostras, sendo que a janela posterior se sobrepõe à anterior em 50%;

- idem para a banda crítica **Bark 2**;
 - (. . .)
 - idem para a banda crítica **Bark 21**, que corresponde à maior banda possível no presente caso, tendo em vista a taxa de amostragem dos sinais de voz;
- **PASSO T_3** : separar os vetores de características correspondentes às vozes normais e às vozes patológicas;
 - **PASSO T_4** : separar o montante de vetores, correspondentes às vozes normais e patológicas, que devem ser utilizados para treinamento do sistema. Os demais, devem ser isolados para a fase de testes;

Algoritmo 3.2 - Passos de Teste / Uso do Sistema Classificador I

- **INÍCIO**
- **PASSO U_1** : disponibilizar para uso os vetores de características, oriundos de vozes normais e patológicas, que foram isolados para testes no passo T_4 do procedimento de treinamento previamente descrito;
- **PASSO U_2** : para cada vetor de testes, \vec{a}_i , com dimensão 42, medir a distância Euclidiana entre ele e todos os vetores de treinamento, \vec{b}_j .
- **PASSO U_3** : toda vez que um vetor \vec{a}_i for comparado com todos os \vec{b}_j , deverá ser armazenado o menor valor, que corresponde ao \vec{a}_i mais próximo, ou similar, com os \vec{b}_j utilizados. Se o vetor \vec{b}_j que gerou o menor valor for normal, então considera-se \vec{a}_i como normal, caso contrário, como patológico.

3.3 Classificador II: Rede Neural RBF e *Kernel* Gaussiano

O treinamento e o teste do sistema classificador com base na rede RBF e no *kernel* Gaussiano ocorrem de acordo com os algoritmos 3.3 e 3.4 que seguem.

Algoritmo 3.3 - Passos de Treinamento do Sistema Classificador II

- **INÍCIO**
- **PASSO T_1** : extrair os “dados brutos”, isto é, as amostras temporais digitalizadas, de todos os arquivos de voz utilizados no experimento, que estão armazenados no formato WAV, amostrados a 22050Hz, 16-bits;
- **PASSO T_2** : para cada sinal de voz, independentemente de ser oriundo de um indivíduo normal ou patológico, produzir o vetor de características representante dos dados brutos. Esse vetor deverá ser composto por $T = 42$ valores, assim dispostos:

- **média** das energias dos J sinais resultantes das filtragens das J janelas do sinal sob análise, utilizando um filtro passa-faixas de ordem 1, com resposta ao impulso finita, projetado para a banda crítica **Bark 1**. Cada janela deverá possuir tamanho igual a 1024 amostras, sendo que a janela posterior se sobrepõe à anterior em 50%;
 - idem para a banda crítica **Bark 2**;
 - (. . .)
 - idem para a banda crítica **Bark 21**, que corresponde à maior banda possível no presente caso, tendo em vista a taxa de amostragem dos sinais de voz;
 - **variância** normalizada das energias dos J sinais resultantes das filtragens das J janelas do sinal sob análise, utilizando um filtro passa-faixas de ordem 1, com resposta ao impulso finita, projetado para a banda crítica **Bark 1**. Cada janela deverá possuir tamanho igual a 1024 amostras, sendo que a janela posterior se sobrepõe à anterior em 50%;
 - idem para a banda crítica **Bark 2**;
 - (. . .)
 - idem para a banda crítica **Bark 21**, que corresponde à maior banda possível no presente caso, tendo em vista a taxa de amostragem dos sinais de voz;
- **PASSO T_3** : separar os vetores de características correspondentes às vozes normais e às vozes patológicas;
 - **PASSO T_4** : separar o montante de vetores, correspondentes às vozes normais e patológicas, que devem ser utilizados para treinamento do sistema. Os demais, devem ser isolados para a fase de testes;
 - **PASSO T_5** : definir a rede RBF a ser utilizada, que deve possuir $T = 42$ entradas, R neurônios na camada oculta, sendo R o número de casos de treinamento, e 1 neurônio na camada final;
 - **PASSO T_6** : utilizando *kernels* Gaussianos, ajustar as saídas dos R neurônios da camada intermediária de modo que o i -ésimo neurônio apresente saída máxima, igual a 1, para o i -ésimo caso de treinamento. Para os demais exemplos, o i -ésimo neurônio apresentará um valor de saída qualquer entre 0 e 1. Para isso, basta ajustar o centro da função Gaussiana respectiva, isto é, μ_i , para o ponto que é idêntico ao vetor de características em questão. Isso fará com que a função exponencial natural do *kernel* resulte no valor 1, que é o seu máximo. Particularmente, a Gaussiana utilizada corresponde à exponencial natural do negativo da distância Euclidiana entre o vetor de entrada e o vetor média μ_i ;
 - **PASSO T_7** : estabelecer um sistema linear, possível e determinado, que apresente R equações e R incógnitas. Cada equação forma-se com base nos valores de saída de cada neurônio da camada oculta, mediante exercitação de cada caso de treinamento. As variáveis correspondem aos pesos existentes entre a camada oculta e a camada final, assim como o uso dos valores 1 ou -1 para igualar as equações, representando, respectivamente, voz normal ou voz patológica;
 - **PASSO T_8** : resolver, por triangularização ou outro método qualquer, o sistema especificado, encontrando os valores dos pesos.

Algoritmo 3.4 - Passos de Teste / Uso do Sistema Classificador II

- **INÍCIO**

- **PASSO U_1 :** disponibilizar para uso os vetores de características, oriundos de vozes normais e patológicas, que foram isolados para testes no passo T_4 do procedimento de treinamento previamente descrito;
- **PASSO U_2 :** para cada vetor de testes, aplicar as características correspondentes na entrada da rede RBF previamente treinada. Coletar os valores presentes nas saídas dos neurônios da camada oculta, fazendo uma combinação linear deles com os pesos determinados na etapa de treinamento.
- **PASSO U_3 :** O valor resultante da combinação linear anterior é interpretado para fornecer a resposta do sistema ao vetor de entrada. Se esse valor estiver mais próximo de 1 do que de -1 , a voz correspondente será considerada normal, caso contrário será considerada patológica.

3.4 Classificador III: Rede Neural RBF e *Kernel* Gaussiano Modificado

O treinamento e o teste deste sistema classificador III são os mesmos do classificador II, anteriormente detalhados, mas trocando a função *kernel*. Particularmente, a variância utilizada comumente no *kernel* Gaussiano foi substituída pela quasi-variância, objetivando a aplicação de um classificador não tendencioso. A função *kernel* modificada é, portanto, a seguinte:

$$\begin{aligned}
& e^{-\left(\frac{\left(\frac{(x-\sum x_i)}{N}\right)^2}{\frac{2\sigma^2 N}{N-1}}\right)} \\
&= e^{-\left((x-\sum x_i)^2 \cdot \frac{N-1}{2\sigma^2 N}\right)} \\
&= e^{-\left(\left(x^2 - \frac{2x \sum x_i}{N} + \left(\frac{\sum x_i}{N}\right)^2\right) \cdot \left(\frac{N}{2\sigma^2 N} - \frac{1}{2\sigma^2 N}\right)\right)} \\
&= e^{-\left(\frac{x^2 N}{2\sigma^2 N} - \frac{x^2}{2\sigma^2 N} - \frac{2x \sum x_i}{2\sigma^2 N} + \frac{2x \sum x_i}{2\sigma^2 N^2} + \frac{(\sum x_i)^2}{2\sigma^2 N^2} - \frac{(\sum x_i)^2}{N 2\sigma^2 N^2}\right)} \\
&= e^{-\left(\frac{x^2}{2\sigma^2} - \frac{x^2 N^{-1}}{2\sigma^2} - \frac{2x \sum x_i}{2\sigma^2 N} + \frac{2x \sum x_i N^{-1}}{2\sigma^2 N} + \frac{(\sum x_i)^2}{2\sigma^2 N^2} - \frac{(\sum x_i)^2 N^{-1}}{2\sigma^2 N^2}\right)} \\
&= e^{-\left(\frac{x^2 - x^2 N^{-1}}{2\sigma^2} - \frac{2x \sum x_i + 2x \sum x_i N^{-1}}{2\sigma^2 N} + \frac{(\sum x_i)^2 - (\sum x_i)^2 N^{-1}}{2\sigma^2 N^2}\right)} \\
&= e^{-\left(\frac{(1-N^{-1})x^2 N^2}{2\sigma^2 N} - \frac{N 2x \sum x_i (1+N^{-1})}{2\sigma^2 N^2} + \frac{(\sum x_i)^2 (1-N^{-1})}{2\sigma^2 N^2}\right)}
\end{aligned}$$

