

**Uma contribuição**  
**sobre a caracterização do**  
**sinal de voz envelhecida**

**UFF - Universidade Federal Fluminense**

**Engenharia de Telecomunicações**

**Aluno: Marco Aurélio Botelho da Silva**

**Orientadores:**

**Edson Cataldo, D.Sc.**

**José Antonio Apolinário Junior, D.Sc.**

**Rio de Janeiro, 22 de janeiro de 2010**

# Resumo

O envelhecimento vocal, também conhecido como presbifonia, consiste na deterioração da voz decorrente de alterações anatômicas e fisiológicas associadas ao envelhecimento.

Esta dissertação tem por objetivo realizar um estudo sobre parâmetros acústicos, extraídos da voz humana, que possibilitam caracterizar uma voz como envelhecida.

Foi realizada uma comparação entre o sinal de voz, o sinal eletroglotográfico (EGG) e o sinal glotal (obtido a partir do sinal de voz por meio de filtragem inversa).

Foram gravadas vozes de locutores masculinos e femininos, em bom estado de saúde física, de diversas idades e divididos de acordo com a faixa etária, em grupos, denominados jovens, adultos e idosos.

Como resultado deste estudo, identificaram-se modificações nos parâmetros acústicos da voz, relacionados com a idade cronológica dos locutores.

Como uma aplicação do estudo, uma rede neural foi utilizada para classificar as faixas etárias, tomando como entrada os parâmetros acústicos da voz.

**Palavras-chave:** envelhecimento da voz, características acústicas da voz

# Abstract

Voice aging, also known as presbyphonia, consists in the deterioration of speech resulting from anatomical and physiological changes associated with aging.

This dissertation aims a study on the parameters of human speech, extracted from the human voice, which may be characterized as an aging voice. A comparison involving speech signal, electroglottographic signal, and glottal signal (obtained from speech signal by means of an inverse filtering) was performed.

We recorded voices from male and female speakers, in good health conditions and different ages; according to their ages, they were divided in age groups named youngsters, adults, and seniors.

As a result of this study, modifications in the acoustic parameters of voice, related to the chronological age of speakers, were identified.

As an application of the the study, a neural network was used to classify the ages, based on acoustic parameters of voice.

**Keywords:** voice aging, speech acoustical features

# Conteúdo

<b>Lista de Figuras</b>	<b>iv</b>
<b>Lista de Tabelas</b>	<b>vii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos da dissertação . . . . .	3
<b>2 Processo de produção da voz humana</b>	<b>4</b>
<b>3 Extração de características da voz</b>	<b>8</b>
3.1 Medidas no sinal de Voz . . . . .	8
3.1.1 Frequência Fundamental - $F_0$ . . . . .	10
3.1.2 <i>Jitter</i> e <i>Shimmer</i> . . . . .	10
3.1.3 <i>HNR</i> - <i>Harmonic-to-Noise Ratio</i> . . . . .	14
3.2 Medidas no sinal Glotal . . . . .	16
3.2.1 O Sinal Glotal . . . . .	16
3.2.2 Filtragem Inversa . . . . .	17
3.2.3 Parâmetros do sinal glotal . . . . .	19
3.3 O Sinal Eletroglotográfico - EGG . . . . .	21
3.4 Os programas Aparat e Praat . . . . .	25

3.4.1	<i>Aparat</i> . . . . .	25
3.4.2	<i>Praat</i> . . . . .	27
<b>4</b>	<b>Gravação dos sinais de voz e extração de características</b>	<b>29</b>
4.1	Comparação de sinais com duas faixas etárias (Grupo I) . . . . .	34
4.1.1	Primeiro caso: Sinais de voz x Sinais do eletroglotógrafo . . . . .	34
4.1.2	Segundo caso: Sinais glotais x Sinais do Eletroglotógrafo . . . . .	49
4.2	Parâmetros obtidos do sinal glotal com três faixas etárias . . . . .	53
4.2.1	Parâmetros usados nas comparações . . . . .	53
4.2.2	Gráficos das análises modificados . . . . .	63
4.3	Dispersão dos parâmetros . . . . .	67
<b>5</b>	<b>Aplicação</b>	<b>76</b>
5.1	Redes Neurais Artificiais - RNA . . . . .	76
5.2	Seleção de atributos e Método <i>Wrapper</i> . . . . .	78
5.3	Montagem da Rede Neural utilizando parâmetros do sinal de voz e sinal glotal . . . . .	80
5.3.1	Construção da Rede Neural . . . . .	81
5.3.2	Resultado . . . . .	83
<b>6</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>85</b>
6.1	Conclusões . . . . .	85
6.2	Trabalhos Futuros . . . . .	89
<b>A</b>	<b>Programa <i>Praat</i></b>	<b>90</b>
<b>B</b>	<b>Programa <i>Aparat</i></b>	<b>101</b>

C Tabelas diversas 108

Bibliografia 114

# Lista de Figuras

2.1	Aparelho fonador [57]. . . . .	5
2.2	Cordas Vocais [57]. . . . .	6
2.3	Laringe [26]. . . . .	6
3.1	Sinal glotal e sua derivada [57]. . . . .	17
3.2	(a)Parâmetros do sinal glotal; (b)Sinal glotal [27]. . . . .	19
3.3	Eletroglotógrafo [43]. . . . .	22
3.4	Forma de onda EGG e DEGG [44]. . . . .	23
3.5	Fases de fechamento e abertura glotal [43]. . . . .	24
3.6	Microfone - LS1 - C1 [57]. . . . .	25
3.7	Exemplo de forma de onda do sinal no pacote Aparat [48]. . . . .	26
3.8	Exemplo de janela “parâmetros do sinal” no pacote Aparat [48]. . . . .	27
3.9	Uma janela de visualização do Praat. . . . .	28
4.1	Interface de gravação e microfone. . . . .	31
4.2	Frequência fundamental(Hz) - Masculino. . . . .	37
4.3	Frequência fundamental(Hz) - Feminino. . . . .	38
4.4	Jitter - Masculino. . . . .	40
4.5	Jitter - Feminino. . . . .	41

4.6	Shimmer - Masculino. . . . .	44
4.7	Shimmer - Feminino. . . . .	45
4.8	HNR (dB)- Masculino. . . . .	47
4.9	HNR (dB)- Feminino. . . . .	48
4.10	F0 (Hz)- Sinal Eletroglotógrafo - Sinal glotal. . . . .	49
4.11	Jitter - Sinal Eletroglotógrafo - Sinal glotal. . . . .	50
4.12	Shimmer - Sinal Eletroglotógrafo - Sinal glotal. . . . .	51
4.13	HNR (dB)- Sinal Eletroglotógrafo - Sinal glotal. . . . .	52
4.14	F0 e Jitter - Masculino e Feminino. . . . .	54
4.15	Shimmer e HNR - Masculino e Feminino. . . . .	55
4.16	pp e K0 - Masculino e Feminino. . . . .	56
4.17	DH12 e HRF - Masculino e Feminino. . . . .	57
4.18	PSP e NAQ - Masculino e Feminino. . . . .	58
4.19	Aq e CIQ - Masculino e Feminino. . . . .	59
4.20	OQ1 e OQ2 - Masculino e Feminino. . . . .	60
4.21	Oqa e Qoq - Masculino e Feminino. . . . .	61
4.22	SQ1 e SQ2 - Masculino e Feminino. . . . .	62
4.23	F0 e Jitter - Masculino e Feminino. . . . .	63
4.24	Shimmer e HNR - Masculino e Feminino. . . . .	64
4.25	pp e K0 - Masculino e Feminino. . . . .	65
4.26	DH12 e HRF - Masculino e Feminino. . . . .	66
4.27	Locutor Masculino Faixa etária 1 e 3 (eixo x-fases do sinal no tempo/eixo y-valores dos parâmetros). . . . .	68
4.28	Locutor Masculino Faixa etária 1 e 3. . . . .	69



