

Id : *abntcite.sty,v1.26*2003/10/1309 : 40 : 41*gweberExp* abntcite
 Chamadas de referencias no texto (NBR10520)



UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA

Análise Espectral de Sinal de Voz

Isaac Teles de Souza

FORTALEZA – CEARÁ

JULHO 2009



UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
CURSO DE GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

Análise Espectral de Sinal de Voz

Autor

Isaac Teles de Souza

Orientadora

Prof^a. Dr(a). Fátima N. Sombra de Medeiros

*Trabalho de conclusão de curso
apresentado à Coordenação do Curso
de Graduação em Engenharia de
Teleinformática da Universidade
Federal do Ceará como parte dos
requisitos para obtenção do grau de
Engenheiro de Teleinformática.*

FORTALEZA – CEARÁ

JULHO 2009

ISAAC TELES DE SOUZA

Análise Espectral de Sinal de Voz

Este Trabalho foi julgado adequado para a obtenção do título de Engenheiro de Teleinformática e aprovado em sua forma final pelo Departamento de Graduação em Engenharia de Teleinformática da Universidade Federal do Ceará.

Isaac Teles de Souza

Banca Examinadora:

Prof^a Dr(a). Fátima N. Sombra de Medeiros
Orientadora

Prof Dr. Francisco José A. de Aquino
Instituto Federal do Ceará

Prof Dr. Sérgio Antenor de Carvalho
Universidade Federal do Ceará

Prof Dr. Walter da Cruz F. Júnior
Universidade Federal do Ceará

Fortaleza, 3 de julho de 2009.

Resumo

A engenharia criou, ao longo da história, inúmeros instrumentos de aplicações e com utilidade médica. Alguns destes servem para auxiliar o médico na identificação de alterações e patologias em órgãos internos do corpo humano, dispensando processos cirúrgicos para tanto. Para detectar e analisar alterações na fala, exames como videolaringoscopia e videoestroboscopia são utilizados, no entanto tais exames permitem interpretações subjetivas de seus resultados. Este trabalho apresenta um estudo realizado sobre o trato vocal e as patologias que o afetam, e um conjunto de técnicas computacionais que medem alterações no espectro do sinal de voz. Estas medidas provêm uma avaliação quantitativa das alterações causadas por patologias que afetam a laringe e podem ser utilizadas, posteriormente, como entradas de uma rede neural artificial na elaboração de um sistema capaz de auxiliar o diagnóstico de doenças da voz.

Este trabalho é dedicado à minha família
cujo amor e sabedoria me inspiram
a melhorar dia após dia.

Agradecimentos

Ao término deste trabalho, deixo aqui meus sinceros agradecimentos:

- a Prof^a. Dr^a. Fátima N. S. de Medeiros e Prof. Dr. Francisco J. A. de Aquino, por toda dedicação, paciência e estímulo;
- a todos os professores e funcionários do Departamento de Engenharia de Telinformática;
- a minha família, pelo incentivo e segurança que me passaram durante todo esse período;
- aos amigos do curso de Engenharia de Teleinformática pelo agradável convívio;
- a todos que direta ou indiretamente contribuíram para a realização deste trabalho.

"Meu sonho não tem fim, e eu tenho muita vida pela frente."

Ayrton Senna

Sumário

Lista de Figuras	vii
Lista de Tabelas	viii
Lista de Siglas	ix
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	2
1.3 Organização do Trabalho	2
2 Fundamentação Teórica	4
2.1 Aparelho Fonador	4
2.2 Patologias da Voz	5
2.2.1 Úlceras de Contato	7
2.2.2 Diplofonia	7
2.2.3 Falsete	7
2.2.4 Afonia Funcional	7
2.2.5 Disfonia Funcional	7
2.2.6 Espessamento de Cordas Vocais	7
2.2.7 Laringite Traumática	8
2.2.8 Disfonia Ventricular	8
2.2.9 Nódulos	8
2.2.10 Cistos	8
2.2.11 Pólipos	8
2.2.12 Edema de Reinke	8
2.2.13 Câncer ou Carcinoma	9
2.2.14 Granuloma	9
2.2.15 Hemangioma	9
2.2.16 Hiperceratose	9
2.2.17 Laringite Infecciosa	9
2.2.18 Leucoplasia	9
2.3 Transformada de Fourier	10

2.3.1	Transformada Discreta de Fourier (TDF)	11
2.3.2	Algumas Propriedades da Transformada de Fourier	12
3	Metodologia	15
3.1	Gravação das Vozes	15
3.2	Análise Preliminar	15
3.3	Medidas de Perturbação	16
3.3.1	Número de Picos Harmônicos	17
3.3.2	Níveis de Ruído	17
3.3.3	Cálculo da Variação dos Picos Harmônicos	18
4	Resultados e Discussões	21
4.1	Vozes Masculinas	21
4.2	Vozes Femininas	23
4.3	Estudo de caso: Edema de Reinke	25
5	Conclusões	28
5.1	Análises e Interpretações	28
5.2	Trabalhos Futuros	29
Apêndice A	Apêndice	32
A.1	Redes Neurais Artificiais	32
A.1.1	Analogia entre neurônios biológicos e artificiais	32
A.1.2	Modelo MCP	33

Lista de Figuras

2.1	Trato vocal	6
2.2	Representação de um sinal a partir de seus componentes frequenciais	11
3.1	Diagrama de blocos da metodologia	16
3.2	Vozes masculinas normal e doente	17
3.3	Vozes femininas normal e doente	18
4.1	Espectro de vozes masculinas normais.	22
4.2	Espectro de vozes masculinas doentes.	22
4.3	Constelações de vozes masculinas normais e doentes	24
4.4	Espectro de vozes femininas normais	24
4.5	Espectro de vozes femininas doentes	25
4.6	Constelações de vozes femininas normais e doentes	26
4.7	Espectro de vozes com edema de Reinke	27
A.1	Neurônio biológico	33
A.2	Modelo de um neurônio artificial	34

Lista de Tabelas

4.1	Vozes masculinas normais.	23
4.2	Vozes masculinas doentes.	23
4.3	Vozes femininas normais.	26
4.4	Vozes femininas doentes.	26

Lista de Siglas

NP	Número de Picos Harmônicos
VP	Variação Média dos Picos Harmônicos
REP	Ruído Entre Picos Harmônicos Médio
RAF	Ruído em Altas Frequências

Capítulo 1

Introdução

O grande progresso experimentado pela engenharia, desde a sua origem aos dias de hoje, tem permitido sua atuação em áreas que normalmente estariam fora de seu alcance, como a área de saúde e medicina. Essa atuação pode ser traduzida, entre outras possibilidades, tanto no desenvolvimento de sofisticadas próteses biomecânicas quanto no uso de técnicas matemáticas e de processamento digital de sinais para auxiliar o médico a estabelecer um diagnóstico mais preciso. Este trabalho de conclusão de curso (TCC) está inserido nessa segunda vertente.

O objetivo principal deste capítulo introdutório é contextualizar o problema, apresentar os objetivos e a estrutura geral do TCC. Com esse intuito, este primeiro capítulo está organizado como segue. Na Seção 1.1 é apresentada a motivação, destacando o problema e algumas idéias para resolvê-lo. Nas Seções 1.2 e 1.3 são descritos, respectivamente, os objetivos e a organização desta monografia.

1.1 Motivação

O exame mais comum realizado por otorrinolaringologistas e fonoaudiólogos é baseado na avaliação perceptual do sistema fonador. Utilizando a audição, o especialista é capaz de analisar o funcionamento fisiológico deste sistema. Características como rouquidão, soprosidade e aspereza são alguns dos indicadores de alterações causadas pelas diversas patologias que afetam a laringe. Estes indicadores são classificados em diferentes graus, de modo que vozes muito soprosas caracterizam a presença de nódulos ou fendas fusiformes, por exemplo. Para este procedimento de graduação dos indicadores, não há uma padronização adequada,

tornando o diagnóstico sujeito a falhas, pois este depende da acuidade auditiva e dos conhecimentos do especialista. Portanto, diferentes diagnósticos podem ser auferidos por médicos diferentes [1].

A visualização do trato vocal é útil na avaliação do sistema fonador, provendo resultados acerca da anatomia deste. O exame de laringoscopia permite a análise do movimento da mucosa que reveste as pregas vocais, a posição das estruturas que dão suporte a fonação, sua coloração e configuração durante a fonação de alguma vogal sustentada; porém, este é um exame desconfortável para o paciente [1].

Com o intuito de auxiliar a elaboração de diagnósticos mais precisos, menos subjetivos e não-invasivos, este trabalho apresenta um conjunto de técnicas computacionais para extrair medidas acústicas de sinais de voz que evidenciem possíveis alterações nas estruturas do trato vocal. Tais medidas acústicas podem, além de auxiliar o médico a diagnosticar várias patologias da voz, apontar o tratamento mais adequado, possibilitando, também, um acompanhamento da evolução do paciente através dos números extraídos de sua voz.