$$\begin{aligned}
&= e^{-\left(\frac{N^{-2}(1-N^{-1}) \cdot (x^2 N^2 + (\sum x_i)^2) - N^{-2} N 2x \sum x_i (1+N^{-1})}{2\sigma^2}\right)} \\
&= e^{-\left(\frac{(N^{-2}-N^{-3}) \cdot (x^2 N^2 + (\sum x_i)^2) - N^{-1} 2x \sum x_i (1+N^{-1})}{2\sigma^2}\right)}
\end{aligned} \tag{3.1}$$

Portanto, o *kernel* parcialmente modificado é dado por:

$$e^{-\left(\frac{t}{2\sigma^2}\right)} \tag{3.2}$$

sendo,

$$t = -(N^{-2} - N^{-3})(x^2 N^2 + (\sum x_i)^2) - N^{-1} 2x \sum x_i (1 + N^{-1}) \tag{3.3}$$

A partir de t , tem-se:

$$\begin{aligned}
t &= (N^{-2} - N^{-3})(x^2 N^2 + (\mu N)^2) - N^{-1} 2x \mu N (1 + N^{-1}) \\
&= (N^{-2} - N^{-3})(x^2 N^2 + \mu^2 N^2) - N^{-1} 2x \mu N (1 + N^{-1}) \\
&= (N^{-2} - N^{-3})(N^2(x^2 + \mu^2)) - N^{-1} 2x \mu N - N^{-1} N^{-1} 2x \mu N \\
&= (N^{-2} - N^{-3})(N^2(x^2 + \mu^2)) - 2x \mu - N^{-1} 2x \mu \\
&= N^{-2} N^2(x^2 + \mu^2) - N^{-3} N^2(x^2 + \mu^2) - 2x \mu - N^{-1} 2x \mu \\
&= (x^2 + \mu^2) - N^{-1}(x^2 + \mu^2) - 2x \mu - N^{-1} 2x \mu \\
&= (x^2 + \mu^2)(1 - N^{-1}) - 2x \mu (1 - N^{-1}) \mu \\
&= (1 - N^{-1})(x^2 + \mu^2) - 2x \mu \\
&= (1 - N^{-1})(x^2 + \mu^2 - 2x \mu) \\
&= (1 - N^{-1})(x - \mu)^2 \\
&= (x - \mu)^2 - \frac{(x - \mu)^2}{N}
\end{aligned} \tag{3.4}$$

Substituindo (3.4) em (3.2), tem-se:

$$e^{-\frac{\left((x-\mu)^2 - \frac{(x-\mu)^2}{N}\right)}{2\sigma^2}} \tag{3.5}$$

As figuras 3.2 e 3.3 servem de apoio, respectivamente, para a compreensão do procedimento

de treinamento e do procedimento de testes, detalhados acima, para todos os classificadores (I, II ou III).

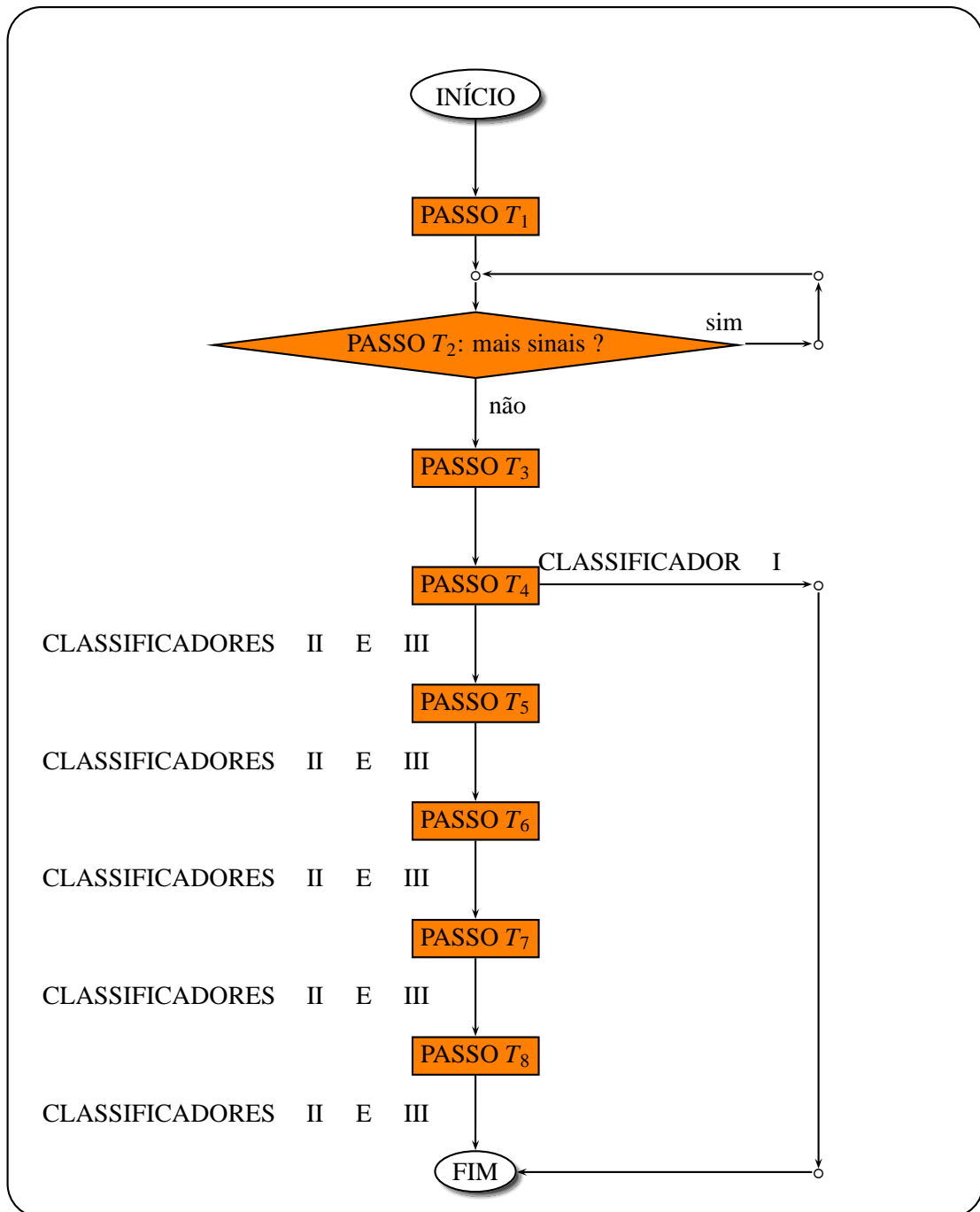


Figura 3.2 – Arquitetura básica para treinamento do sistema de detecção e classificação de patologias do trato vocal.