4.29	Locutor Feminino Faixa etária 1 e 3. . . . .	70
4.30	Locutor Feminino Faixa etária 1 e 3. . . . .	71
5.1	Abordagem Wrapper [28] . . . . .	80
5.2	Rede Neural MLP para o grupo feminino . . . . .	82
5.3	Rede Neural MLP para o grupo masculino . . . . .	83
A.1	Pulsos do sinal . . . . .	92
A.2	Janela Objeto e Figura . . . . .	94
A.3	Janela gravação . . . . .	95
A.4	Janela Itens Objetos . . . . .	96
A.5	Janela de Edição . . . . .	97
A.6	Espectrograma do sinal . . . . .	98
A.7	Frequências do sinal . . . . .	98
A.8	Intensidade do sinal . . . . .	99
A.9	Formantes do sinal . . . . .	99
A.10	Pulsos do sinal . . . . .	100
B.1	Espectro Locutor 1 . . . . .	102
B.2	Plano-z Locutor 1 . . . . .	103
B.3	Fase Locutor 1 . . . . .	104
B.4	Trato vocal Locutor 1 . . . . .	104

# Lista de Tabelas

4.1	<i>Jitter/Shimmer/HNR</i> - Locutora 99 anos . . . . .	30
4.2	Frequência Fundamental (Hz) - Locutores masculinos com microfone. . . . .	35
4.3	Frequência Fundamental (Hz) - Locutores masculinos com eletroglotógrafo. . . . .	35
4.4	Frequência Fundamental(Hz) - Locutores femininos com microfone. . . . .	35
4.5	Frequência Fundamental(Hz) - Locutores femininos com eletroglotógrafo. . . . .	36
4.6	<i>Jitter</i> - Locutores masculinos com microfone. . . . .	39
4.7	<i>Jitter</i> - Locutores masculinos com eletroglotógrafo. . . . .	39
4.8	<i>Jitter</i> - Locutores femininos com microfone. . . . .	39
4.9	<i>Jitter</i> - Locutores femininos com eletroglotógrafo. . . . .	40
4.10	<i>Shimmer</i> - Locutores masculinos com microfone. . . . .	42
4.11	<i>Shimmer</i> - Locutores masculinos com eletroglotógrafo. . . . .	42
4.12	<i>Shimmer</i> - Locutores femininos com microfone. . . . .	43
4.13	<i>Shimmer</i> - Locutores femininos com eletroglotógrafo. . . . .	43
4.14	<i>HNR</i> (dB) - Locutores masculinos com microfone. . . . .	46
4.15	<i>HNR</i> (dB) - Locutores masculinos com eletroglotógrafo. . . . .	46
4.16	<i>HNR</i> (dB) - Locutores femininos com microfone. . . . .	46
4.17	<i>HNR</i> (dB) - Locutores femininos com eletroglotógrafo. . . . .	47

4.18	Parâmetros Grupo I. . . . .	74
4.19	Parâmetros Grupo II. . . . .	75
C.1	Informações cadastrais- Locutores . . . . .	108
C.2	Tabela de <i>Jitter</i> . . . . .	109
C.3	Tabela de <i>Shimmer</i> . . . . .	110
C.4	Tabela de <i>HNR</i> . . . . .	111
C.5	Tabela de Frequência Fundamental . . . . .	112
C.6	Parâmetros Locutora 99 anos . . . . .	113

# Capítulo 1

## Introdução

O envelhecimento do ser humano é um processo natural que, com o avanço da idade, pode acarretar alguns problemas vocais [10] [23]. Esse assunto é de grande interesse de diversos autores e pesquisadores na área de voz, pois o envelhecimento torna o organismo humano vulnerável, o que reflete diretamente na produção da voz [26] [52] [65] [76].

Por ser o envelhecimento um processo fisiológico, contínuo e irreversível [65], grandes transformações físicas ocorrem nas pessoas idosas. Não podemos definir o envelhecimento somente pela cronologia, mas também pelas condições física e social nas quais se apresentam os indivíduos [65].

Uma série de modificações na voz aparecem com a idade e alguns dos seguintes fatores contribuem para seu envelhecimento:

- Modificação do posicionamento da laringe;
- Mobilidade da laringe devido a calcificações das cartilagens e atrofia muscular [65];
- Elasticidade muscular devido à redução de massa [66];

- Problemas respiratórios devido à modificação da cavidade pulmonar;
- Modificação do comprimento das cordas vocais;
- Problemas auditivos, visuais, clínicos e psicológicos.

Como causas do envelhecimento da voz, podemos citar as alterações que ocorrem no organismo como, por exemplo, perda de capacidade pulmonar, alterações no epitélio que reveste a laringe, atrofia da musculatura laríngea, calcificação das cartilagens e mudanças nas vibrações das pregas vocais [76].

Problemas patológicos são responsáveis também pelo envelhecimento da voz, tais como as fendas glóticas - espaços entre as cordas vocais que dificultam o fechamento por completo [23], assimetria das cordas vocais, redução vibratória causada pela atrofia muscular, voz rouca, tremor na voz, dificuldade de sustentação da frequência fundamental, volume, soprosidade, pressão aérea subglótica, redução na velocidade da fala, pausas articulatórias e outras alterações [23] [47].

O sinal glotal e seus parâmetros vem sendo utilizados em diversas pesquisas e aplicações clínicas, tais como codificação, síntese e reconhecimento automático de locutor. Muitos desses parâmetros são obtidos utilizando técnicas invasivas e não invasivas [63] [73], e vamos utilizá-los para discutir o envelhecimento da voz.

Os parâmetros serão extraídos de três tipos de sinais distintos: o sinal de voz, o sinal eletroglotográfico e o sinal glotal.

O sinal de voz é aquele obtido diretamente na gravação da voz através do microfone, o sinal eletroglotográfico é obtido através do Eletroglotógrafo e o sinal glotal é obtido através da filtragem inversa do sinal de voz.

O método de eletroglotografia [44] é uma técnica não invasiva que permite analisar o contato entre as cordas vocais em fonação, medindo as variações de

impedância entre elas, identificando o momento de abertura e fechamento da glote. Aplicam-se dois eletrodos no pescoço gerando uma corrente de baixa intensidade, que não produz nenhum desconforto durante a fonação [43] [44] [63]. Outras técnicas não invasivas, como filtragem inversa, fotografia estroboscópica, imagens de alta velocidade e medidas de pressão subglotal [57] [63] são utilizadas na estimação do sinal glotal e permitem obter informações sobre o contato das cordas vocais.

Neste trabalho, procuramos identificar as características de uma voz envelhecida utilizando técnicas não invasivas para gravar vogais sustentadas e formar um grande banco de dados, do qual foram extraídos alguns parâmetros acústicos que nos permitiram obter resultados relacionados com a idade.

## **1.1 Objetivos da dissertação**

O objetivo principal dessa dissertação é o de levantar características da voz humana que permitam classificar uma voz como envelhecida.

## Capítulo 2

# Processo de produção da voz

## humana

Para entender o processo de envelhecimento da voz é preciso, primeiramente, compreender sua produção. A voz é uma ferramenta necessária para a comunicação do ser humano, e sua produção ocorre pelo trabalho em conjunto dos sistemas respiratório, digestivo e nervoso, além das cartilagens, músculos, ligamentos e ossos [73].

São três os subsistemas que atuam sucessivamente na produção da fala: o **respiratório**, responsável pelo fluxo de ar gerado, o **laringeal**, que é a parte do sistema onde se encontra a glote (espaço entre as cordas vocais) e o **supralaringeal**, que atua como filtro modificando o espectro da voz para cada indivíduo.

Devido à contração e à expansão dos pulmões, um fluxo de ar é gerado criando uma diferença de pressão entre o ar nos pulmões e a pressão do ar na boca, provocando, assim, um escoamento de ar [73]. Após passar pelas cordas vocais, esse escoamento transforma-se em uma série de pulsos de ar, formando o sinal glotal, que é um sinal quase periódico e de baixa intensidade, cuja frequência é denominada “frequência fundamental da voz”. Quando as cordas vocais estão vibrando,

os pulsos de ar sofrem influência do sistema de ressonância formado pelos tratos vocal e nasal, que funcionam como um filtro, gerando frequências de ressonância chamadas de formantes, e têm um papel importante na inteligibilidade da fala [11]. A geração dos diferentes sons que emitimos quando falamos só é possível devido a essa filtragem.