1.2 Objetivos

Este trabalho tem como objetivos principais:

- i. realizar um estudo comparativo entre os espectros de sinais de vozes doentes de pacientes e vozes normais;
- ii. extrair medidas de perturbações existentes em sinais de voz;
- iii. analisar e interpretar os resultados obtidos, de modo que possíveis trabalhos futuros sejam elaborados na perspectiva do desenvolvimento de um sistema de inteligência artificial baseado em redes neurais.

1.3 Organização do Trabalho

Este trabalho está dividido em cinco capítulos e um apêndice. Os capítulos são: introdução, fundamentação teórica, metodologia, resultados e discussões e conclusões. No capítulo de fundamentação teórica, serão apresentados conceitos acerca do funcionamento e da estrutura do aparelho fonador, bem como patologias e seus sintomas associados. No capítulo de metodologia, serão apresentados os passos e critérios utilizados no desenvolvimento das técnicas computacionais que

serão aplicadas na análise espectral de sinal de voz. Nos capítulos seguintes, estarão, respectivamente, os resultados e discussões e as conclusões. No apêndice são introduzidos conceitos básicos de redes neurais artificiais.

Fundamentação Teórica

Neste capítulo serão abordados os seguintes temas: aparelho fonador, as patologias associadas e seus sintomas e a transformada de Fourier, que é a principal ferramenta matemática utilizada, neste trabalho, para a análise dos sinais de voz. Na Seção 2.1 são apresentadas as principais características quanto à anatomia e fisiologia do aparelho fonador humano. Na Seção 2.2 são citadas algumas das patologias que acometem o trato vocal. A transformada de Fourier direta e inversa, a transformada discreta de Fourier e algumas das propriedades desta transformada serão apresentadas na Seção 2.3.

2.1 Aparelho Fonador

Para facilitar o entendimento de seu funcionamento e anatomia, o aparelho fonador, mostrado na Figura 2.1, foi dividido em três sistemas [1,2].

- i. Respiratório ou subglotal: consiste nos pulmões, músculos pulmonares, tubos brônquios e traquéia. Este sistema, além de produzir a respiração, é responsável por gerar o fluxo de ar que se inicia nos pulmões, passa pela traquéia e chega à laringe.

A função respiratória é fundamental na produção da voz, pois as ondas sonoras são resultado do fluxo de ar criado nos pulmões. Como a respiração é limitada pela capacidade de armazenamento de ar dos pulmões, a fonação também é limitada, tanto em altura quanto em duração.

- ii. Fonatório ou laringal: é constituído pela laringe, um anel cartilagenoso situado

na parte superior da traquéia. Dentro da laringe, encontram-se as cordas ou pregas vocais, órgão fundamental à fonação.

As cordas vocais são como dois lábios altamente flexíveis esticados horizontalmente ao longo da abertura central da laringe. Na frente, as cordas estão presas na tireóide e atrás são ligadas às cartilagens aritenóides. Quando as pregas estão separadas, exibem entre elas um espaço chamado de glote e, desta forma, permitem que o ar vindo do pulmão passe livremente, não havendo produção de som. Com a glote fechada, o fluxo de ar vindo do sistema respiratório é forçado a passar e faz com que as cordas vocais vibrem, gerando uma onda sonora de baixa intensidade quase periódica. O período entre duas aberturas sucessivas determina a frequência fundamental (ou *pitch*) da onda. O aumento e a redução do comprimento das cordas vocais modificam a densidade de seu tecido, gerando sons mais altos ou mais baixos respectivamente [1,2].

- iii. Articulatório ou supralaringeal: é formado pelas cavidades ou caixas de ressonância oral e nasal, faringe, língua, palato, nariz, dentes e lábios.

Modificações nas estruturas do sistema articulatório modificam (amplificam e modulam) a onda formada no sistema fonatório, produzindo fonemas e tornando o som inteligível. Este fenômeno é conhecido como ressonância e é similar ao que ocorre em instrumentos musicais. Os elementos ressonadores produzem os diferentes dialetos regionais e auxiliam na identificação de pessoas pela voz [9].

2.2 Patologias da Voz

As patologias da voz são decorrentes do mau funcionamento das estruturas do trato vocal. Isto implica em uma voz diferente do normal, o que caracteriza a existência de algum tipo de patologia. As patologias que afetam o aparelho fonador podem ocorrer em qualquer um de seus três sistemas e podem ser classificadas, quanto à origem, em [1]:

- i. origem funcional: são oriundas do mau uso do aparelho fonador;
- ii. origem orgânica: resultado de anomalias físicas no sistema de produção da fala.

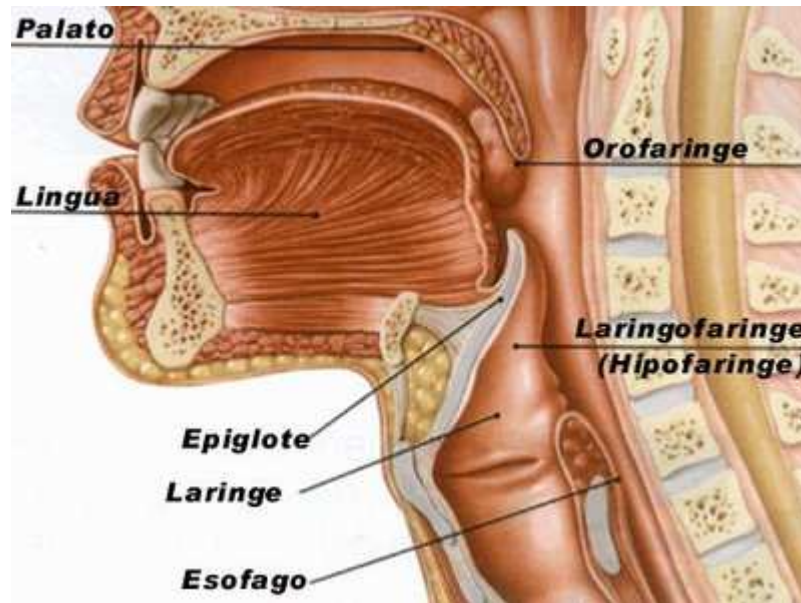


Figura 2.1: Trato vocal.

O mau uso das estruturas de fonação, como produção constante de voz alta, pode provocar o aparecimento de alterações orgânicas pelo uso excessivo de determinadas estruturas, pregas vocais e musculaturas anexas principalmente.

Os diferentes graus de rouquidão, aspereza e soprosidade fazem a diferenciação entre vozes normais e doentes e entre as diversas patologias. Uma voz áspera caracteriza-se por ser estridente e ocorre devido a um elevado estiramento das cordas vocais, exigindo maiores níveis de ar vindos do pulmão e provocando variações na frequência fundamental e amplitude do sinal de voz.

A soprosidade resulta de um escape de ar na região glotal devido ao mau funcionamento dos músculos adutores das pregas, ou por alguma alteração orgânica destas. Espectralmente, a soprosidade apresenta-se como ruído ao longo de todas as frequências, principalmente nas altas.

A rouquidão, por sua vez, é fruto da combinação da aspereza com a soprosidade. As pregas não vibram periodicamente e ocorre escape de ar, produzindo, portanto, alterações na frequência fundamental e ruído espectral.

Algumas das patologias que afetam o trato vocal serão definidas a seguir de acordo com a percepção acústica inerente a cada uma [9].

2.2.1 Úlceras de Contato

Doença de origem funcional provocada por um uso freqüente de limpeza de garganta e tosse, associado com a utilização de freqüência fundamental baixa (voz grave) e com ataque vocal excessivamente brusco. Acusticamente caracteriza-se por rouquidão e aspereza.

2.2.2 Diplofonia

Tem origem funcional e é causada por uma alteração de massa-tamanho-tensão nas pregas vocais. Estas passam a vibrar de forma diferente uma da outra, ocasionando o aparecimento de um sinal com duas componentes harmônicas distintas, não necessariamente múltiplas entre si.

2.2.3 Falsete

Patologia de origem funcional na qual a voz gerada tem altura acima do normal. Acusticamente, essa altura elevada indica escape de ar, ou seja, soprosidade. Ocorre também quebras de freqüência fundamental para baixo.

2.2.4 Afonia Funcional

Como o nome sugere, a afonia têm origem funcional. Ocorre normalmente em pacientes que se tornaram afônicos temporários antes de sua condição permanente e também após alguma patologia laríngea ou outra doença.

2.2.5 Disfonia Funcional

Esta patologia de origem funcional ocorre devido à aproximação das cordas vocais de modo relaxado ou à alta tensão exercida sobre as cordas vocais (aspereza). A percepção acústica pode ser rouquidão, aspereza ou soprosidade, de acordo com o movimento que as cordas vocais realizam.

2.2.6 Espessamento de Cordas Vocais

Patologia de origem funcional que ocorre como consequência do aumento volumétrico ao longo das bordas das cordas vocais. Tem a rouquidão como sintoma

acústico.