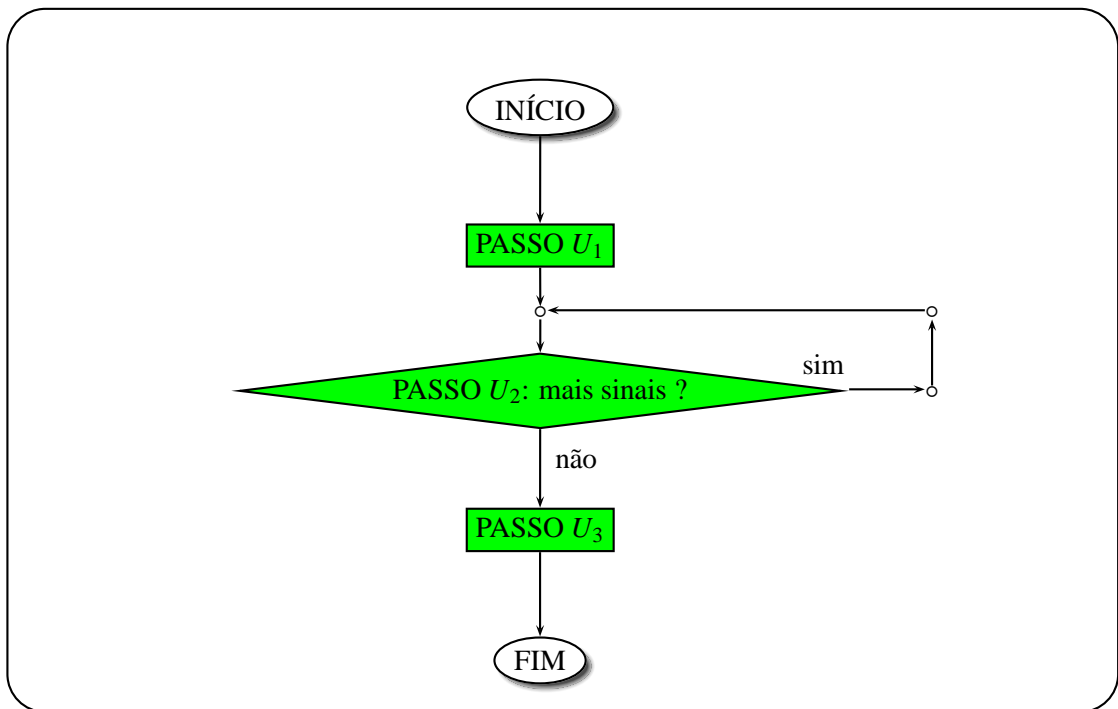


Figura 3.3 – Arquitetura para testes e uso do sistema de detecção de patologias.

3.5 Implementação

Em ambas as etapas, de treinamento e de testes / aplicação, os algoritmos foram implementados em linguagem C/C++, sob ambiente LINUX Conectiva 7.0 e processador *Intel Core i5*, em uma máquina com 4 Gb de memória RAM. O compilador utilizado foi o GNU g++, associado ao editor de textos *Kwrite*.

O próximo capítulo apresenta detalhes referentes aos testes e aos resultados obtidos com o sistema projetado e implementado.

Capítulo 4

Testes e Resultados

Neste capítulo são apresentados os testes realizados, que foram divididos em duas partes, assim como os seus respectivos resultados obtidos.

4.1 Métodos e Materiais Utilizados

Para o desenvolvimento deste projeto, todos os sinais de vozes utilizados correspondem a diversas combinações de parâmetros extraídos dos fonemas vozeados /a/, tais como na palavra *amor*, sustentados em média por cinco segundos, que foram extraídas do banco de vozes do Hospital das Clínicas da USP, de Ribeirão Preto. Essa base de dados possui 118 vozes com características normais e outras 33 pertencentes a indivíduos com as seguintes patologias na laringe: nódulo/cistos/calos nas pregas vocais, compondo um tipo de patologia analisado, e edemas de Reinke, compondo outro tipo de patologia procurado. Todos os indivíduos foram previamente examinados por profissionais da área médica para confirmar seu estado saudável ou patológico. A base está separada por pastas nomeadas com o nome de cada paciente examinado. Em cada uma das pastas existem arquivos com a extensão .WAV, amostrados a uma taxa de 22050Hz, 16-bits.

4.2 Procedimentos de Testes e Respostivos Resultados com o Uso do Classificador I - Distâncias Euclidianas

Segue a bateria dos testes realizados, assim como os resultados obtidos, que comparadas com as demais levam às conclusões do presente trabalho. Particularmente, os testes foram realizados de forma a experimentar:

- 10% da base para treinamento e 90% restantes para testes;
- 20% da base para treinamento e 80% restantes para testes;
- (. . .)
- 90% da base para treinamento e 10% restantes para testes.

Em cada caso acima, 100 combinações aleatórias foram utilizadas para separar o montante de sinais destinado para o treinamento e o montante destinado para testes, sendo os conjuntos

mutuamente excludentes. Os resultados apresentados a seguir, expressos na forma de matrizes de confusão, correspondem às médias de cada um dos 100 testes.

- **GRUPO 1: utilizando 10% da base para treinamento e 90% para testes:**

		Patológica	Normal
Médias:	Patológica	16	90
	Normal	9	20

- **GRUPO 2: utilizando 20% da base para treinamento e 80% para testes:**

		Patológica	Normal
Médias:	Patológica	15	79
	Normal	9	18

- **GRUPO 3: utilizando 30% da base para treinamento e 70% para testes:**

		Patológica	Normal
Médias:	Patológica	18	65
	Normal	12	13

- **GRUPO 4: utilizando 40% da base para treinamento e 60% para testes:**

		Patológica	Normal
Médias:	Patológica	20	51
	Normal	10	13

- **GRUPO 5: utilizando 50% da base para treinamento e 50% para testes:**

		Patológica	Normal
Médias:	Patológica	17	42
	Normal	10	7

- **GRUPO 6: utilizando 60% da base para treinamento e 40% para testes:**

		Patológica	Normal
Médias:	Patológica	5	42
	Normal	7	6

- **GRUPO 7: utilizando 70% da base para treinamento e 30% para testes:**

		Patológica	Normal
Médias:	Patológica	3	32
	Normal	6	4

- **GRUPO 8: utilizando 80% da base para treinamento e 20% para testes:**

		Patológica	Normal
Médias:	Patológica	3	21
	Normal	2	3

- **GRUPO 9: utilizando 90% da base para treinamento e 10% para testes:**

		Patológica	Normal
Médias:	Patológica	2	10
	Normal	2	2

Observando os resultados anteriores, é possível constatar que o número de vozes classificadas corretamente como sendo normais é bastante relevante, mesmo para treinamentos com um número reduzido de vozes. Nos casos em que vozes normais são apontadas como patológicas, o risco oferecido pelo sistema de pré-diagnóstico não é considerável, tendo em vista que um exame médico complementar é capaz de dirimir possíveis dúvidas ou apontar erros nas classificações. Um ponto negativo que pode ser observado nos resultados é que há um número considerável de vozes patológicas classificadas como sendo normais. Tais erros são mais graves e constituem uma desvantagem do presente Classificador I.

4.3 Procedimentos de Testes e Respectivos Resultados com o Uso do Classificador II - Rede Neural RBF e *Kernel* Gaussiano

Segue a bateria dos testes realizados, assim como os resultados obtidos, que comparados com os demais levam às conclusões do presente trabalho. Particularmente, os testes foram realizados da mesma forma que no caso anterior, isto é, foram experimentados:

- 10% da base para treinamento e 90% restantes para testes;
- 20% da base para treinamento e 80% restantes para testes;
- (. . .)
- 90% da base para treinamento e 10% restantes para testes.