Os formantes são os efeitos ressonantes no trato vocal relacionados à amplificação da energia do som no subsistema supralaríngeo. Estão relacionados à anatomia e às configurações específicas do aparelho fonador de cada indivíduo, e têm um papel importante na inteligibilidade da fala [73].

A Fig 2.1 mostra um esquema do aparelho fonador.

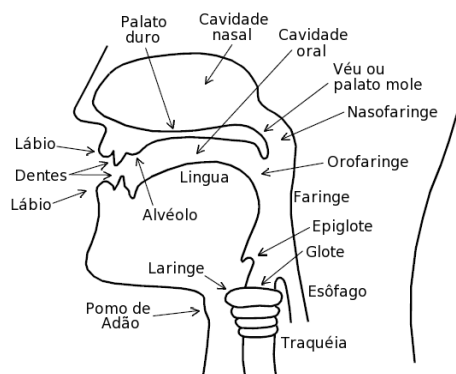


Figura 2.1: Aparelho fonador [57].

As cordas vocais, ilustradas na Fig 2.2, são as principais responsáveis pela geração da voz e são constituídas por duas membranas situadas na laringe, unidas às cartilagens tiróide e aritenóides e aos músculos intrínsecos e extrínsecos.



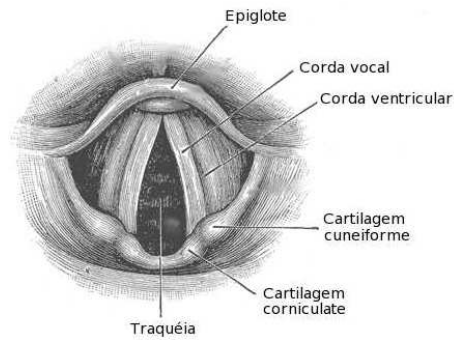


Figura 2.2: Cordas Vocais [57].

Na infância, os tecidos epiteliais da laringe são densos, abundantes e mais vascularizados, com tendência à obstrução e ao edema. As cordas vocais medem de 6 a 8 milímetros, as cartilagens são delicadas e os ligamentos frouxos. Até a puberdade, há semelhanças no órgão em ambos os sexos, daí a dificuldade de se discriminar a voz de um menino ou de uma menina. Na adolescência, com a puberdade, associado ao crescimento corporal por ação hormonal, ocorre o crescimento da laringe, ou seja, a laringe infantil transforma-se em laringe adulta. A Fig 2.3 mostra uma laringe na fase adulta.

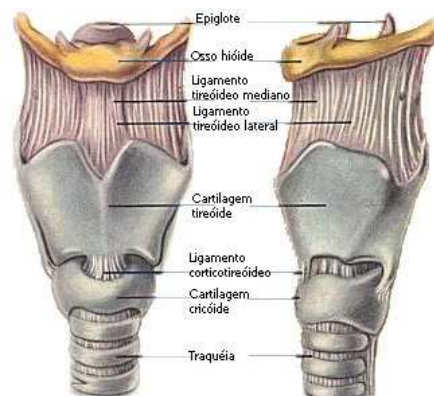


Figura 2.3: Laringe [26].

No adulto, a laringe atinge o estágio de conclusão de seu desenvolvimento a partir dos 18 anos, quando as cordas vocais medem 17 milímetros no sexo feminino e de 17 a 23 milímetros no sexo masculino. A voz passa a ser considerada madura, sendo caracterizada pelo pleno controle da intensidade vocal, podendo produzir variações na frequência e na qualidade.

# Capítulo 3

## Extração de características da voz

### 3.1 Medidas no sinal de Voz

Através de algumas medidas acústicas da voz, informações sobre a saúde vocal podem ser obtidas, distinguindo aquelas que ocorrem naturalmente com a idade daquelas associadas a alguma patologia. Diversas perturbações ocorrem na produção da voz, baseadas na frequência [75], na amplitude e na forma de onda e, muitas vezes, por serem pequenos distúrbios ou mudanças temporárias não alteram o padrão de forma visual e nem definitivo do sinal. Citaremos alguns exemplos de perturbações ocorridas na voz, embora nem todas sejam utilizadas no nosso estudo:

- *Flutuação*: reflete uma instabilidade do sistema. São flutuações na frequência fundamental ou na amplitude do sinal de voz.
- *Variabilidade*: é a habilidade de modificar a voz por intenção ou acidente.
- *Jitter*: é um tipo de perturbação ou variabilidade da frequência fundamental da voz.

- *Shimmer*: é um tipo de perturbação ou variabilidade da amplitude da frequência fundamental da voz.
- *Tremor*: é uma pequena flutuação na amplitude ou na frequência do sinal de voz ou em ambas no sinal. Originalmente, é um problema neurológico ou fisiológico.
- *Vibrato*: é diagnosticado como sendo um tremor fisiológico estável nos músculos da laringe [73].
- *Aspereza*: é uma irregularidade na energia do sinal. Podemos ter uma rigidez da mucosa que pode causar uma irregularidade vibratória. A voz fica sem projeção, com ruídos nas frequências altas.
- *Respiração*: encontram-se momentos de pausa para respirações durante a fonação.
- *Rouquidão*: encontra-se em uma combinação de *Aspereza* e *Respiração* durante a fonação. Ocorre irregularidade na vibração das cordas vocais. presença de ruído de baixa frequência.
- *Soprosidade*: presença de ruído de fundo, audível. Refere-se ao fechamento glótico incompleto.
- *Astenia*: é um termo empregado em medicina para designar uma fraqueza orgânica, porém sem perda real da capacidade muscular. Reduz a energia de emissão do som.

Muitas medidas acústicas têm sido utilizadas na tentativa de mostrar a evolução do mecanismo vocal ao longo da vida e, entre as mais conhecidas, podemos citar a Frequência Fundamental  $F_0$ , *Jitter*, *Shimmer* e *HNR* (*Harmonic-to-Noise Ratio*). Há também outras, que serão discutidas ao longo do capítulo.

### 3.1.1 Frequência Fundamental - $F_0$

A frequência de vibração das cordas vocais determina frequência fundamental da voz, sendo representada por  $F_0$ , e tem relação direta com a massa, elasticidade e comprimento das cordas vocais [65]. Quanto mais rígido for o tecido e quanto mais curta for a vibração das cordas vocais, maior será a frequência fundamental. Por outro lado, quanto maior a massa da corda vocal, menor a frequência fundamental [47].

No sexo masculino, a frequência de vibração das cordas vocais situa-se entre 50 e 250Hz, enquanto que, no sexo feminino, de 120 a 300Hz. Na infância, pode chegar a 500Hz [73].

Um estudo clínico da frequência fundamental pode contribuir para a identificação da idade, sexo e faixa etária do locutor [65]. Com o avanço da idade ocorre um aumento de  $F_0$  para os locutores do sexo masculino e uma diminuição para os do sexo feminino [65].

### 3.1.2 *Jitter* e *Shimmer*

*Jitter* e *Shimmer* são medidas que têm contribuição nas comparações entre vozes jovens e envelhecidas, mas não fornecem informações sobre a idade cronológica do locutor. No sexo masculino, vozes com altos valores de *Jitter* e/ou *Shimmer* são

percebíveis pelo ouvido humano com aspereza, indicando avanço da idade ou problemas patológicos [52] [65]. No sexo feminino, um alto valor de *Jitter* não permite concluir que essas vozes sejam envelhecidas, pode estar dentro das características normais [26] [65].

A seguir, citamos algumas fórmulas para o cálculo de diversos tipos de *Jitter* [13]:

- *Jitter (local)*: é a média do módulo da diferença entre dois períodos consecutivos da Pitch, dividido pelo período médio.

$$Jitter_{local} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |P_{i+1} - P_i|}{\frac{1}{N} \sum_{i=1}^N P_i}$$

onde N é o número de amostras e  $P_i$  o período.