2.2.7 Laringite Traumática

A laringite traumática caracteriza-se pelo aumento de massa e volume das pregas vocais devido de tal forma que a vibração das cordas vocais é inadequada, provocando voz rouca de baixa intensidade. É uma doença de origem funcional.

2.2.8 Disfonia Ventricular

A disfonia ventricular tem origem funcional e resulta da adução das pregas ventriculares (falsas pregas) que vibram com a passagem do ar oriundo do pulmão, enquanto as pregas verdadeiras estão ligeiramente abertas. Os sintomas acústicos são parecidos com os da diplofonia, ou seja, o sinal de voz apresenta duas frequências fundamentais diferentes e não múltiplas entre si.

2.2.9 Nódulos

Os nódulos surgem como uma protuberância esbranquiçada na borda glótica das cordas vocais devido ao uso contínuo e abusivo destas e são, portanto, de origem funcional. Produzem aperiodicidade no sinal glotal.

2.2.10 Cistos

Protuberância nas pregas vocais resultante do acúmulo de fluidos em uma estrutura do tipo saco, permitindo escape de ar e, conseqüentemente, dando aspecto de soprosidade à voz do doente. É uma doença de origem orgânica.

2.2.11 Pólipos

Doença de origem funcional que está relacionada com hiperfuncionamento da vocalização. Como sintoma acústico, a voz do paciente apresenta soprosidade e, dependendo do tamanho da lesão, pode causar rouquidão.

2.2.12 Edema de Reinke

As cordas vocais aumentam de tamanho como conseqüência de um edema no corpo gelatinoso de Reinke, que fica logo abaixo da mucosa. Com isso, a frequência fundamental é reduzida, tornando a voz mais grave, instável e acompanhada de

soprosidade. Somados, estes efeitos podem dar uma impressão de rouquidão na voz. Esta doença tem origem funcional.

2.2.13 Câncer ou Carcinoma

Este tipo de patologia grave pode afetar qualquer porção do trato vocal, produzindo efeitos diversos sobre a voz, desde soprosidade leve, até severa rouquidão.

2.2.14 Granuloma

A mucosa das pregas vocais apresenta hiperplasia¹ e granulação de aspecto rígido como resultado de traumatismos nas estruturas da laringe. É classificada como patologia de origem orgânica e seus sintomas são níveis de rouquidão e soprosidade elevados.

2.2.15 Hemangioma

Doença semelhante ao granuloma, apresentando as mesmas características etiológicas e acústicas.

2.2.16 Hiperkeratose

Decorrente da irritação contínua do tecido laríngeo, esta doença, que tem origem orgânica, é caracterizada por placas rosadas que aparecem sobre as pregas vocais. Os efeitos podem ser desprezíveis ou severos (soprosidade e rouquidão).

2.2.17 Laringite Infecciosa

Pacientes com laringite funcional podem vir a ter esta doença. Tem origem orgânica e apresenta como sintoma intensa soprosidade.

2.2.18 Leucoplasia

Acréscimo de massa esbranquiçada sobre a membrana superficial da mucosa. Há redução da altura da voz e diminui a intensidade da frequência fundamental. As cordas vocais podem se tornar assimétricas, gerando diplofonia. É uma doença de origem orgânica considerada pré-cancerosa.

Diante das patologias citadas e dos sintomas por elas apresentados, percebe-se que, acusticamente, há dificuldade em distinguir uma doença de outra se o médico

¹Aumento do número de células em um órgão ou tecido

usar apenas a audição, já que muitas das patologias possuem sintomas acústicos em comum. Além da laringoscopia, que possibilita a visualização do trato vocal, a quantização dos níveis de perturbação do sinal de voz pode ser útil, auxiliando o médico a descobrir a doença da qual o paciente sofre.

2.3 Transformada de Fourier

A transformada de Fourier é uma ferramenta que decompõe um determinado sinal nas suas componentes exponenciais. Para o sinal no tempo $f(t)$, a transformada de Fourier direta $F(\omega)$ representa as amplitudes relativas das várias componentes de frequência do sinal, como ilustrado na Figura 2.2. Portanto, $F(\omega)$ é a representação de $f(t)$ no domínio da frequência [3,4]. $F(\omega)$ é calculada por:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt, \quad (2.1)$$

onde $j = \sqrt{-1}$.

Obtém-se $f(t)$ a partir de $F(\omega)$ através da transformada inversa de Fourier, que é dada por:

$$f(t) = F^{-1}(F(\omega)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega, \quad (2.2)$$

As equações 2.1 e 2.2 são denominadas pares de transformadas de Fourier. Estas transformadas existem se $f(t)$ é uma função contínua e integrável e se $F(\omega)$ é integrável [3,4].

A função $F(\omega)$ geralmente é complexa, como nas equações 2.3 e 2.4, e necessita de dois gráficos para a sua representação: um com a magnitude $|F(\omega)| = [R^2(\omega) + I^2(\omega)]^{1/2}$, chamada de espectro de Fourier de $f(t)$, e outro com a fase $\phi(\omega) = \tan^{-1} \left[\frac{I(\omega)}{R(\omega)} \right]$.

$$F(\omega) = R(\omega) + jI(\omega) \quad (2.3)$$

$$F(\omega) = |F(\omega)| e^{j\theta(\omega)} \quad (2.4)$$

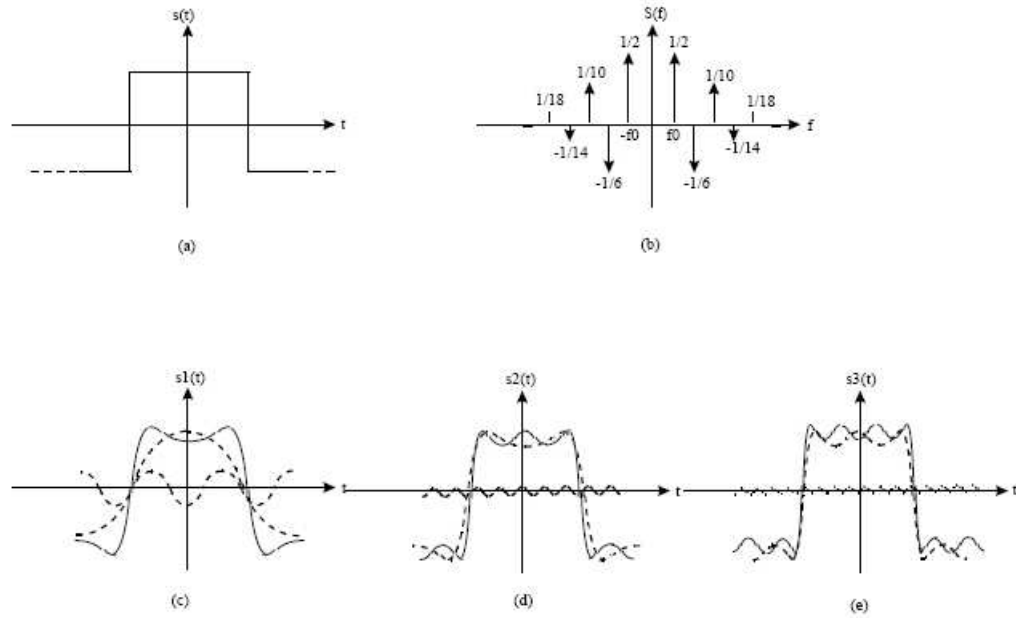


Figura 2.2: Representação de um sinal a partir de seus componentes frequenciais: (a) função pulso; (b) espectro de fourier de (a); (c) resultado de $\cos(2\pi f_0 t) - \frac{1}{3}\cos(6\pi f_0 t)$; (d) resultado de $\cos(2\pi f_0 t) - \frac{1}{3}\cos(6\pi f_0 t) + \frac{1}{5}\cos(10\pi f_0 t)$; (e) resultado de $\cos(2\pi f_0 t) - \frac{1}{3}\cos(6\pi f_0 t) + \frac{1}{5}\cos(10\pi f_0 t) - \frac{1}{7}\cos(14\pi f_0 t)$.

2.3.1 Transformada Discreta de Fourier (TDF)

Considerando-se uma função contínua $f(t)$ e discretizando-a em N seqüências amostradas entre intervalos Δt , esta forma a função discreta $f[n]$ representada pelo conjunto $\{f(t_0), f(t_0 + \Delta t), f(t_0 + 2\Delta t), \dots, f(t_0 + [N - 1]\Delta t)\}$ [4,5].

Portanto, $f[n] = f(t_0 + n\Delta t)$, onde n assume os valores inteiros $0, 1, 2, 3, \dots, N - 1$.