Em cada caso acima, 50 combinações aleatórias foram utilizadas para separar o montante de sinais destinado para o treinamento e o montante destinado para testes, sendo os conjuntos mutuamente excludentes. Os resultados apresentados a seguir são expressos na forma de matrizes de confusão, apontando-se o melhor e o pior caso.

- **utilizando 10% da base para treinamento e 90% para testes:**

	Patológica	Normal
Melhor (Combinação 6 - 108 acertos):	Patológica $\left(\begin{array}{cc} 13 & 16 \\ 11 & 95 \end{array} \right)$	Normal

	Patológica	Normal
Pior (Combinação 46 - 65 acertos):	Patológica $\left(\begin{array}{cc} 5 & 24 \\ 46 & 60 \end{array} \right)$	Normal

- **utilizando 20% da base para treinamento e 80% para testes:**

	Patológica	Normal
Melhor (Combinação 19 - 93 acertos):	Patológica $\left(\begin{array}{cc} 8 & 19 \\ 9 & 85 \end{array} \right)$	Normal

		Patológica	Normal
Pior (Combinação 42 - 69 acertos):	Patológica	(13	14)
	Normal	(38	56)

- **utilizando 30% da base para treinamento e 70% para testes:**

		Patológica	Normal
Melhor (Combinação 8 - 80 acertos):	Patológica	(15	8)
	Normal	(18	65)

		Patológica	Normal
Pior (Combinação 6 - 56 acertos):	Patológica	(14	9)
	Normal	(41	42)

- **utilizando 40% da base para treinamento e 60% para testes:**

		Patológica	Normal
Melhor (Combinação 19 - 77 acertos):	Patológica	(6	14)
	Normal	(30	71)

		Patológica	Normal
Pior (Combinação 2 - 49 acertos):	Patológica	(6	14)
	Normal	(28	43)

- **utilizando 50% da base para treinamento e 50% para testes:**

		Patológica	Normal
Melhor (Combinação 16 - 57 acertos):	Patológica	(10	7)
	Normal	(12	47)

		Patológica	Normal
Pior (Combinação 24 - 37 acertos):	Patológica	(9	8)
	Normal	(31	28)

- **utilizando 60% da base para treinamento e 40% para testes:**

		Patológica	Normal
Melhor (Combinação 42 - 46 acertos):	Patológica	(4	9)
	Normal	(5	42)

		Patológica	Normal
Pior (Combinação 14 - 23 acertos):	Patológica	(4	9)
	Normal	(28	19)

- **utilizando 70% da base para treinamento e 30% para testes:**

		Patológica	Normal
Melhor (Combinação 27 - 33 acertos):	Patológica	(4	6)
	Normal	(6	29)

		Patológica	Normal
Pior (Combinação 7 - 22 acertos):	Patológica	(5	5)
	Normal	(18	17)

- **utilizando 80% da base para treinamento e 20% para testes:**

		Patológica	Normal
Melhor (Combinação 23 - 23 acertos):	Patológica	(3	3)
	Normal	(4	20)

		Patológica	Normal
Pior (Combinação 48 - 12 acertos):	Patológica	(1	5)
	Normal	(13	11)

- **utilizando 90% da base para treinamento e 10% para testes:**

		Patológica	Normal
Melhor (Combinação 48 - 14 acertos):	Patológica	(3	1)
	Normal	(1	11)

		Patológica	Normal
Pior (Combinação 50 - 4 acertos):	Patológica	(1	3)
	Normal	(9	3)

Observando agora os resultados anteriores, é possível constatar ainda que, como no caso anterior, o número de vezes classificadas corretamente como sendo normais é bastante relevante, mesmo para treinamentos com um número reduzido de vezes. Embora de forma mais modesta, observa-se ainda um ponto negativo: há um número considerável de vezes patológicas classificadas como sendo normais.

4.4 Procedimentos de Testes e Respectivos Resultados com o Uso do Classificador III - Rede Neural RBF e *Kernel* Gaussiano Modificado

Segue a bateria dos testes realizados para o Classificador II, considerando os resultados obtidos por meio de um *kernel* modificado para os neurônios da camada oculta da rede RBF. Novamente, os testes foram realizados de forma a experimentar:

- 10% da base para treinamento e 90% restantes para testes;
- 20% da base para treinamento e 80% restantes para testes;
- (. . .)
- 90% da base para treinamento e 10% restantes para testes.

Em cada caso acima, novamente, 50 combinações aleatórias foram utilizadas para separar o montante de sinais destinado para o treinamento e o montante destinado para testes, sendo os conjuntos mutuamente excludentes. Os resultados apresentados a seguir são expressos na forma de matrizes de confusão de cada um dos testes, mais uma vez, apontando o melhor e o pior resultado em cada caso.

- **utilizando 10% da base para treinamento e 90% para testes:**

		Patológica	Normal
Melhor (Combinação 10 - 107 acertos):	Patológica	8	21
	Normal	7	99

		Patológica	Normal
Pior (Combinação 34 - 60 acertos):	Patológica	15	14
	Normal	61	45

- **utilizando 20% da base para treinamento e 80% para testes:**

		Patológica	Normal
Melhor (Combinação 19 - 95 acertos):	Patológica	10	17
	Normal	9	85

		Patológica	Normal
Pior (Combinação 25 - 66 acertos):	Patológica	13	14
	Normal	41	53

- **utilizando 30% da base para treinamento e 70% para testes:**

		Patológica	Normal
Melhor (Combinação 3 - 78 acertos):	Patológica	7	16
	Normal	12	71

	Patológica	Normal
Pior (Combinação 24 - 55 acertos):	Patológica (9	4)
	Normal (27	56)

- **utilizando 40% da base para treinamento e 60% para testes:**

	Patológica	Normal
Melhor (Combinação 37 - 70 acertos):	Patológica (10	10)
	Normal (11	60)

	Patológica	Normal
Pior (Combinação 19- 45 acertos):	Patológica (6	14)
	Normal (32	39)

- **utilizando 50% da base para treinamento e 50% para testes:**

	Patológica	Normal
Melhor (Combinação 9 - 57 acertos):	Patológica (12	5)
	Normal (14	45)

	Patológica	Normal
Pior (Combinação 24 - 37 acertos):	Patológica (11	6)
	Normal (33	26)

- **utilizando 60% da base para treinamento e 40% para testes:**

	Patológica	Normal
Melhor (Combinação 42 - 46 acertos):	Patológica (4	9)
	Normal (5	42)

	Patológica	Normal
Pior (Combinação 14 - 23 acertos):	Patológica (4	9)
	Normal (28	19)

- **utilizando 70% da base para treinamento e 30% para testes:**

	Patológica	Normal
Melhor (Combinação 49 - 33 acertos):	Patológica (5	5)
	Normal (7	29)

		Patológica	Normal
Pior (Combinação 2 - 21 acertos):	Patológica	5	5
	Normal	19	16

- **utilizando 80% da base para treinamento e 20% para testes:**

		Patológica	Normal
Melhor (Combinação 6 - 22 acertos):	Patológica	4	2
	Normal	6	18

		Patológica	Normal
Pior (Combinação 4 - 12 acertos):	Patológica	1	5
	Normal	13	11

- **utilizando 90% da base para treinamento e 10% para testes:**

		Patológica	Normal
Melhor (Combinação 28 - 14 acertos):	Patológica	3	1
	Normal	1	11

		Patológica	Normal
Pior (Combinação 50 - 4 acertos):	Patológica	1	3
	Normal	9	3

Uma observação muito importante é que, para todos os três classificadores testados, I, II e III, foram considerados como melhor e pior casos, os que produzem, respectivamente, o maior e o menor número **total** de acertos. Assim, não se percebe vantagem quando o classificador III, por exemplo, é comparado com o classificador II, isto é, o número **total** de acertos é similar. Entretanto, quando se comparam os falsos negativos, isto é, vozes patológicas reconhecidas erroneamente como sendo normais, percebe-se vantagem no uso do classificador III em comparação com o classificador II, principalmente quando mais de 50 % da base é utilizada para treinamento. Tal fato é relevante, tendo em vista que os falsos negativos constituem a maior preocupação dos clínicos.