- *Jitter (rap)*: é a média do módulo da diferença entre um período e o período médio em uma janela, que contém este período e seus vizinhos anterior e posterior, dividido pelo período médio.

$$Jitter_{rap} = \frac{\frac{1}{N-J+1} \sum_{i=1}^{N-J+1} \left| \frac{1}{J} \left( \sum_{j=1}^J P_{i+j-1} \right) - P_{i+0,5(J-1)} \right|}{\frac{1}{N} \sum_{i=1}^N P_i}$$

onde J é o tamanho da janela e a utilizada é igual a 3, N é o número de amostras,  $P_i$  é o período.

- *Jitter (ppq5)*: é a média do módulo da diferença entre um período e o período médio em uma janela, que contém este período e seus quatro períodos vizinhos mais próximos, ou seja, dois períodos anteriores e dois posteriores, dividida pelo período médio.

$$Jitter_{ppq5} = \frac{\frac{1}{N-J+1} \sum_{i=1}^{N-J+1} \left| \frac{1}{J} \left( \sum_{j=1}^J P_{i+j-1} \right) - P_{i+0,5(J-1)} \right|}{\frac{1}{N} \sum_{i=1}^N P_i}$$

onde J é o tamanho da janela e a utilizada é igual a 5, N é o número de amostras,  $P_i$  é o período.

- *Jitter (ddp)*: é a média do módulo da diferença entre as diferenças consecutivas de períodos consecutivos, dividida pelo período médio.

$$Jitter_{ddp} = \frac{\frac{1}{N-3} \sum_{i=1}^{N-3} \left| |P_{i+3} - P_{i+2}| - |P_{i+1} - P_i| \right|}{\frac{1}{N} \sum_{i=1}^N P_i}$$

onde N é o número de amostras,  $P_i$  é o período.

Citamos a seguir, algumas fórmulas para o cálculo de vários tipos de Shimmer [13]:

- *Shimmer (local)*: é a média do módulo da diferença entre as amplitudes de dois períodos consecutivos dividida pela amplitude média.

$$Shimmer_{local} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_{i+1} - A_i|}{\frac{1}{N} \sum_{i=1}^N A_i}$$

onde  $N$  é o número de amostras,  $A_i$  é a amplitude.

- *Shimmer (apq3)*: é a média do módulo da diferença entre a amplitude de um período e a amplitude média em uma janela que contém esta amplitude e as amplitudes vizinhas anterior e posterior, dividida pela amplitude média.

$$Shimmer_{apq3} = \frac{\frac{1}{N-J+1} \sum_{i=1}^{N-J+1} \left| \frac{1}{J} \left( \sum_{j=1}^J A_{i+j-1} \right) - A_{i+0,5(J-1)} \right|}{\frac{1}{N} \sum_{i=1}^N A_i}$$

onde  $J$  é o tamanho da janela e a utilizada é igual a 3,  $N$  é o número de amostras,  $A_i$  é a amplitude.

- *Shimmer (apq5)*: é a média do módulo da diferença entre a amplitude de um período e a amplitude média em uma janela, que contém esta amplitude e as suas quatro amplitudes vizinhas mais próximas, ou seja, duas amplitudes anteriores e duas posteriores, dividida pela amplitude média.

$$Shimmer_{apq5} = \frac{\frac{1}{N-J+1} \sum_{i=1}^{N-J+1} \left| \frac{1}{J} \left( \sum_{j=1}^J A_{i+j-1} \right) - A_{i+0,5(J-1)} \right|}{\frac{1}{N} \sum_{i=1}^N A_i}$$

onde  $J$  é o tamanho da janela e a utilizada é igual a 5,  $N$  é o número de amostras,  $A_i$  é a amplitude.

- *Shimmer (apq11)*: é a média do módulo da diferença entre a amplitude de um período e a amplitude média em uma janela que contém esta amplitude e as suas dez amplitudes vizinhas mais próximas, ou seja, cinco amplitudes anteriores e cinco posteriores, dividida pela amplitude média.



$$Shimmer_{apq11} = \frac{\frac{1}{N-J+1} \sum_{i=1}^{N-J+1} \left| \frac{1}{J} \left( \sum_{j=1}^J A_{i+j-1} \right) - A_{i+0,5(J-1)} \right|}{\frac{1}{N} \sum_{i=1}^N A_i}$$

onde J é o tamanho da janela e a utilizada é igual a 11, N é número de amostras,  $A_i$  é a amplitude.

- *Shimmer (ddp)*: é a média do módulo da diferença entre diferenças consecutivas de amplitudes de períodos consecutivos, dividida pela amplitude média.

$$Shimmer_{ddp} = \frac{\frac{1}{N-3} \sum_{i=1}^{N-3} \left| |A_{i+3} - A_{i+2}| - |A_{i+1} - A_i| \right|}{\frac{1}{N} \sum_{i=1}^N A_i}$$

onde N é o número de amostras,  $A_i$  é o período.

### 3.1.3 *HNR - Harmonic-to-Noise Ratio*

*HNR* é a medida que quantifica a relação entre a energia da parcela do sinal de voz (sonoro) devido aos impulsos glotais, e a energia da parcela de ruído glótico, ambos filtrados pelo trato vocal. Tal ruído decorre da turbulência do fluxo de ar gerado na glote durante a fonação e pode ocorrer, por exemplo, quando a glote tem um fechamento inadequado. A presença do ruído no sinal glotal é resultado da uma vibração não-periódica das cordas vocais.

O *HNR* é obtido através da razão entre os níveis do sinal periódico (decorrente da vibração das cordas vocais) e o sinal não periódico (decorrente de ruído glótico), expresso em decibéis (dB), oferecendo um índice que relaciona o componente harmônico com o componente ruído [15]. Este parâmetro traduz de certa

forma a eficiência do processo de fonação. Uma voz saudável deve caracterizar-se por um *HNR* elevado, sendo que um valor baixo de *HNR* denotaria uma voz astenica ou disfônica.

Estudos das características espectrais mostram que idosos em perfeitas condições físicas apresentam menores espectros de ruído que idosos com condições físicas ruins e o *HNR* é uma das melhores medidas para identificar a deterioração da voz em adultos [15]. Estudos sugerem que o *HNR* apresente-se estável na idade jovem e mediana, e diminui na idade avançada [66].

Na prática, o cálculo do espectro é realizado através de técnicas eficientes como o *Fast Fourier Transform* (FFT), que consiste em um método de cálculo rápido da Transformada de Fourier. Deste modo, o espectro é calculado não como um função contínua, mas como uma amostragem desta função [53].

$$HNR = 10 \log \frac{\sum_k |H(w_k)|^2}{\sum_k |N(w_k)|^2}$$

cálculo do HNR com a interação do trato vocal.

$$GHNR = 10 \log \frac{1}{N} \sum_k \frac{|H(w_k)|^2}{|N(w_k)|^2}$$

cálculo do HNR na glote, sem a interação do trato vocal.

## 3.2 Medidas no sinal Glotal

### 3.2.1 O Sinal Glotal

A voz humana é um sinal de pressão acústica que varia com o tempo. O fluxo de ar faz com que as cordas vocais oscilem, gerando assim um trem de pulsos que pode ser descrito pela teoria Aerodinâmica Mioelástica [74]. Esta teoria refere-se aos movimentos de abertura e fechamento das cordas vocais regidos por propriedades mecânicas ao longo da laringe durante a fonação.

A estimação do sinal glotal tem sido muito estudada, e várias técnicas como: eletroglotografia, fotoglografia, ultrassom, filtragem inversa, microfone de contato e laringoscopia, foram desenvolvidas tentando obter informações a respeito da formação e modelagem desse sinal. Um método que pode ser utilizado para obter o sinal glotal é a filtragem inversa a partir do sinal de voz, que se baseia na idéia de cancelar os efeitos dos formantes na voz [63].