A transformada discreta de Fourier de $f[n]$ e sua inversa são dadas respectivamente pelas equações 2.5 e 2.6:

$$F(\omega) = \frac{1}{N} \sum_{n=0}^{N-1} f[n] e^{-2\pi i \omega n / N}, \quad (2.5)$$

onde ω assume os valores inteiros $0, 1, 2, 3, \dots, N - 1$;

$$f[n] = \sum_{\omega=0}^{N-1} F(\omega) e^{2\pi i \omega n / N}. \quad (2.6)$$

Analogamente a $f(t)$, a sua transformada de Fourier $F(\omega)$ será amostrada entre intervalos $\Delta\omega$. A relação entre Δx e $\Delta\omega$ é dada por [4,5]:

$$\Delta\omega = \frac{1}{N\Delta x}. \quad (2.7)$$

2.3.2 Algumas Propriedades da Transformada de Fourier

Propriedade de linearidade

Se

$$f_1 \leftrightarrow F_1(\omega)$$

$$f_2 \leftrightarrow F_2(\omega),$$

então, para quaisquer constantes arbitrárias a_1 e a_2 ,

$$a_1 f_1 + a_2 f_2 \leftrightarrow a_1 F_1(\omega) + a_2 F_2(\omega) \quad (2.8)$$

Propriedade de escalonamento

Se

$$f(t) \leftrightarrow F(\omega),$$

então, para uma constante real a ,

$$f(at) \leftrightarrow \frac{1}{|a|} F\left(\frac{\omega}{a}\right) \quad (2.9)$$

Desta forma, a propriedade de escalonamento estabelece que a compressão no domínio do tempo é equivalente à expansão no domínio da frequência e vice-versa [3].

Propriedade de deslocamento em frequência

Se

$$f(t) \leftrightarrow F(\omega),$$

então

$$f(t)e^{j\omega_0 t} \leftrightarrow F(\omega - \omega_0). \quad (2.10)$$

O teorema estabelece que um deslocamento de ω_0 no domínio da frequência é equivalente à multiplicação por $e^{j\omega_0 t}$ no domínio do tempo [3].

Propriedade de deslocamento em tempo

Se

$$f(t) \leftrightarrow F(\omega),$$

então

$$f(t - t_0) \leftrightarrow F(\omega)e^{-j\omega t_0}. \quad (2.11)$$

Esse teorema estabelece que, se uma função é deslocada no domínio do tempo de t_0 segundos, então, o seu espectro de amplitude $|F(\omega)|$ permanece inalterado; porém seu espectro de fase é alterado de uma quantidade $-\omega t_0$ [3].

Propriedade da diferenciação e integração em tempo

Se

$$f(t) \leftrightarrow F(\omega),$$

então

$$\frac{df}{dt} \leftrightarrow (j\omega)F(\omega) \quad (2.12)$$

e

$$\int_{-\infty}^t f(\tau)d\tau \leftrightarrow \frac{1}{j\omega}F(\omega). \quad (2.13)$$

Propriedade da diferenciação em frequência

Se

$$f(t) \leftrightarrow F(\omega),$$

então

$$-jtf(t) \leftrightarrow \frac{dF}{d\omega}. \quad (2.14)$$

Propriedade da convolução no tempo e na frequência

Se

$$f_1(t) \leftrightarrow F_1(\omega)$$

e

$$f_2(t) \leftrightarrow F_2(\omega),$$

então

$$f_1 * f_2 \leftrightarrow F_1(\omega)F_2(\omega) \tag{2.15}$$

e

$$f_1 f_2 \leftrightarrow \frac{1}{2\pi} F_1(\omega) * F_2(\omega). \tag{2.16}$$

Capítulo 3

Metodologia

Neste Neste capítulo será apresentada a metodologia adotada no desenvolvimento deste trabalho e ilustrada na Figura 3.1 . Na Seção 3.1 é descrito o processo de gravação dos sinais de voz utilizados neste trabalho. A Seção 3.2 explica como foi realizada a análise preliminar que determinou o desenvolvimento das técnicas computacionais que serão aplicadas aos sinais de voz. Estas técnicas são apresentadas na Seção 3.3.

3.1 Gravação das Vozes

Neste trabalho, foi utilizado um conjunto de 500 vozes, 289 normais e 211 doentes. Estas vozes foram gravadas diretamente pelo computador, através de um microfone modelo LeSon SM58, que minimiza informações espúrias do ambiente. A conversão analógico-digital foi realizada por uma placa SoundBlaster SB16 a uma taxa de amostragem de 22050 Hz e discretização das amostras em 16 bits.

3.2 Análise Preliminar

A partir dos 500 sinais de voz disponíveis, foram selecionados 24 para a realização de uma análise primária, todos correspondentes ao fonema /a/ sustentado [6]. Deste grupo, 12 são sinais de voz feminina e os outros 12 são de voz masculina. Em cada um dos subgrupos de 12, metade corresponde a vozes doentes e o restante a vozes normais. A análise primária consistiu em verificar características inerentes às vozes que sofrem de alguma patologia e às vozes saudáveis, objetivando uma classificação de padrões posterior. As seguintes peculiaridades foram observadas nos espectros

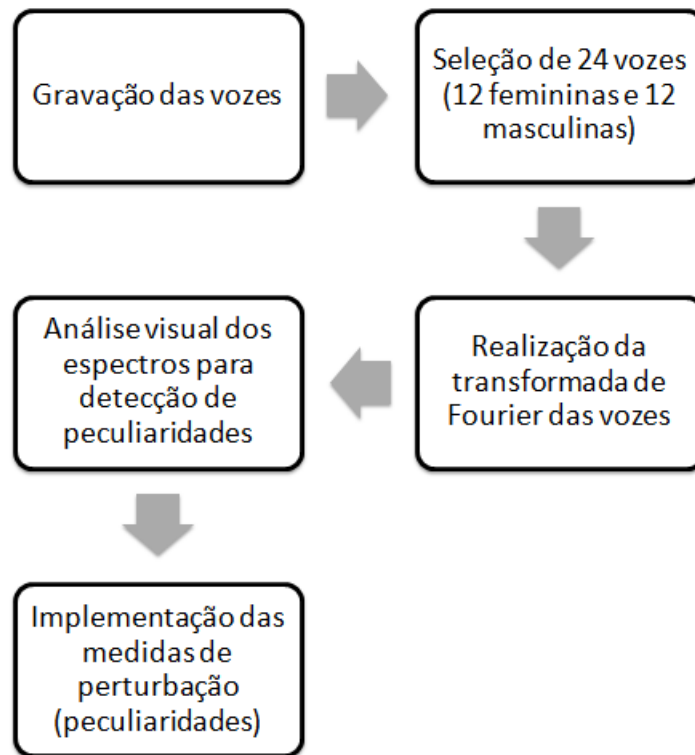


Figura 3.1: Diagrama de blocos da metodologia

de Fourier das Figuras 3.2 e 3.3:

- ▶ picos harmônicos tendem a apresentar menor variação em vozes normais;
- ▶ em vozes doentes, maiores níveis de ruído são percebidos entre picos harmônicos;
- ▶ vozes doentes masculinas tendem a apresentar ruído em altas frequências;
- ▶ algumas vozes doentes apresentaram poucos picos harmônicos.

3.3 Medidas de Perturbação

As técnicas computacionais descritas a seguir foram implementadas para calcular as perturbações percebidas visualmente nos espectros de Fourier dos sinais.

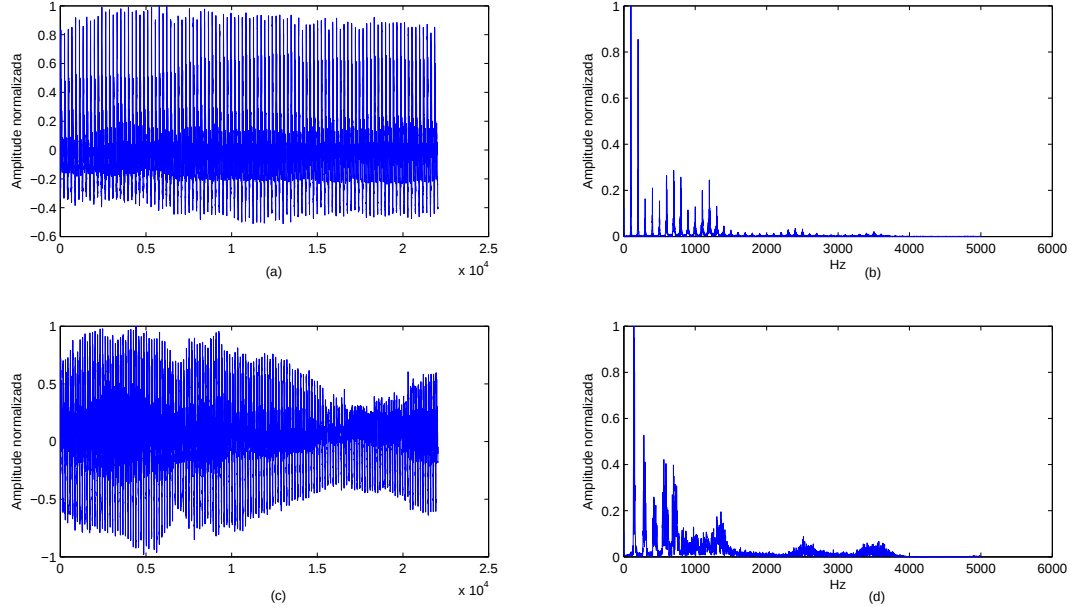


Figura 3.2: (Vozes masculinas: (a) sinal no tempo de uma voz normal; (b) espectro de (a); (c) sinal no tempo de uma voz doente; (d) espectro de (c).