Capítulo 5

Conclusões e Trabalhos Futuros

Feita a revisão bibliográfica de importantes conceitos, o presente trabalho apresenta um sistema de detecção de patologias na laringe, não-invasivo, visando servir como um método de pré-diagnóstico. O sistema foi projetado com base em um módulo de extração de parâmetros, seguido por um classificador. Este foi implementado com base em três abordagens, que foram implementadas, testadas e comparadas: distâncias Euclidianas, rede RBF com *kernels* Gaussianos e redes RBFs como *Kernels* Gaussianos modificados.

Principalmente no que se refere a detecção de casos falsos negativos, o uso da classificador III mostrou-se mais vantajoso do que os classificadores I e II e, assim, confirma-se que o emprego da quasi-variância foi eficaz, de acordo com os testes e resultados descritos no capítulo anterior.

Os resultados aqui apresentados são comparáveis aos resultados obtidos por outros autores. Exemplos como em CHILDERS [6], em que os autores separaram normais de patológicas com um índice de acerto de 75,9%, usando rede neural treinada com o algoritmo LPC (*Linear Predictive Coding*), 29 vozes patológicas e 52 vozes normais. Outro exemplo encontra-se em PARAGA [15], que abordou, para a classificação automática de vozes em patológicas ou normais, a aplicação da *Transformada Wavelet Packet* e do algoritmo *Best Basis*, por meio da qual obteve um índice de classificação de 85,42% para as patológicas (7 pacientes falsos negativos para 48 amostras) e 76,92% normais (3 pacientes falsos positivos para 13 amostras). Um outro exemplo também é SOUZA [3] que usou uma estrutura composta por duas Redes SVMs (*Support Vector Machines*) treinadas com um procedimento semi-supervisionado, recebendo parâmetros obtidos com base na DWT (*Discrete Wavelet Transform*) do sinal de voz do paciente. É possível notar que o uso de 35% a 40% da base de vozes para treinar o sistema já foi suficiente para a obtenção de um alto índice de acertos nas classificações: 92% de acertos para as normais e 94,11% de acertos para as patológicas. Alguns resultados foram superiores, como 96,1%, em ACCARDO [33]. Porém, este resultado foi alcançado com condições restritas, pois investiga a existência de uma patologia específica. Já em ROSA [34], foram utilizados índices acústicos e uma rede neural artificial para discriminar entre vozes patológicas e normais, também com um índice de acertos similar ao do presente trabalho.

Para trabalhos futuros, pode-se tentar trabalhar com uma base de dados maior, o que pode ser não ser tão simples, pois para levantar essas informações depende-se de profissionais da área médica, de hospitais. Uma outra opção seria utilizar outro tipo de classificador inteligente, a própria lógica fuzzy é uma alternativa.

Referências

- [1] GUIDO, R. C. ; PEREIRA, J ; SLAETS, J . Introduction to the special issue on Emergent Applications of Fractals and Wavelets in Biology and Biomedicine. *Applied Mathematics and Computation*, v. 207, p. 3-4, 2009.
- [2] BARBON JR, S. *Dynamic time warping baseado na transformada wavelet*. 2007. 113p. Dissertação (Mestrado). Instituto de Física de São Carlos da Universidade de São Paulo, São Carlos, 2007.
- [3] SOUZA, L.M. Detecção inteligente de patologias na laringe baseada em máquinas de vetores de suporte e na transformada wavelet. 2010. 102p. Dissertação (Mestrado em Ciências). Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.
- [4] SCALASSARA, P. R. Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias de Laringe. 2009. 267 f. Tese (Doutorado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.
- [5] BAKEN R. J. *Clinical Measurement of Speech and Voice*. Allyn and Bacon, a division of Simon and Schuster, Inc., Needham Heights, MA. 1987.
- [6] CHILDERS D. G., Bae K. S. Detection of Laryngeal Function Using Speech and Electroglottographic Data, *IEEE Transactions on Biomedical Engineering*, Vol. 39, Issue 1, Pages 19-25, Janeiro 1992.
- [7] FROHLICH M., Michaelis D., Werner S. H. Acoustic breathiness measures in the description of pathologic voices, *Proceedings of the 1998 IEEE International*.
- [8] DAJER, M. E. Análise de sinais de voz por padrões visuais de dinâmica vocal. 2010. pp. 154. Tese (Doutorado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2010.
- [9] RITCHINGS RT., McGillion M., Moore CJ. Pathological voice quality assesment using artificial neural networks, *proceedings of 2nd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze-Italy, Setembro, 2001.
- [10] JO C., Kim K., Kim D., Wang S. Screening of Pathological Voice from ARS using Neural Networks, *proceedings of 2nd International Workshop on Models and Analysis of vocal emissions for biomedical applications*, Firenze-Italy, Setembro, 2001.
- [11] MARTINEZ C. E., Rufiner H. L. *Acoustic Analysis of speech for detection of Laryngeal Pathologies*, *Information Technology Applications in Biomedicine*. *Proceedings*. 2000 IEEE EMBS International Conference on, 2000.

- [12] KENT R. D. and Read C. *The Acoustics Analysis of Speech*. Singular Publishing Group, Inc., San Diego, CA, 1992.
<http://www.webciencia.com>
- [13] Esquema da laringe. Disponível em: <http://www.webciencia.com>. Acesso em: Janeiro 2013.
- [14] Localização da laringe. Disponível em: <http://www.mediscan.com.br/medicina-e-saude/cancer-de-laringe/>. Acesso em: Janeiro 2013.
- [15] PARRAGA, A. *Aplicação da Transformada Wavelet-Packet na Análise e Classificação de Sinais de Vozes Patológicas*. 2002. 163p. Dissertação (Mestrado). Departamento de Engenharia Elétrica, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2002.
- [16] DELLER Jr., P. R.; Proakis, J. G. and Hansen, J. H. L. *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, New York, NY, USA. 1993.
- [17] Cistos. Disponível em: <http://www.fonoclin.com.br/artigo/polipos/vocais.php>. Acesso em: Abril 2012.
- [18] Edemas (inchaços) de Reinke. Disponível em: <http://vivianverissimocantar.blogspot.com.br/2009/06/01-archive.html>. Acesso em: Abril 2012.
- [19] Edemas (inchaços) de Reinke. Disponível em: <http://www.capeotorrino.com.br/doencas/vocais.html>. Acesso em: Abril 2012.
- [20] Tabela. Disponível em: http://www.rc.unesp.br/pef/2003_projetos/Pedro/trabalho.html. Acesso em: Abril 2012.
- [21] COLTON R. H., Casper J. K., *Compreendendo os problemas de voz*. Artes Médicas, Porto Alegre, 1996.
- [22] FONSECA, E. S.; GUIDO, R. C.; SCALASSARA, P. R.; MACIEL, C.D.; PEREIRA, J. C.; *Speech processing: Wavelet Time-frequency Analysis and Least Squares Support Vector Machines for the Identification of Voice Disorders*. EE/UCLA - School of Engineering and Applied Sciences, University of California at Los Angeles, CA, USA, 2006.
- [23] MUSSOI, F. L. R.; *Resposta em Frequência: filtros passivos*. 2.ed. Florianópolis, Julho 2004.
- [24] HAYKIN, S.; VEEN, B. V. *Sinais e sistemas*. Porto Alegre: Bookman, 2002.
- [25] SILVA I. N. DA.; SPATTI, D. H. ; FLAUZINO, R. A. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. São Paulo, Artliber, 2010.
- [26] BORS, A. G., GABBOUJ, G., “Minimal topology for a radial basis function neural network for pattern classification”, *Digital Signal Processing: a review journal*, vol. 4, no. 3, pp. 173-188.(1994)
- [27] Cavidade Nasal. Disponível em: <http://trabalhodecienciascsa.blogspot.com.br>. Acesso em: Dezembro 2012.