Podemos usar, também, a derivada do sinal glotal. Através dele consegue-se obter os instantes de máxima abertura e máximo fechamento da glote. Esses instantes mostram o momento em que as cordas vocais estão totalmente separadas ou não. A Fig.3.1 mostra um exemplo do sinal glotal e sua derivada:

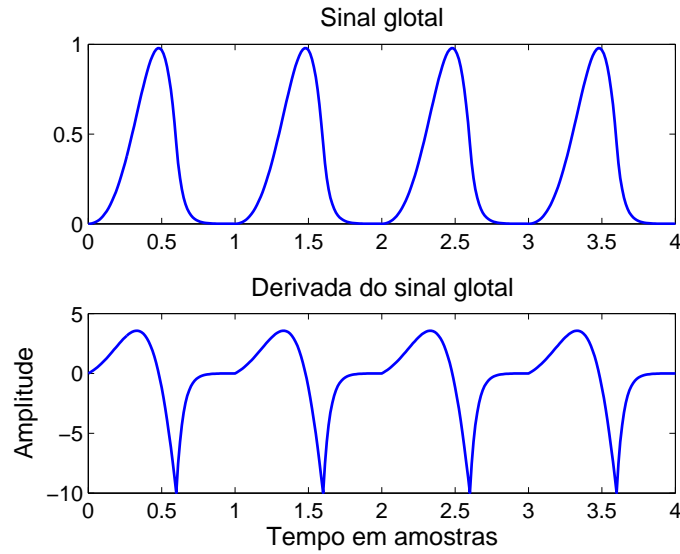


Figura 3.1: Sinal glotal e sua derivada [57].

### 3.2.2 Filtragem Inversa

O sinal glotal, após ser gerado, é amplificado pelo trato vocal e modificado pelas cavidades oral e nasal, modulado pela língua, dentes e lábios produzindo o som que ouvimos. A filtragem inversa é uma técnica bastante empregada para se estimar o sinal glotal, tendo a vantagem de partir do sinal de voz produzido. A técnica baseia-se na ideia de que o pulso glotal é obtido cancelando os efeitos dos formantes da voz, e para que isso seja possível, é necessário que o trato vocal seja modelado para ocorrer uma filtragem do sinal de voz através do inverso do trato vocal [8].

Primeiro, procede-se à filtragem de pré-ênfase que serve para atenuar as componentes de baixa frequência e incrementar as componentes de alta frequência do sinal. Desta forma diminui a instabilidade numérica e minimiza o efeito dos lábios e da glote. Depois, procede-se à filtragem inversa, propriamente dita.

Podemos utilizar os algoritmos IAIF e PSIAIF [63] têm sido usados para realizar a filtragem inversa no sinal de voz [57]. O método semi-automático de

filtragem inversa, IAIF, utiliza o sinal de voz como entrada a fim de obter, na saída, uma estimação do fluxo glotal correspondente, sendo composto por três blocos fundamentais sendo eles: análise LPC, filtragem e integração. A análise LPC é responsável pela filtragem de pré-ênfase, pela estimação do trato vocal e da contribuição glotal, definidas através da ordem de seus coeficientes, a filtragem inversa é responsável pela eliminação do trato vocal e da contribuição glotal no sinal da voz e a integração pela eliminação da radiação dos lábios. O sinal de entrada passa por um filtro passa alta para eliminar as frequências baixas que provocam flutuações na saída, servindo como entrada para o próximo passo. A frequência de corte deve ser ajustada de modo que não seja maior que a frequência fundamental do sinal de voz; caso contrário, poderá perder informações relevantes.

O método IAIF é baseado no prévio conhecimento da função de transferência do trato vocal; logo, se todo o efeito da fonte glotal for eliminado do espectro da voz, o trato vocal pode ser estimado, mais precisamente, por análise LPC ou outro método de predição linear. Essa estimação da contribuição glotal e a função de transferência do trato vocal é computada pelo algoritmo IAIF em uma estrutura que se repete duas vezes. A primeira estimativa é obtida do sinal de voz por análise LPC de ordem um e, posteriormente, eliminada por filtragem inversa. A ordem da análise LPC, se for maior que um, pode modelar os formantes [57]. A primeira estimativa da excitação glotal é obtida cancelando o efeito do trato vocal e da radiação dos lábios, por filtragem inversa e integração, respectivamente. O resultado desta primeira estrutura é o sinal glotal (excitação glotal ou contribuição glotal) que é usado como entrada da segunda estrutura a fim de estimá-lo de forma mais precisa.

No método PSIAIF, a forma do pulso glotal é obtida aplicando-se o algoritmo IAIF duas vezes, ao mesmo sinal, servindo o resultado da primeira aplicação apenas para identificar o período fundamental que será a base para o cálculo do novo janelamento, antes da segunda aplicação do IAIF.

### 3.2.3 Parâmetros do sinal glotal

A partir do sinal glotal, podemos extrair parâmetros que serão discutidos a seguir.

A Fig. 3.2 ilustra alguns parâmetros que podem ser obtidos.

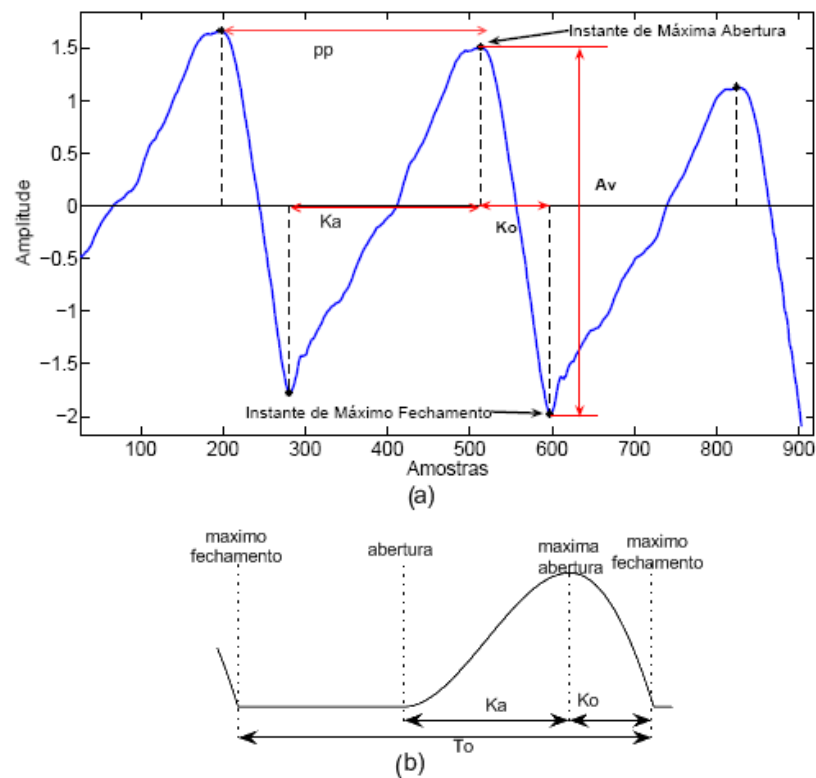


Figura 3.2: (a)Parâmetros do sinal glotal; (b)Sinal glotal [27].