3.3.1 Número de Picos Harmônicos

Como visto na análise primária, o número de picos harmônicos pode vir a ser útil na classificação das vozes em um módulo de reconhecimento de padrões, já que uma voz com número reduzido desses picos foi previamente classificada como doente. O pseudocódigo a seguir mostra como esse cálculo foi realizado.

3.3.2 Níveis de Ruído

Para estimar a quantidade de ruído presente no sinal, considerou-se que há apenas componentes ruidosas entre os picos harmônicos e na faixa das altas frequências. Desta forma, o nível de ruído nesses intervalos do espectro é calculado por [7]:

$$N_R = \frac{\sum_{f=i}^j |S_f(\omega)|^2}{\sum_{n=1}^{NP} |P_n(\omega)|^2}, \quad (3.1)$$

onde $S_f(\omega)$ é a amplitude da componente do sinal com frequência ω e $P_n(\omega)$ é a magnitude dos picos harmônicos. Os valores de i e j definem os intervalos de frequência onde se quer calcular o ruído e NP é o número total de picos.

O pseudocódigo a seguir descreve como os intervalos de frequência $[i,j]$, utilizados

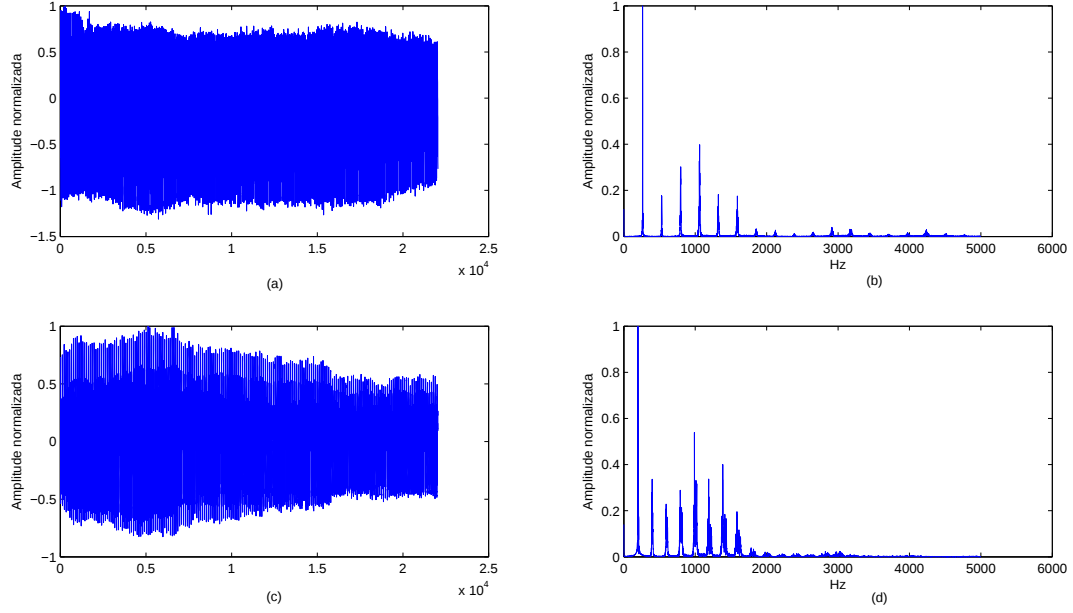


Figura 3.3: Vozes femininas: (a) sinal no tempo de uma voz normal; (b) espectro de (a); (c) sinal no tempo de uma voz doente; (d) espectro de (c).

para calcular o nível de ruído entre os picos harmônicos, são definidos.

3.3.3 Cálculo da Variação dos Picos Harmônicos

Para estimar a variação dos picos harmônicos presente nas componentes fundamentais do sinal, as frequências do espectro normalizado com magnitude acima de um determinado limiar são extraídas do sinal, compondo a matriz $\mathbf{M}$ como é mostrado abaixo [8]:

$$\mathbf{M} = [S_1(\omega_1), S_1(\omega_2), S_1(\omega_3), \dots, S_N(\omega_n)], \quad (3.2)$$

que, para $S_N(\omega_n)$, N corresponde a um pico harmônico e n é o número correspondente à componente harmônica da transformada discreta de Fourier.

Saltos no valor de ω maiores que 90 Hz, caracterizam um novo pico harmônico. Considerando $S_N(\omega_{n_i})$ a primeira componente harmônica próxima do pico N e $S_N(\omega_{n_f})$ a última componente próxima deste, o cálculo da variação da harmônica é feito por:

$$VP = \omega_{n_f} - \omega_{n_i}. \quad (3.3)$$

```
//função para calcular o número de picos harmônicos
```

```
Função [número de picos] = contpicos(coeficientes de Fourier do sinal de voz)  
frequências=busca(coeficientes>(2*media(coeficientes)+desviopadrão(coeficientes)));  
//Procura valores de amplitude de frequência acima do limiar e guarda  
as frequências na matriz frequências.
```

```
Para i = 1:tamanho do vetor frequências
```

```
    Se (frequências(i) - frequências(i-1) > 95)  
        número de picos++;
```

```
    fim Se
```

```
fim Para
```

```
fim Função
```

```
//função que define os intervalos de frequência entre picos harmônicos
```

```
Função [número de picos] = contpicos(coeficientes de Fourier do sinal de voz)  
frequências=busca(coeficientes>(2*media(coeficientes)+desviopadrão(coeficientes)));  
//Procura valores de amplitude de frequência acima do limiar e guarda  
as frequências na matriz frequências.
```

```
Para i = 1:tamanho do vetor frequências
```

```
    Se (frequências(i) - frequências(i-1) > 95)  
        número de picos++;  
        p2 = frequências(i-1);  
        valor = busca(coeficientes == maximo(coeficientes(p1:p2)));  
        frequenciamax(número de picos) = valor;  
        //Guarda os valores das frequências com maior amplitude  
        em cada pico harmônico
```

```
    fim Se
```

```
fim Para
```

```
fim Função
```

```
//função que calcula a variação dos picos harmônicos
Função [número de picos] = contpicos(coeficientes de Fourier do sinal de voz)
frequências=busca(coeficientes>(media(coeficientes)+desviopadrão(coeficientes)));

i1 = frequências(2);
i2 = 1;
cont = 1;
Para i = 1:tamanho do vetor frequências
    Se (frequências(i)-i1 >90)
        i2 = frequências(i-1);
        VP(cont) = i2 - i1 + 1;
        i1 = frequências(i);
        cont = cont+1;
    fim Se
fim Para
saida = VP;
fim Função
```

A função descrita pelo pseudocódigo acima tem como entrada os coeficientes da transformada de Fourier do sinal de voz e gera uma saída com os valores de VP de cada pico harmônico do sinal.

Resultados e Discussões

Neste capítulo são apresentados os resultados e discussões das medidas acústicas aplicadas aos sinais de voz. Na seção 4.1 temos os resultados dos cálculos obtidos a partir das vozes masculinas selecionadas para o estudo. Na seção 4.2 são apresentados os resultados das medidas de perturbação extraídas das vozes femininas. Por fim, na Seção 4.3 um breve estudo de caso é realizado sobre quatro vozes femininas que sofrem de edema de Reinke.

4.1 Vozes Masculinas

Os resultados dos cálculos realizados sobre as 12 vozes masculinas analisadas, cujo espectros de frequência são exibidos nas Figuras 4.1 e 4.2, são apresentados nas Tabelas 4.1 e 4.2. A voz masculina normal identificada pela letra "a" tem $VP = 29,50$, $REP = 0,33$ e $RAF = 0,0092$, enquanto a voz masculina doente identificada pela letra "c" tem $VP = 59,54$, $REP = 8,68$ e $RAF = 0,3436$. Tais valores são coerentes com o que é observado nos espectros de cada voz e se mostram adequados para classificação de padrões. Através destas tabelas, também observa-se que, para vozes masculinas, tanto os valores da variação média dos picos harmônicos (VP), quanto os níveis de ruído em altas frequências (RAF) distinguem melhor vozes normais de doentes.