- [28] Cavidade Nasal. Disponível em: <http://www.portalsaofrancisco.com.br/alfa/respiracao-cutanea/funcao-do-sistema-respiratorio.php>. Acesso em: Dezembro 2012.
- [29] CUNHA COSTA, S.; COSTA, W.; AGUIAR NETO, B.; Análise Dinâmica Não-Linear de Sinais de Vozes Afetados por Patologias na Laringe. Disponível em: <http://connepi.ifal.edu.br/ocs/index.php/connepi/CONNepi2010/paper/view/1827/463>, 2010.
- [30] Patologias. Disponível em: <http://www.medicinanet.com.br/conteudos/revisoes/1389/rouquidao.htm>. Acesso em: Dezembro 2012.
- [31] Câncer. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/laringe>. Acesso em: Dezembro 2012.
- [32] Filtros. Disponível em: <http://blogdopicco.blogspot.com.br/2010/09/filtros-parte-01.html>. Acesso em: Dezembro 2012.
- [33] ACCARDO A. P.; ENZO M. *An algorithm for the automatic differentiation between the speech of normals and patients with Friedreich's ataxia based on the short-time fractal dimension* . Computers in Biology and Medicine, Vol. 28, Issue 1, Elsevier Pages 75-89, January 1998.
- [34] ROSA M. de O.; PEREIRA C. J. and CARVALHO.P.L.F.A. *Evaluation of Neural Classifiers using Statistic Methods for Identification of Laryngeal Pathologies*. Proceedings of the Vth Brazilian Symposium on Neural Networks, 1998.
- [35] Matriz de Confusão. Disponível em: <http://wikipedia.qwika.com/en2pt/Confusion matrix>. Acesso em: Janeiro 2014.

Apêndice I - Código fonte do sistema proposto.

```

#include<iostream>
#include<stdio.h>
#include<math.h>
#include<string.h>
using namespace std;
//-----
main()
{
double distancia_euclidiana(double*, double*);

double normais_treinamento[][21]=
{
0.2397936628,0.3108997610,0.6824147052,0.5586255743,0.6263353902,...
0.1130863362,0.1274333787,0.2576308350,0.2069796325,0.1998145847,...
0.1763302655,0.2121607820,0.4158173797,0.3218354567,0.3740480019,...
0.5431588283,0.5236377499,1.0000000000,0.7732592181,0.7380088952,...
0.0497800748,0.0670808469,0.1351393999,0.0940642254,0.1164481856,...
0.0906583736,0.2327451059,1.0000000000,0.8501417055,0.2877699385,...
0.1709627281,0.3505728825,1.0000000000,0.5551628555,0.3495708081,...
0.2196687700,0.1397497780,0.2235429689,0.1514758987,0.1451268161,...
0.6167926094,0.5065888905,1.0000000000,0.6733688663,0.2827582302,...
0.0167054791,0.0294017343,0.0952911643,0.0557099076,0.0465493165,...
0.8344038102,0.7152564421,1.0000000000,0.5248069370,0.2599113196,...
1.0000000000,0.7720013861,0.9339116846,0.6037318934,0.4196112727,...
0.2019180060,0.2623761255,0.5753202747,0.4702420337,0.7021806416,...
0.1465915031,0.1489877631,0.2416534136,0.1302777544,0.1033655912,...
1.0000000000,0.3114909219,0.1297625849,0.0913523702,0.0338480833,...
0.4652008117,0.5314471566,0.9457566051,0.6417258690,0.5502218374,...
0.1657638837,0.3239194110,0.8721875801,0.6459833408,0.2790182236,...
0.0475235461,0.0583204361,0.1616403675,0.1379186814,0.0722114799,...
0.5557248696,0.4159706643,1.0000000000,0.5018731705,0.1585546147,...
0.2050146873,0.0745426075,0.0484189894,0.0251057339,0.0752560971,...
1.0000000000,0.6334355071,0.4735713921,0.1952457890,0.1385368515,...
0.3694107434,0.1227533044,0.0329452706,0.0278053041,0.0766436890,...
1.0000000000,0.2099196445,0.0340715926,0.0498694992,0.0034421966,...
0.1136413895,0.3084655829,1.0000000000,0.6442872952,0.1256246731,...
1.0000000000,0.2749252121,0.0711594333,0.0861422942,0.0313912120,...
0.0987631175,0.0882161717,0.1307758979,0.0651697222,0.0508190315,...
0.0973094192,0.1827004392,0.4191679819,0.2309734021,0.0632973768,...
0.5608031176,0.5758474699,0.9253623609,0.5861385444,0.4724599234,...
0.3542883208,0.4952603294,1.0000000000,0.6334999681,0.3409792394,...
0.0776898848,0.2132992708,0.6967830078,0.4404425785,0.0870108828,...
1.0000000000,0.2240211090,0.0303899291,0.0585357538,0.0165524322,...