- **Instantes de Máxima abertura e máximo fechamento glotal** - o instante de máxima abertura corresponde ao valor máximo de fluxo glotal, enquanto que o de máximo fechamento corresponde o valor mínimo [57];
- **$Ko$**  (duração da fase de fechamento) - corresponde ao intervalo de tempo que as cordas vocais levam da fase totalmente aberta até a fase de fechamento [57];
- **$Av$**  (amplitude de vozeamento) - corresponde à amplitude do sinal de pico a pico do sinal glotal, e está sujeito a interferências externas na hora da gravação do sinal [57];
- **$Ka$**  (duração da fase de abertura) - identifica o tempo que as cordas vocais levam da fase totalmente fechada até a fase de abertura [27];
- **$To$**  (período Fundamental) - esse parâmetro está associado ao tempo entre os ciclos de abertura e fechamento da glote. O inverso deste valor corresponde à frequência fundamental  $Fo$  [27];
- **$pp$**  (distâncias entre os instantes de máxima abertura glotal) - esse parâmetro é definido como a distância entre os valores máximos de pico do sinal glotal [63];
- **$Fa$**  (fase de abertura) - corresponde à fase do ciclo da glote em que, estando as cordas vocais separadas, o ar passa através desta -  $Fa = Ka + Ko$  [63];
- **$OQ$**  (quociente de abertura) - é definido como a relação entre a fase de abertura e o comprimento total do ciclo glotal -  $Oq = Fa/T$  [63];
- **$CIQ$**  (quociente de fechamento) - é definido como a relação entre a fase de fechamento e o comprimento total do ciclo glotal -  $CIQ = Ko/T$  [63];

- ***SQ*** (quociente de velocidade) - é definido como a relação entre a fase de abertura e a fase de fechamento -  $Sq = Ka/Ko$  [63];
- ***AQ*** (quociente da amplitude)- relaciona a amplitude do sinal glotal ( $A_v$ ) com a amplitude negativa da primeira derivada do sinal glotal;
- ***QoQ*** (quociente de abertura durante o pulso glotal) - relaciona a abertura da glote como sendo o início do sinal glotal;
- ***NAQ*** (quociente de amplitude normalizado) - relaciona a amplitude máxima do sinal glotal com o produto entre um período do sinal glotal e a amplitude negativa da primeira derivada do sinal glotal.
- ***DH12*** - é a diferença entre os dois primeiros formantes e é expresso em decibéis (dB) [63];
- ***HRF*** - expressa o fator de riqueza harmônica do sinal. Relaciona o primeiro harmônico com o somatório da energia dos harmônicos restantes [63];
- ***PSP*** - é o parâmetro que corresponde a um espectro polinomial de segunda ordem, montado no harmônico para obter uma estimativa do declive espectral.

### 3.3 O Sinal Eletroglotográfico - EGG

O Eletroglotógrafo (ver Fig.3.3) foi inventado por Fabre em 1956 [44] e se tornou um instrumento usado para investigar a área de contato entre as cordas vocais, em fonação, de uma forma não invasiva. Desta forma, podem-se captar características do sinal glotal sem a influência do trato vocal.



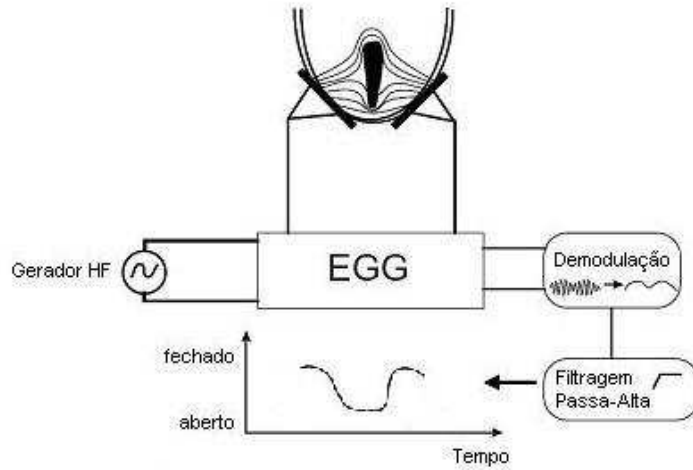


Figura 3.3: Eletroglotógrafo [43].

Dois eletrodos são colocados no pescoço e entre eles há uma variação na impedância, que ocorre com a abertura e fechamento da glote [58]. Utilizando a abertura e o fechamento da glote como fases, pode-se analisar a forma de onda do pulso glotal, uma vez que os tecidos conduzem eletricidade melhor que o ar, o que facilita a medição. Quando as cordas vocais estão fechadas, uma corrente elétrica passa através delas e, quando estão separadas, a impedância entre elas é alta. Com a mudança da impedância, temos a indicação da variação da área de contato entre as cordas vocais. O sinal obtido pelo eletroglotógrafo é denominado *EGG*.

Os parâmetros obtidos pelo sinal *EGG* expressam os instantes de máxima abertura e máximo fechamento, enquanto que na derivada do sinal *EGG* (indicada por *DEGG*), indicam o exato momento de abertura e fechamento glotal [43]. Por ser o fechamento glotal muito rápido, a forma de onda *DEGG* facilita a visualização precisa desse movimento.

Fazendo uma análise da forma de onda pode-se dizer que o tempo de duração entre dois fechamentos glotais corresponde ao período fundamental (o inverso seria a frequência fundamental) da voz. Na Fig.3.4 vemos a forma de onda EGG e DEGG descrevendo as fases de abertura e fechamento da glote.

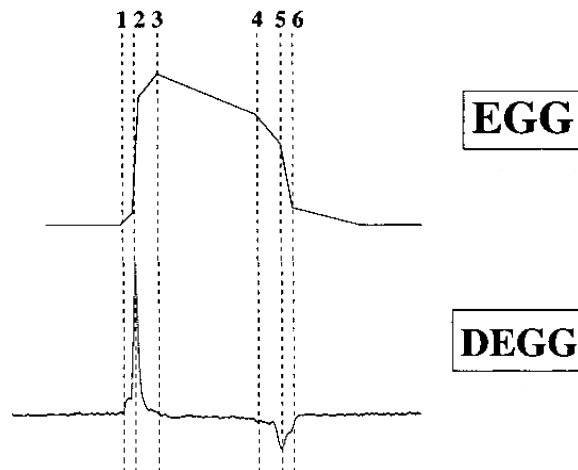


Figura 3.4: Forma de onda EGG e DEGG [44].

A seguir a descrição das fases medidas pelos sinais EGG e DEGG.

- (1)-(3) fase de fechamento glotal;
- (3)-(4) fase glote fechada;
- (4)-(6) fase de abertura;
- (6)-(1) fase glote aberta;

Na Fig.3.5 podemos visualizar todas as fases de abertura e fechamento glotal com suas respectivas formas de onda EGG e DEGG [40].

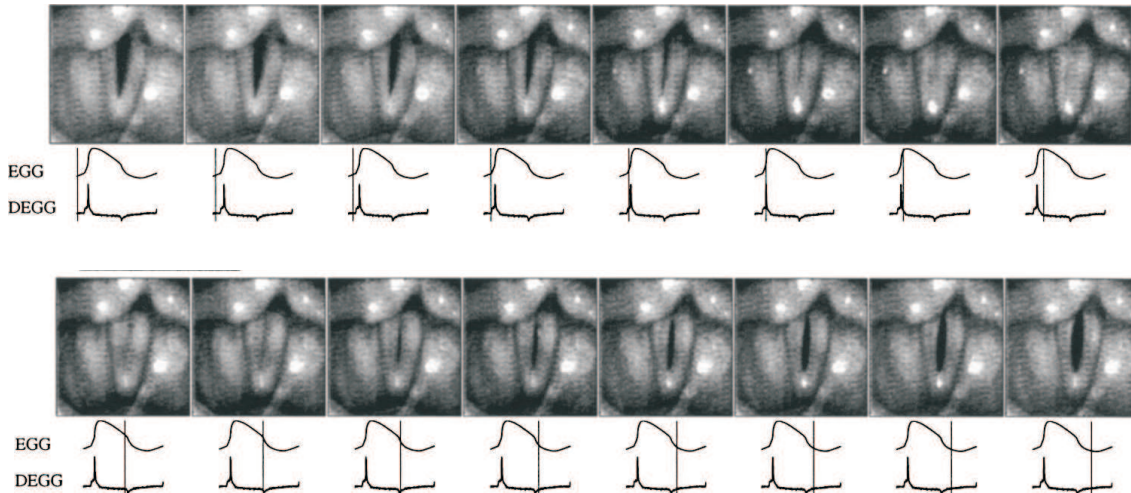


Figura 3.5: Fases de fechamento e abertura glotal [43].