Na Figura 4.3 são plotados os valores de VP e de RAF das 12 vozes masculinas analisadas. As vozes normais são marcadas com "x" e as vozes doentes com "*". Por esse gráfico, verifica-se que a constelação formada por vozes normais pode ser separada por uma reta da constelação formada por vozes doentes. Tal separação

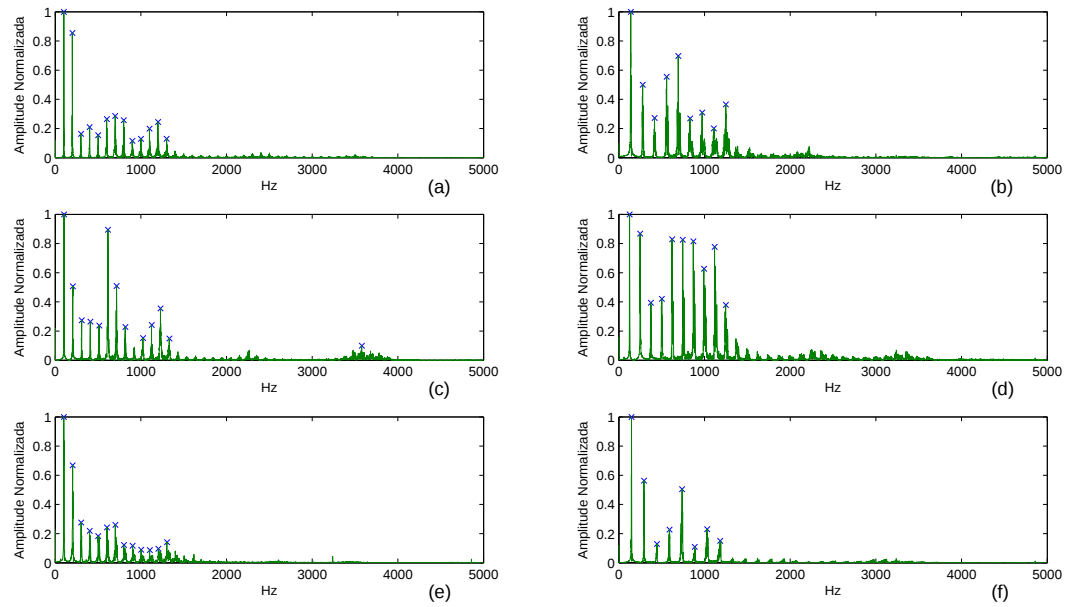


Figura 4.1: Espectro de vozes masculinas: (a),(b),(c),(d),(e) e (f) são normais.

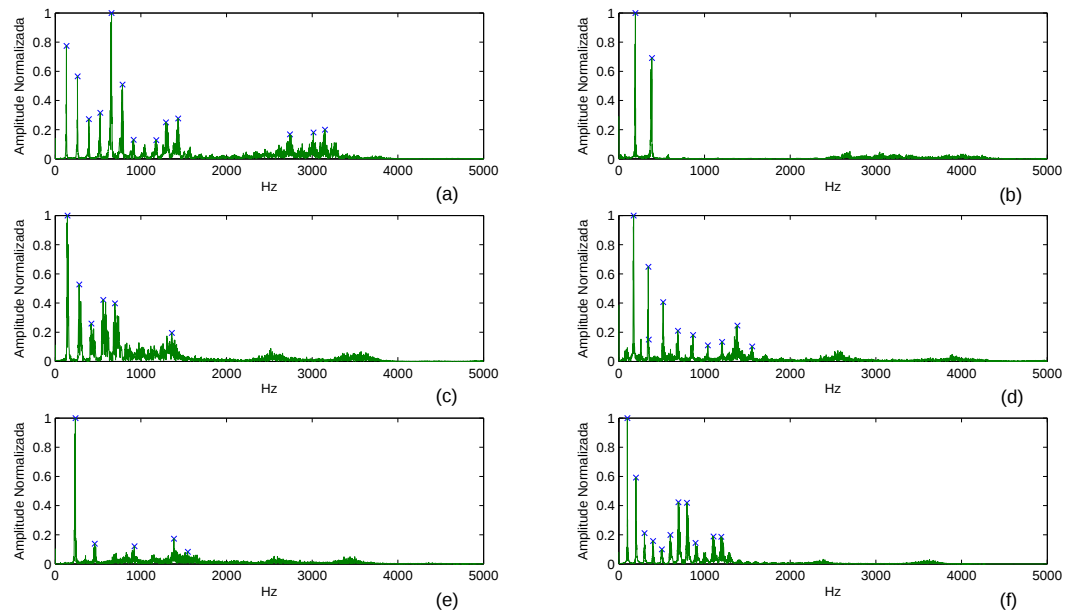


Figura 4.2: Espectro de vozes masculinas doentes: (a) laringite crônica; (b) edema de Reinke; (c) eperemia; (d) pólipos; (e) cordectomia parcial; (f) espessamento

Tabela 4.1: Vozes masculinas normais.

Voz	NP	VP	REP	RAF
a	13	29,50	0,33	0,0092
b	9	37,09	2,33	0,0076
c	13	17,41	1,21	0,0933
d	10	24,60	2,62	0,0496
e	13	25,26	1,62	0,0113
f	8	22,14	1,17	0,0238

Tabela 4.2: Vozes masculinas doentes.

Voz	NP	VP	REP	RAF
a	13	40,64	3,58	0,8095
b	2	31,66	1,70	0,2057
c	6	59,54	8,68	0,3436
d	10	50,23	1,65	0,1845
e	5	51,38	1,11	0,2260
f	11	41,91	2,11	0,0221

pode ser implementada por uma rede neural baseada no modelo perceptron simples, como é explicado no apêndice A.

4.2 Vozes Femininas

Através dos cálculos realizados sobre os espectros de frequência exibidos nas das Figuras 4.4 e 4.5, mostrados nas Tabelas 4.3 e 4.4, observa-se que, para vozes femininas, a variação média dos picos harmônicos (VP) é o parâmetro que mais diferencia vozes normais de doentes. Os níveis médios de ruído entre picos harmônicos (REP), embora menos determinantes que o VP, também se mostram adequados para a discriminação dos padrões.

Na Figura 4.6 o mesmo estudo realizado anteriormente com as vozes masculinas é feito com as vozes femininas. Ao invés dos níveis de ruído em altas frequências (RAF), os níveis de ruído entre picos harmônicos (REP) foram utilizados como

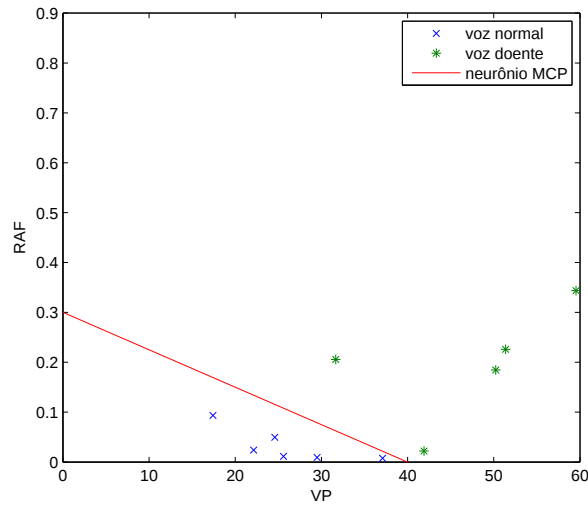


Figura 4.3: Constelações de vozes masculinas normais e doentes.

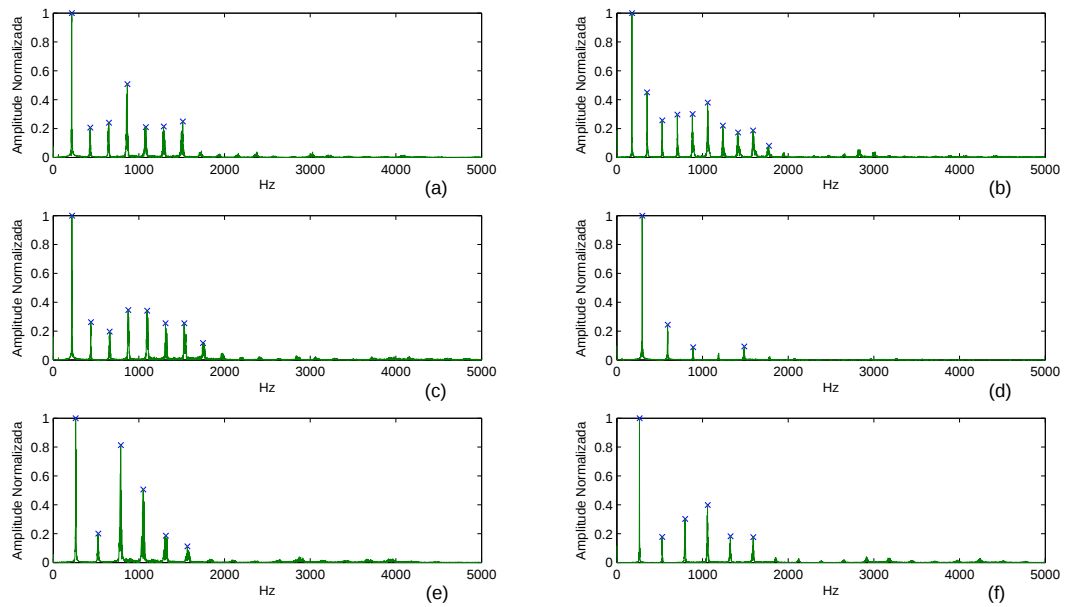


Figura 4.4: Espectro de vozes femininas: (a),(b),(c),(d),(e) e (f) são normais.