```

0.0378138146,0.0490250831,0.0916226402,0.0516672166,0.0201020255,...
0.5438689726,0.4894082506,1.0000000000,0.6633906615,0.3195506639,...
0.0569006231,0.0654787137,0.1221013400,0.0692652609,0.0355159720,...
1.0000000000,0.3374853271,0.0628642071,0.0578301106,0.0668281836,...
0.3454074098,0.2753372299,0.3381624329,0.1930485730,0.0921807668,...
0.2175307775,0.2388606640,0.4014414971,0.1733230584,0.0647895759,...
0.4488495904,0.2444722745,0.1633379528,0.0660078038,0.0376356891,...
0.0339762209,0.0408694371,0.0769497192,0.0578247397,0.0617211289,...
0.0468942324,0.0310160024,0.1298789083,0.1036354517,0.1377162471,...
0.5614654722,0.5119474613,1.0000000000,0.6954242456,0.4282467000,...
0.2258626059,0.3280737499,0.6354868073,0.3446282700,0.2426199062,...
0.1980941237,0.2315701190,0.5255750455,0.3945521887,0.1912851249,...
0.1452683143,0.0651414129,0.0534985627,0.0369637820,0.0203012494,...
0.0629177104,0.0279253115,0.0224283151,0.0088338798,0.0110347623,...
0.4866283914,0.3246135489,0.4654543135,0.3165329919,0.2495979423,...
0.1509349973,0.0868187491,0.0933447637,0.0441410320,0.0397342138,...
0.4636013461,0.3978752783,0.7580255146,0.5585940174,0.4487285916,...
0.2546405187,0.2683550805,0.4367532970,0.2373337911,0.1803538291,...
0.6074912555,0.4547847533,0.8647905131,0.6214558382,0.4142852357,...
0.0649980527,0.0627424844,0.1034458293,0.0744718667,0.1054642937,...
0.9302611870,0.3647475696,0.1603567754,0.1042696762,0.1763935956,...
0.4999160677,0.5367938784,0.9391793597,0.6813829035,0.4572238892,...
1.0000000000,0.2498210102,0.0386759408,0.0540471942,0.0147346608,...
0.1769828799,0.3350912668,1.0000000000,0.8506538867,0.4225048692,...
0.2773144015,0.4021837191,1.0000000000,0.6905106633,0.1797230209,...
1.0000000000,0.3670669588,0.0947816021,0.0590813865,0.0531541173,...
1.0000000000,0.2212536782,0.0430460277,0.0758006693,0.0311401793,...
0.3585601267,0.3147598272,0.4516377448,0.3113937147,0.2433181748,...
0.0134613879,0.0172493150,0.0431645669,0.0233663967,0.0787388401,...
0.6885123682,0.6470084163,1.0000000000,0.6429760940,0.3152435629,...
0.5896201315,0.6065490845,1.0000000000,0.6251465740,0.3063762516,...
0.0728422636,0.0507341247,0.0740733136,0.0374497020,0.0212140747,...
0.4874467336,0.5220779247,1.0000000000,0.7046817335,0.3280486744,...
0.2904075259,0.2936921582,0.5834353158,0.6682504690,0.7382430675,...
1.0000000000,0.4440966413,0.2049662595,0.0932266554,0.0620236755,...
0.2921745757,0.1849937545,0.1867243346,0.1257738905,0.1726615492,...
0.1747727454,0.3782410911,1.0000000000,0.5620704970,0.1202822885,...
0.0970231749,0.1672065520,0.3235110593,0.1555499547,0.1610167914,...
0.7538794772,0.5581695926,1.0000000000,0.6886494538,0.3727733421,...
1.0000000000,0.5243621106,0.4197293130,0.2434838565,0.0994443490,...
0.0981929988,0.0197112789,0.0100886404,0.0100758264,0.0097534203,...
0.2855625590,0.3664593565,0.9486626852,1.0000000000,0.9418834037,...
1.0000000000,0.3780056020,0.1554202300,0.1107370121,0.0564259724,...
0.6878844487,0.6547560447,1.0000000000,0.6706140318,0.5798232096,...
0.2299617496,0.2591911669,0.5111696611,0.3500454804,0.2462938563,...
0.6747746864,0.5531525789,1.0000000000,0.4709565069,0.2958085334,...

```

1.0000000000,0.3418486231,0.1349818367,0.1269141879,0.0628732857,...
0.0111660356,0.0150011657,0.0548186484,0.0475582302,0.0178318945,...
0.7281375717,0.6723726660,1.0000000000,0.6208081002,0.3849409977,...
0.9677149107,0.7749673210,1.0000000000,0.7671282867,0.7905511425,...
0.3556226849,0.4000049211,0.7154857000,0.5150563323,0.5049257801,...
0.4273827669,0.5164201706,1.0000000000,0.7455319320,0.5487706111,...
1.0000000000,0.3710786084,0.1382559979,0.1124840768,0.0802292885,...
0.7304898214,0.3881990851,1.0000000000,0.7649861529,0.4457716371,...
0.3074630598,0.3660926817,0.7475964207,0.5868776863,0.4472025164,...
1.0000000000,0.2244588593,0.0386667225,0.0624144474,0.0151183625,...
1.0000000000,0.4189370238,0.1946480441,0.1174671868,0.0673856139,...
0.7099667094,0.2397082648,0.1306816328,0.1102142501,0.1362729474,...
0.3453545074,0.4168462526,0.8169333720,0.5595635934,0.2465603078,...
1.0000000000,0.4355293473,0.9520069248,0.4796861801,0.1437630633,...
0.3055912109,0.3504082160,0.6542573182,0.4905312678,0.4482273319,...
0.1404739929,0.3113532768,0.9660648166,0.5603077943,0.1665809084,...
0.0701596911,0.0825125866,0.1510986891,0.0921503967,0.0960626833,...
0.3683877363,0.3080849933,0.4231907892,0.2846992606,0.1983976699,...
0.0289859789,0.0776748871,0.2360303299,0.1152378224,0.0339915889,...
0.0186298058,0.0395234065,0.1639889419,0.1651710513,0.1781699339,...
1.0000000000,0.7872942318,0.9234600871,0.4801059976,0.1228633929,...
0.3754340150,0.3587090392,0.5658591479,0.3306792915,0.0670247566,...
1.0000000000,0.5690958048,0.4023024543,0.1904293132,0.0921559712,...
0.5952283866,0.3500507052,0.4429982243,0.2997138598,0.2912887970,...
0.7068435835,0.2281482946,0.0554971903,0.0294107362,0.0282674371,...
0.9571581362,0.6506543941,1.0000000000,0.7094850336,0.5105559790,...
0.0568436093,0.0673621593,0.1335166052,0.0647190609,0.0724702844,...
0.1485701762,0.2129262217,0.4005412394,0.1826921330,0.1028815143,...
0.1369794936,0.1473490167,0.2471693695,0.1309507163,0.1048639164,...

```

```
double nodulos_treinamento[][21]=
```

```

{
0.7371261247,0.6632172006,1.0000000000,0.5144438511,0.1444965182,...
0.1738342915,0.0317172103,0.0310842832,0.0501395133,0.1311531351,...
0.2325191988,0.2839356298,0.5964131491,0.4211845084,0.2038075718,...
0.4141306167,0.2101030893,0.1213339835,0.0512471882,0.0811954942,...
0.7881505296,0.6988034639,1.0000000000,0.6202739728,0.2741884000,...
0.4194632870,0.4336321568,0.6131168527,0.1914625586,0.2012048757,...
0.3193641297,0.4350394857,1.0000000000,0.6272956286,0.1602636024,...
0.2991067632,0.3808269552,1.0000000000,0.7929473222,0.3871887262,...
0.3066450899,0.4838283620,1.0000000000,0.4628108029,0.1302572812,...
0.3066450899,0.4838283620,1.0000000000,0.4628108029,0.1302572812,...
0.1290889815,0.0342036556,0.0541247913,0.0595344901,0.0229328927,...
1.0000000000,0.2198498215,0.0310705987,0.0720931710,0.0418301746,...
0.2680687436,0.2678041413,0.4892342571,0.3551506502,0.3972367960,...
0.2424313927,0.0712831158,0.0114913095,0.0111339682,0.0290629928,...