Existem diversos tipos de eletroglotógrafos utilizados para fazer a medição do sinal eletroglotográfico. Neste estudo, será utilizado um equipamento da Glottal Enterprises modelo EG2-PCX, que contém uma interface USB para ser ligada diretamente no computador, eletrodos para serem fixados no pescoço, microfone *shot gun*, LS-1 (*larynx simulator*), C-1 (*compensator*) e cabos diversos.

Para a gravação do sinal eletroglotográfico utilizamos o programa *SoundForge* (<http://www.sonycreativesoftware.com/soundforge>).

Analisando partes do sistema:

- Interface - interface do sistema que possibilita a conexão entre os eletrodos com o computador;
- Eletrodos - são microfones que ficam em contato com o pescoço, chamados de *throat microphone* (um disco piezoelétrico cerâmico, geralmente utilizado na captação de instrumentos acústicos);
- LS-1 (*Larynx Simulator*) - Simulador da Laringe;

- C-1 (*Compensator Phase*) - Corretor de fase.

O eletroglotógrafo possui equipamentos externos para fazer a calibração (ver Fig.3.6). O LS-1 é um simulador de laringe e o C-1 um compensador de fase que são usados para simular as pequenas variações de resistência do pescoço e na verificação de desempenho do equipamento respectivamente. O C-1 foi desenvolvido para diminuir as distorções de fase quando inseridas pela filtragem passa alta.



Figura 3.6: Microfone - LS1 - C1 [57].

## 3.4 Os programas Aparat e Praat

### 3.4.1 *Aparat*

O *Aparat* [48] foi desenvolvido na Universidade de Tecnologia de Helsinki - Laboratório de Acústica e Processamento de Sinal de Áudio (<http://www.acoustics.hut.fi>) e pode ser definido como um conjunto de ferramentas e interface gráfica utilizada para análise de sinais através da filtragem inversa do sinal glotal e extração de características do sinal glotal. O *Aparat* funciona como um *toolbox* do programa Matlab<sup>®</sup> sendo necessário que este seja da versão 6.5 em diante. Através do *Aparat*, podemos obter, também, informações sobre o sinal de voz, além do sinal glotal.

Diferente partes do sinal podem ser selecionados e podemos ter acesso as suas respectivas análises e formas de onda conforme ilustrado na Fig.3.7. Todos os dados referentes ao sinal selecionado podem ser salvos, no padrão Matlab<sup>®</sup> (.mat) [48].

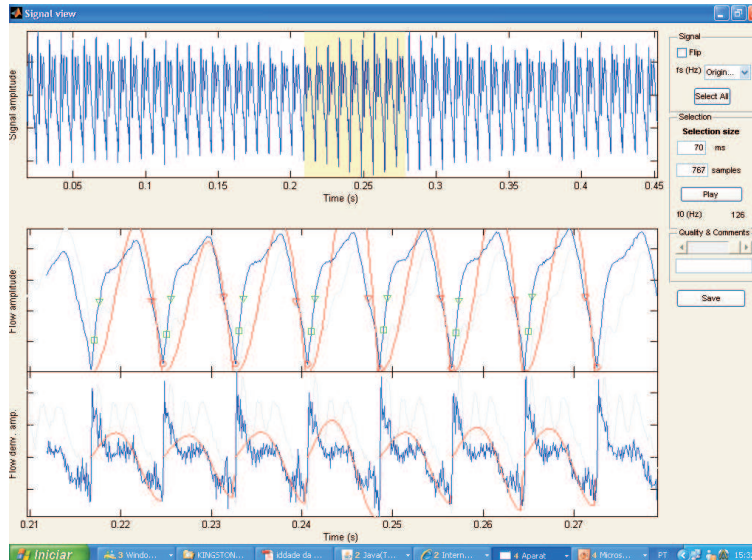


Figura 3.7: Exemplo de forma de onda do sinal no pacote Aparat [48].

Uma janela chamada de *parameter view* nos permite visualizar os gráficos de cada parâmetro selecionado (ver Fig.3.8). Podemos obter diferentes parâmetros na base de tempo, frequência e modelo LF (Liljencrants and Fant). Este modelo baseia-se na derivada da velocidade do volume glotal e permite a associação dos parâmetros do modelo com características físicas, obtendo maior qualidade da voz [51].

O programa **Aparat** está melhor detalhado no apêndice.

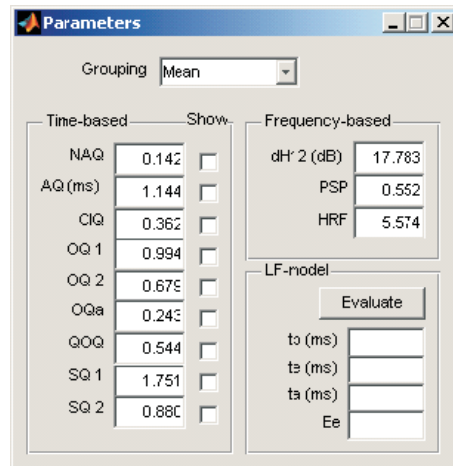


Figura 3.8: Exemplo de janela “parâmetros do sinal” no pacote Aparat [48].

### 3.4.2 Praat

O Praat é um programa utilizado para análise de voz, desenvolvido por **Paul Boerma e David Weenink do Institute of Phonetic Sciences** - Universidade de Amsterdã, e pode ser adquirido na internet. É um *software* livre que permite sua aplicação em diferentes tipos de plataformas [13].

Todas as análises feitas aparecem ao longo da forma de onda e podem ser mostradas simultaneamente. A Fig.3.9 mostra um exemplo de janela de visualização do sinal de voz: o sinal e seu espectrograma. A partir desses sinais, podemos obter:

- *Intensity* - permite o controle das definições da intensidade.
- *Formant* - permite o controle das definições dos formantes.
- *Pulses* - permite a extração de informação específica relacionada com a qualidade de voz (*jitter* e *shimmer*).

No apêndice, há uma descrição detalhada do programa **Praat**.

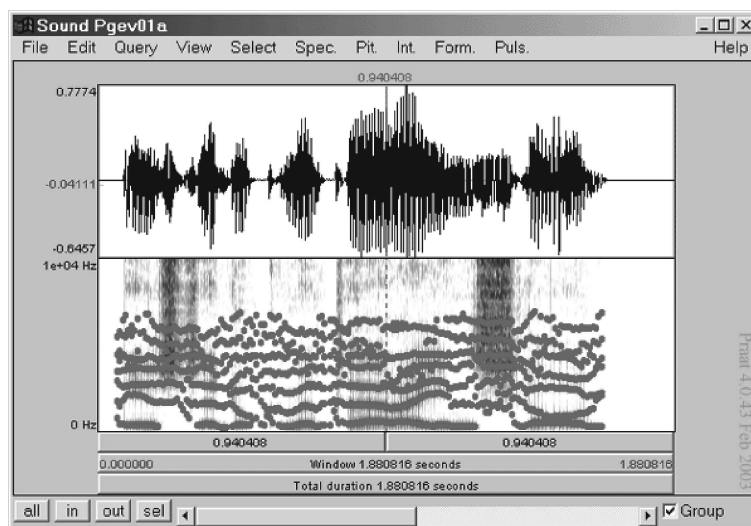


Figura 3.9: Uma janela de visualização do Praat.

## Capítulo 4

# Gravação dos sinais de voz e extração de características

Neste capítulo, serão descritos os procedimentos usados para a criação de uma base de dados e para a extração de características (ou parâmetros) dos sinais de voz, dos sinais do eletroglotógrafo e dos sinais glotais. Estaremos analisando grupos de faixas etárias distintas, sendo que, o ideal seria analisar as mesmas pessoas com diferentes idades, mas isso não foi possível.