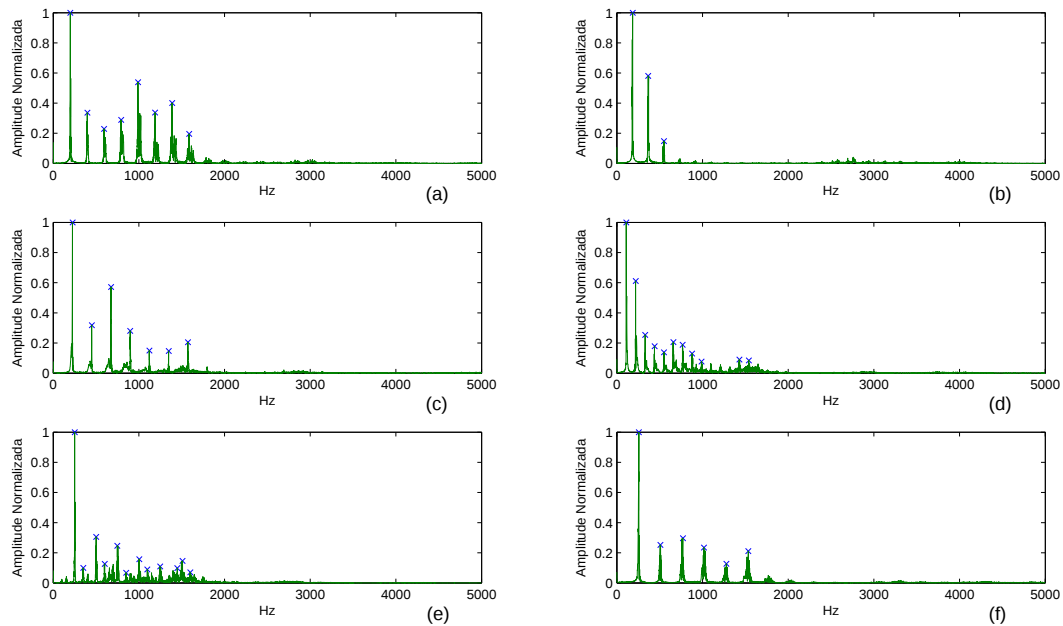


Figura 4.5: Espectros de vozes femininas doentes: (a) cisto; (b) pólipo; (c) edema de Reinke; (d) edema de Reinke; (e) nódulo; (f) cisto.

atributo de discriminação dos padrões. Observa-se que as constelações não podem ser separadas por uma reta e, portanto, uma rede neural baseada no modelo perceptron simples não seria eficaz na classificação destas vozes como normais ou doentes. A utilização de um maior número de atributos, neste caso, certamente é necessário para discriminar os padrões corretamente.

4.3 Estudo de caso: Edema de Reinke

Nesta seção, são apresentados os resultados dos cálculos obtidos sobre sinais voz com edema de Reinke. Esse estudo foi realizado a partir de quatro vozes femininas correspondentes ao fonema /a/ sustentado e cujo objetivo é destacar semelhanças presentes nos espectros das vozes que sofrem desta patologia. A Figura 4.7 mostra os espectros das quatro vozes analisadas, os resultados estão dispostos abaixo de seus respectivos espectros e mostram valores semelhantes para a variação média dos picos harmônicos (VP), ruído médio entre os picos (REP) e ruído em altas frequências (RAF).

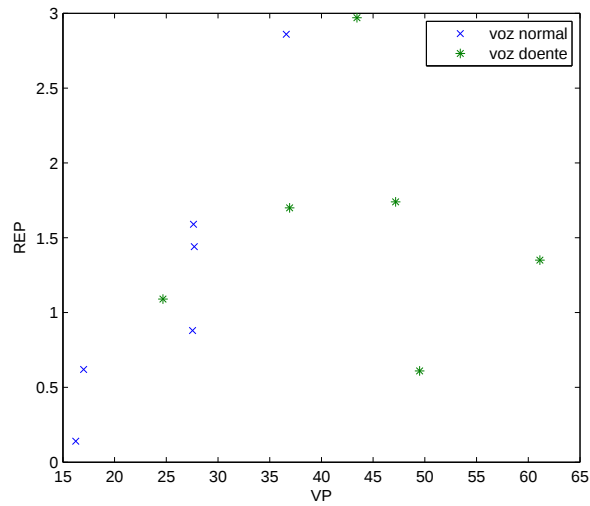


Figura 4.6: Constelações de vozes femininas normais e doentes.

Tabela 4.3: Vozes femininas normais.

Voz	NP	VP	REP	RAF
a	7	27,71	1,44	0,0199
b	10	27,54	0,88	0,0516
c	8	27,63	1,59	0,0411
d	4	16,25	0,14	0,0038
e	6	36,60	2,86	0,0514
f	6	17,00	0,62	0,0439

Tabela 4.4: Vozes femininas doentes.

Voz	NP	VP	REP	RAF
a	8	43,43	2,97	0,0190
b	3	24,67	1,09	0,0520
c	7	49,50	0,61	0,0168
d	11	36,92	1,70	0,0123
e	12	61,13	1,35	0,0126
f	6	47,17	1,74	0,0225

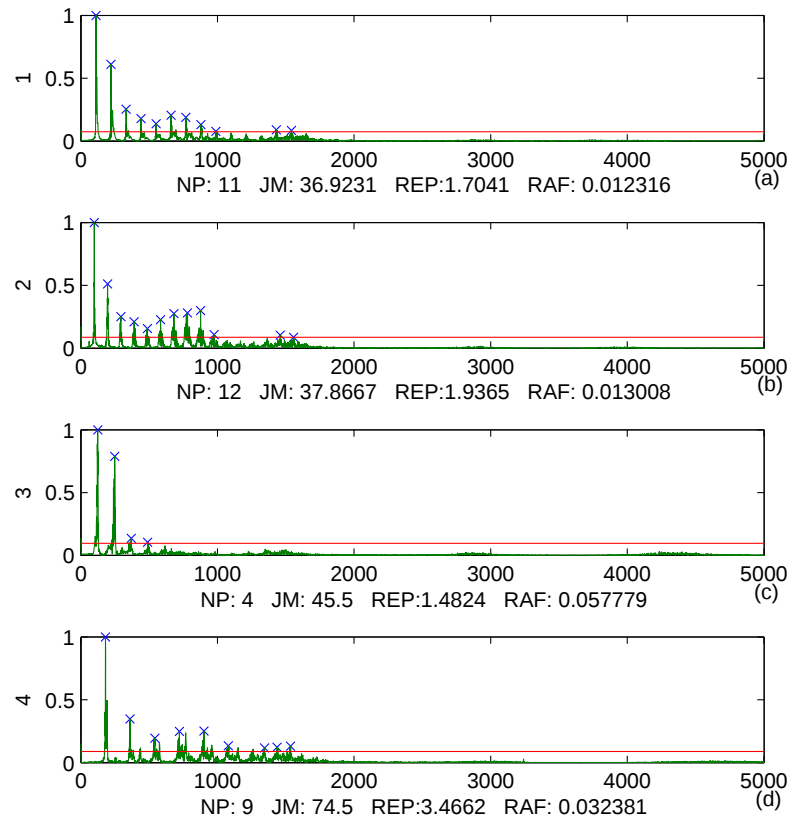


Figura 4.7: Espectro de (a), (b), (c) e (d) são de vozes com edema de Reinke.

Conclusões

Neste capítulo serão apresentadas na Seção 5.1 as análises e interpretações da metodologia adotada e dos resultados obtidos ao final desta monografia. Na Seção 5.2 são propostas novas diretrizes de pesquisa para a continuidade deste trabalho.

5.1 Análises e Interpretações

Este trabalho é resultado de um estudo realizado sobre o aparelho fonador, as patologias que o afetam e as dificuldades encontradas por médicos na tarefa de diagnosticar doenças da voz. Com o intuito de minimizar estas dificuldades de diagnóstico, foi adotada uma metodologia que é fruto de uma análise primária realizada em 24 vozes correspondentes ao fonema /a/ sustentado, das quais metade são vozes femininas e o restante são vozes masculinas. Para a análise das vozes no domínio da frequência, foi utilizada a ferramenta matemática transformada de Fourier, que é descrita na Seção 2.3 no capítulo 2. As diferenças observadas nos espectros de vozes doentes e normais (número de picos harmônicos, ruído entre picos harmônicos, variação dos picos harmônicos e ruído em altas frequências) foram consideradas como alterações conseqüentes das doenças e serviram como diretrizes para o desenvolvimento de técnicas computacionais que pudessem mensurar tais alterações.