```



```
};
```

```
double edemas_treinamento[][21]=
```

```
{
1.0000000000,0.3310853693,0.1003065529,0.0553093381,0.0101599380,...
0.4046180336,0.0871375754,0.0672189894,0.1222733522,0.1504301237,...
1.0000000000,0.3005887559,0.0815018176,0.0609314337,0.0256204104,...
1.0000000000,0.3532709968,0.0850053418,0.0447339394,0.0268125612,...
1.0000000000,0.1727323708,0.1925125980,0.1738620183,0.1670194213,...
0.9728634299,0.5429554833,0.3498027604,0.1446005066,0.0822641063,...
0.1561609262,0.1159389966,0.1776890984,0.1160651615,0.0500039725,...
0.5237722961,0.5408090380,1.0000000000,0.6744254636,0.4085126382,...
1.0000000000,0.3511923398,0.0788351444,0.0541502487,0.0306587416,...
0.6215199514,0.2383804382,0.1091098952,0.0763785230,0.1238109988,...
0.0380806161,0.0014680714,0.0143629646,0.0123137980,0.0037467706,...
1.0000000000,0.3433794771,0.1470611364,0.1247036575,0.0508098327,...
0.9549694312,0.3634676980,0.1154387696,0.0519106660,0.0300873701,...
0.1277781098,0.0607302434,0.1141160557,0.0674524215,0.0287952599,...
1.0000000000,0.1817617974,0.0820989026,0.0830682533,0.0626236621,...
};
```

```
double normais_testes[][21]=
```

```
{
0.5166233430,0.5495127697,1.0000000000,0.7001120737,0.4546378002,...
0.5046543298,0.5061929514,1.0000000000,0.6713039347,0.3436357697,...
1.0000000000,0.2536167150,0.1244104472,0.0991599198,0.0241321132,...
0.0909878121,0.1077546865,0.1909452795,0.1206474967,0.1200831526,...
0.3757550936,0.4711067686,1.0000000000,0.6987545282,0.2997822164,...
1.0000000000,0.5587738554,0.5425802811,0.3254389183,0.1047722474,...
0.0425940448,0.0477848326,0.1124990472,0.0554229968,0.0538803898,...
1.0000000000,0.3249766972,0.0893253163,0.0596086595,0.0219604913,...
0.5165002804,0.5522973974,1.0000000000,0.6526041844,0.3257191211,...
0.8955084032,0.7426249827,1.0000000000,0.5484926165,0.2178162101,...
1.0000000000,0.6119968701,0.6426219731,0.4290076295,0.2736701437,...
0.1206018802,0.1436784930,0.2486111801,0.1347712202,0.0971297345,...
};
```

```
double nodulos_testes[][21]=
```

```
{
0.0325141588,0.0171428171,0.0639367447,0.0517278405,0.0235130187,...
0.0387559115,0.0248799916,0.0596573652,0.0364975020,0.0356325683,...
};
```

```
double edemas_testes[][21]=
```

```
{
0.3416136493,0.3884076173,1.0000000000,0.8060765021,0.4478236408,...
```

```
1.00000000000,0.3056643826,0.0507765549,0.0405966337,0.0487840708,...
};
```

```
int mc[3][3];
for(int i=0;i<3;i++)
for(int j=0;j<3;j++)
mc[i][j]=0;
int resultado=0;
double menor=999999999;

//comparar normais_testes c/ todos do treinamento
for(int i=0;i<12;i++)
{
for(int j=0;j<106;j++)//comparando com os normais treinamento
if(distancia_euclidiana(normais_testes[i],normais_treinamento[j])<menor)
{
menor=distancia_euclidiana(normais_testes[i],normais_treinamento[j]);
resultado=0;
}
for(int j=0;j<14;j++)//comparando com os nodulos treinamento
if(distancia_euclidiana(normais_testes[i],nodulos_treinamento[j])<menor)
{
menor=distancia_euclidiana(normais_testes[i],nodulos_treinamento[j]);
resultado=1;
}
for(int j=0;j<15;j++)//comparando com os edemas treinamento
if(distancia_euclidiana(normais_testes[i],edemas_treinamento[j])<menor)
{
menor=distancia_euclidiana(normais_testes[i],edemas_treinamento[j]);
resultado=2;
}
if(resultado==0) //e' normal
mc[0][0]++;
else if (resultado==1) //e'nodulo
mc[0][1]++;
else
mc[0][2]++;
}

//comparar nodulos_testes c/ todos do treinamento
menor=999999999;
for(int i=0;i<2;i++)
{
for(int j=0;j<94;j++)//comparando com os normais treinamento
```

```

if(distancia_euclidiana(nodulos_testes[i],normais_treinamento[j])<menor)
{
menor=distancia_euclidiana(nodulos_testes[i],normais_treinamento[j]);
resultado=0;
}
for(int j=0;j<14;j++)//comparando com os nodulos treinamento
if(distancia_euclidiana(nodulos_testes[i],nodulos_treinamento[j])<menor)
{
menor=distancia_euclidiana(nodulos_testes[i],nodulos_treinamento[j]);
resultado=1;
}
for(int j=0;j<15;j++)//comparando com os edemas treinamento
if(distancia_euclidiana(nodulos_testes[i],edemas_treinamento[j])<menor)
{
menor=distancia_euclidiana(nodulos_testes[i],edemas_treinamento[j]);
resultado=2;
}
if(resultado==0) //e' normal
mc[1][0]++;
else if (resultado==1) //e'nodulo
mc[1][1]++;
else
mc[1][2]++;
}

//comparar edemas_testes c/ todos do treinamento
menor=999999999;
for(int i=0;i<2;i++)
{
for(int j=0;j<106;j++)//comparando com os normais treinamento
if(distancia_euclidiana(edemas_testes[i],normais_treinamento[j])<menor)
{
menor=distancia_euclidiana(edemas_testes[i],normais_treinamento[j]);
resultado=0;
}
for(int j=0;j<14;j++)//comparando com os nodulos treinamento
if(distancia_euclidiana(edemas_testes[i],nodulos_treinamento[j])<menor)
{
menor=distancia_euclidiana(edemas_testes[i],nodulos_treinamento[j]);
resultado=1;
}
for(int j=0;j<15;j++)//comparando com os edemas treinamento
if(distancia_euclidiana(edemas_testes[i],edemas_treinamento[j])<menor)
{
menor=distancia_euclidiana(edemas_testes[i],edemas_treinamento[j]);
resultado=2;
}
}

```

```

}
if(resultado==0) //e' normal
mc[2][0]++;
else if (resultado==1) //e'nodulo
mc[2][1]++;
else
mc[2][2]++;
}
//escrever mc
for(int i=0;i<3;i++)
{
for(int j=0;j<3;j++)
printf("\t%d",mc[i][j]);
printf("\n");
}

FILE* f=fopen("saida_90_treino_10_teste_normais.txt","a");
for(int i=0;i<3;i++)
{
for(int j=0;j<3;j++)
fprintf(f,"\t%d",mc[i][j]);
fprintf(f,"\n");
}
fprintf(f,"\n\n");
printf("\n");
printf("\nPORCENTAGEM ACERTOS NORMAIS: %.5f", (100*mc[0][0]/9.0));
printf("\nPORCENTAGEM ACERTOS NODULOS: %.5f", (100*mc[1][1]/8.0));
printf("\nPORCENTAGEM ACERTOS EDEMAS: %.5f ", (100*mc[2][2]/8.0));
printf("\n");
printf("\n");
}
////////////////////////////////////
////////////////////////////////////
double distancia_euclidiana(double* v1, double* v2)
{
double r=0;
for(int i=0;i<21;i++)
r+=(v1[i]-v2[i])*(v1[i]-v2[i]);
r=sqrt(r);
return(r);
}

```


Apêndice II - Publicações durante o mestrado.

Durante o curso de mestrado da autora, os seguintes artigos foram publicados:

- GUIDO R.C. ; SOLGON,R. D. S.B. ; PAULO, K. C. S. ; BARBON JUNIOR, S. ; RODRIGUES, L. C. ; ESCOLA, J. P. L. ; SILVA, I. N. . Introducing the Discriminative Paraconsistent Machine (DPM). Journal Information Sciences. Volume 221. Pages 389-402. Elsevier Science Inc. New York, NY, USA. DOI - 10.1016/j.ins.2012.09.028. Fevereiro, 2013.
- DEZANI H.; BASSI, R. D. S.; MARRANGHELLO, N. ; GOMES, L.; DAMIANI, F.; SILVA, I. N. . Optimizing urban traffic flow using Genetic Algorithm with Petri net analysis as fitness function. Neurocomputing, 124 (2014) 162 - 167. p. <http://dx.doi.org/10.1016/j.neucom.2013.07.015>, 2013.
- GOUVEIA P. S. S ; SOLGON,R. D. S.B. . Problema de quadros de horários escolares sob a ótica da pesquisa operacional. Revista Unorp - Ciências Exatas, v. 3, p. 9-112, 2012.
- SOLGON,R. D. S.B. ; SILVA, I. N.; DEZANI H.; GENARI L.; MARRANGHELLO, N. ; Identificação Inteligente de Patologias no Trato Vocal. CMAC - Sudeste - Congresso de Matemática Aplicada e Computacional. Unesp - Bauru, 2013.
- COTES L. V.; TAVARES M. R.. Transformada Discreta Wavelet Aplicada a Sinais de Áudio. 2012. Iniciação Científica. (Graduandos em Engenharia da Computação) - Centro Universitário do Norte Paulista. Orientadora: Regiane Denise Solgon Bassi.
- GENARI L.; SOLGON R. D. B.; DELORME A. L.; PAULO K. C. S.. Processamento de Imagens Digitais - Detecção de Face. V Encontro sobre Ciência popularização da Ciência e Alfabetização Científica. Semana Nacional de Ciência e Tecnologia. CICC - Centro Integrado de Ciência e Cultura, 2013.