As medidas de perturbação foram desenvolvidas no software de aplicações científicas Matlab e algumas destas implementadas renderam resultados adequados, porém insuficientes para discriminação de vozes como normais ou doentes. Para

tal discriminação, que poderia ser realizada por um sistema de inteligência artificial formado por redes neurais ou outros classificadores de padrões, é necessário um número maior de atributos e uma base de dados maior, mais abrangente e mais consistente do que a disponível para este trabalho.

5.2 Trabalhos Futuros

Como trabalhos futuros, é proposto:

- ▶ um estudo mais detalhado sobre o assunto, abordando não somente o espectro, mas também o sinal no domínio do tempo;
- ▶ construir uma base de dados que seja mais adequada e suficiente para o desenvolvimento de um sistema que seja capaz de diagnosticar automaticamente as patologias do trato vocal;
- ▶ uso de transformadas, como, por exemplo, a transformada de Gabor, para destacar detalhes do espectro de voz.

Referências Bibliográficas

[1] ROSA, M. O., *Análise Acústica da Voz para Pré-diagnóstico de Patologias da Laringe*. São Carlos, 1998. Dissertação de mestrado, Escola de Engenharia de São Carlos, Universidade de São Paulo.

[2] FRANÇA, I. F. *Resumo sobre o aparelho fonador e suas correspondências articulatórias*, Disponível em: <http://www.acesin.lettras.ufrj.br/pratica_de_dados/resumo_fonetica.pdf>, acesso : 13/05/2009

[3] LATHI, B. P. *Sistemas de Comunicação*. Rio de Janeiro: Guanabara Dois, 1983.

[4] MINATEL, E. R., *Desenvolvimento de Algoritmo para Reconstrução e Visualização Tridimensional de Imagens Tomográficas com uso de Técnicas Freqüenciais e Wavelets*. São Carlos, 1997. Dissertação de mestrado, Centro de Ciências Exatas e de Tecnologia, Universidade Federal de São Carlos.

[5] HAYKIN, F.; VEEN, B. V. *Sinais e Sistemas*. Porto Alegre: Bookman, 2001.

[6] SAZ, O., SIMÓN, J., RODRIGUEZ, R. W., LLEIDA, E. e VAQUERO, C. "Analysis of Acoustic Features in Speakers with Cognitive Disorders and Speech Impairments",

[7] KASUYA, H., OGAWA, S., MASHIMA, K. e EBIHARA, S. "Normalized Noise Energy as a Acoustic Measure to Evaluate Pathologic Voice," *Journal of the Acoustical Society of America*, v. 80, pp. 1329–1334, 1986.

[8] HORII, Y. "Fundamental Frequency Perturbation Observed in Sustained Phonation," *Journal of Speech and Hearing Research*, v. 22, pp. 5–19, 1979.

- [9] BARRETO, G. A. B. *Perceptron Simples*, 2007. Disponível em:
<<http://www.deti.ufc.br/guilherme/TIP705/perceptron.pdf>>, acesso: 20/05/2009
- [10] DENCKER F. A., *Detecção de Falha na Montagem de Compressores Herméticos por Redes Neurais Artificiais*. Florianópolis, 2002. Dissertação de mestrado, Engenharia Mecânica, Universidade Federal de Santa Catarina.
- [11] MENDES, D. Q. e OLIVEIRA, M. S. *Tutorial de Redes Neurais*, Disponível em: <<http://www.lncc.br/labinfo/tutorialRN/>>, acesso: 11/06/2009

Apêndice

A.1 Redes Neurais Artificiais

Neste apêndice, serão introduzidos conceitos básicos de redes neurais artificiais, fazendo primeiramente uma analogia de um neurônio artificial com o biológico e por último descrevendo o modelo de neurônio artificial proposto por McCulloch e Pitts. Tal estudo é importante devido à potencialidade de uso dos atributos extraídos do espectro de voz como entradas de um sistema baseado neste tipo de inteligência artificial.

A.1.1 Analogia entre neurônios biológicos e artificiais

Pode-se verificar na Figura A.1 que um neurônio biológico é dividido basicamente em três partes [9,10,11]:

- i. dentritos - ramificações correspondentes aos canais de entrada, recebem informações, ou impulsos, vindos de outros neurônios e os conduzem até o corpo celular;
- ii. corpo celular - local onde é feito o balanço energético da célula nervosa (soma das contribuições de energia);
- iii. axônio - caminho de propagação dos impulsos nervosos em direção a outros neurônios.

O ponto de contato entre a terminação axônica de um neurônio e o dentrito de outro é chamado de sinapse.

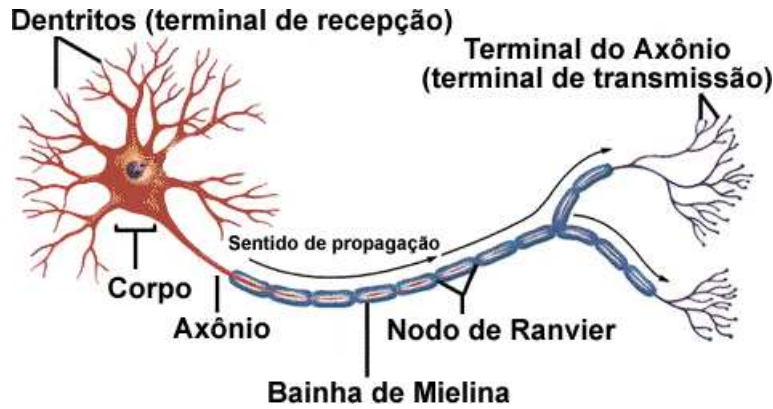


Figura A.1: Neurônio biológico

Modelando os mesmos mecanismos de funcionamento do neurônio biológico, o neurônio artificial da Figura A.2 funciona da seguinte forma [9]:

- i. os dentritos são modelados como linhas ou canais de transmissão por onde fluem as informações de entrada $(X_1, X_2, \dots, X_n)$;
- ii. a força (ou eficiência) das conexões sinápticas de uma certa árvore dentrítica é modelada como um fator (peso sináptico), cujo papel é modular o fluxo dos sinais passantes. Estes pesos estão representados por $(W_1, W_2, \dots, W_n)$;
- iii. a função do corpo celular de realizar o balanço energético é modelada por um somatório sobre as entradas moduladas pelos pesos sinápticos;
- iv. o axônio é modelado por um mecanismo que, após comparar o valor da soma obtida no corpo celular com o limiar de excitação do neurônio, decide se este deve ou não disparar.

A.1.2 Modelo MCP

No modelo proposto por McCulloch e Pitts (MCP), por exemplo, a função de transferência é dada pela função de limiar descrita pela equação 2.1. O neurônio MCP será disparado, ou seja $y = 1$, quando [9,10]

$$S = \sum_{i=1}^n X_i W_i \geq \theta, \quad (\text{A.1})$$

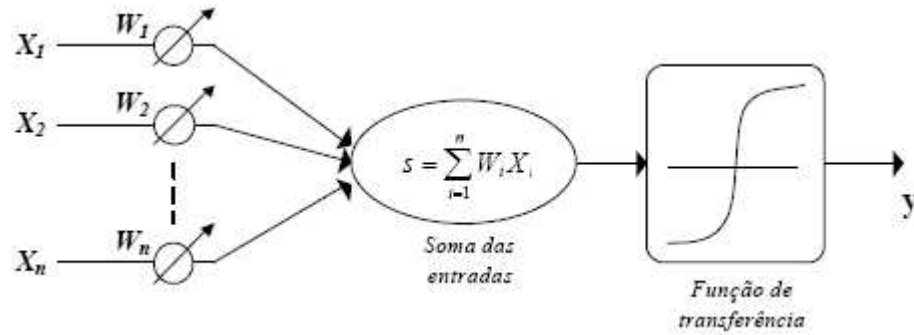


Figura A.2: Neurônio artificial

onde n é o número de entradas, W_i é o peso sináptico associado à entrada X_i e θ é o limiar do neurônio. Quando não é disparado, $y = 0$.

O neurônio MCP, do ponto de vista geométrico, pode ser interpretado como uma reta (2D), um plano (3D) ou um hiperplano que é usado para separar dois pontos ou duas constelações em um gráfico onde cada eixo corresponde a uma entrada (atributo) da rede. Um ponto corresponde a um objeto e uma constelação corresponde a uma classe de objeto. Desta forma, uma rede neural que seja capaz de separar estas constelações será eficaz na classificação dos objetos.

Os pesos sinápticos definem a inclinação das retas, superfícies ou hiperplanos que irão separar as diferentes classes de objetos. Para que o neurônio seja capaz de aprender sozinho, é necessário dotá-lo de uma regra de aprendizagem, que consiste em uma equação que altera os valores dos pesos e do limiar em função dos erros cometidos durante a execução da tarefa de classificação [9].