Uma grande base de dados foi criada com gravações de sinais de várias vogais, sustentadas e concatenadas, e frase. Gravamos locutores de diversas faixas etárias e com certo grau de parentesco (mãe, filha e neta), separados da seguinte forma: um grupo de moradores de casas de repouso, um grupo da casa Retiro dos Artistas e um grupo de voluntários (gravados em suas próprias casas e no consultório da fonoaudióloga). O grupo de moradores das casas de repouso foi constituído somente de locutores do sexo feminino (os locutores do sexo masculino eram portadores de doenças avançadas, impossibilitando as gravações) e todas as gravações para esse grupo foram somente de sinais de voz. As gravações dos voluntários e da casa Retiro



dos Artistas foram realizadas com homens e mulheres, e foram tanto de sinais de voz como de sinais eletroglotográficos.

Após recolhermos todos os dados, decidimos nos concentrar em sinais de uma única vogal e proceder as análises com essa vogal. Todos os locutores autorizaram a gravação de suas vozes.

A Tab.4.1 mostra um exemplo dos valores obtidos da análise da voz de uma senhora de 99 anos utilizando o *Praat*:

Tabela 4.1: *Jitter/Shimmer/HNR* - Locutora 99 anos

Locutora 99 anos					
Vogais	A	E	I	O	U
Jitter(local)	0.744%	1.668%	0.681%	0.591%	0.336%
Jitter(local,abs)	33.3E-6s	96.6E-6s	32.1E-6s	29.7E-6s	15.6E-6s
Jitter(rap)	0.300%	0.846%	0.328%	0.281%	0.148%
Jitter(ppq5)	0.512%	1.259%	0.361%	0.276%	0.175%
Jitter(ddp)	0.899%	2.537%	0.985%	0.843%	0.445%
Shimmer(local)	6.223%	13.730%	4.177%	1.946%	3.199%
Shimmer(local, Db)	0.555db	1.123db	0.440db	0.251db	0.319db
Shimmer(apq3)	2.858%	6.467%	2.381%	0.916%	1.811%
Shimmer(apq5)	4.112%	10.419%	2.575%	1.094%	2.000%
Shimmer(apq11)	6.074%	7.620%	2.875%	1.525%	2.133%
Shimmer(dda)	8.575%	19.402%	7.142%	2.749%	5.432%
Harm.-to-noise ratio	15.336dB	9.499dB	14.640dB	21.235dB	25.948dB

Para as gravações, solicitamos que vogais sustentadas e frases fossem produzidas em tom e intensidade confortáveis para o locutor, por um período de 5 segundos, aproximadamente. Todos os locutores estavam sentados corretamente com o microfone posicionado a uma distância de 20cm da boca. Contamos com o auxílio de uma fonoaudióloga para a gravação das vogais sustentadas e frases de 103 locutores, para certificar sobre o estado de saúde da voz de cada um. Com isso, 29 pessoas foram

descartadas.

Para a captação das vozes, utilizamos um sistema com *hardware* dedicado para trabalhar com o programa *Pro Tools* e, para a interface do eletroglotógrafo o programa *SoundForge*. O programa *SoundForge* é fabricado pela Sony, tendo como função a gravação do áudio e sua correspondente edição.

O *hardware* utilizado foi um Laptop PC Vaio com 512 MRAM, processador Intel Pentium 4 - 1.18GHz com interface *MBox mini* (ver Fig.4.1) ligada no computador via USB e um microfone AKG 421 dinâmico. Esse sistema integrado denomina-se *workstation* e permite fazer várias gravações em diferentes canais facilitando, caso necessário, a edição do sinal.



Figura 4.1: Interface de gravação e microfone.

Para a análise das vozes, após a captação, utilizamos o programa *Praat* para a extração dos parâmetros dos sinais de voz e eletroglotográfico. Os parâmetros extraídos foram: Frequência Fundamental, *Jitter*, *Shimmer* e *HNR* (*Harmonic-to-Noise Ratio*).

## **Grupo I - Sinal de voz, sinal eletroglotográfico e sinal glotal**

Os locutores que tiveram a gravação dos sinais de voz e, também, sinais eletroglotográficos, foram divididos em faixas etárias da seguinte forma:

### 1. Locutores Masculinos

- (a) **Faixa etária 1** - 5 locutores de 15 a 49 anos
- (b) **Faixa etária 2** - 6 locutores de 50 a 100 anos

### 2. Locutores Femininos

- (a) **Faixa etária 1** - 5 locutores de 15 a 49 anos
- (b) **Faixa etária 2** - 4 locutores de 50 a 100 anos

Para esse estudo, foram consideradas as análises dos sinais de voz, eletroglotográfico e glotal (que foram obtidas por filtragem inversa usando o programa *Aparat*). Os parâmetros foram extraídos para cada caso e efetuada uma comparação entre eles.

De todas as vozes disponíveis no grupo I, tivemos as seguintes situações:

- 4 vozes obtidas em casas de repouso;
- 16 vozes de voluntários em suas residências;

## **Grupo II - Sinal de voz e sinal glotal**

Com os locutores que tiveram a gravação somente do sinal de voz, utilizamos o programa *Aparat* para obter o sinal glotal e extrair os parâmetros. Dividimos as faixas etárias da seguinte forma:

## 1. Locutores Masculinos

- (a) **Faixa etária 1** - 19 locutores de 15 a 30 anos
- (b) **Faixa etária 2** - 19 locutores de 31 a 59 anos
- (c) **Faixa etária 3** - 14 locutores de 60 anos em diante

## 2. Locutores Femininos

- (a) **Faixa etária 1** - 15 locutores de 15 a 30 anos
- (b) **Faixa etária 2** - 39 locutores de 31 a 59 anos
- (c) **Faixa etária 3** - 18 locutores de 60 anos em diante

Como no estudo anterior, também obtivemos os sinais glotais por filtragem inversa, e foi realizada avaliação da variação dos valores entre as faixas etárias.

De todas as vozes disponíveis do grupo II, tivemos as seguintes situações:

- 16 vozes obtidas em casas de repouso;
- 10 vozes obtidas na casa Retiro dos Artistas;
- 30 vozes de voluntários em suas residências;
- 74 vozes de voluntários no consultório da fonoaudióloga.

## 4.1 Comparação de sinais com duas faixas etárias (Grupo I)

### 4.1.1 Primeiro caso: Sinais de voz x Sinais do eletroglotógrafo

Para essa comparação, usamos apenas os locutores do grupo I, com o objetivo de comparar os valores dos parâmetros usando o microfone e o eletroglotógrafo. Os resultados das análises, assim como das comparações, serão apresentados no decorrer desta seção. As estatísticas utilizadas para as comparações foram: média, mediana, desvio padrão, variância e curtose dos parâmetros extraídos. Para as análises das vozes e confecção dos gráficos utilizamos, além do Praat, o programa Matlab<sup>®</sup>.

A média e a mediana são medidas de posição ou de tendência central, o desvio padrão e a variância são medidas de dispersão ou afastamento e indicam a diferença entre os valores de uma série e sua média aritmética, a curtose indica o grau de concentração das frequências absolutas sob a curva de frequência tomando como base a distribuição normal [23]. Da análise, extraímos os parâmetros: Frequência Fundamental, *Jitter*, *Shimmer* e *HNR*.

#### **Frequência Fundamental - $F_0$**

Na análise das vozes, a média da Frequência Fundamental para os homens, utilizando o microfone, foi de 143,09Hz com um desvio padrão de 33,02Hz, enquanto que com o eletroglotógrafo foi de 140,64Hz com um desvio padrão de 31,75Hz. Nas vozes das mulheres, para a captação em microfone, encontramos uma média de 174,56Hz com um desvio padrão de 35,61Hz, enquanto que com o eletroglotógrafo foi 180,11Hz com um desvio padrão de 40,87Hz.

As Tab.4.2, Tab.4.3, Tab.4.4 e Tab.4.5, mostram as estatísticas obtidas para os homens e mulheres, da Frequência Fundamental, referente à emissão da vogal /e/.

Tabela 4.2: Frequência Fundamental (Hz) - Locutores masculinos com microfone.

<b>Frequência Fundamental Masculino - Microfone</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	143,09	133,00	33,02	1090,49	-1,07
15-49 anos (N1=5)	122,60	123,00	9,07	82,30	-2,71
50-100 anos (N2=6)	160,17	171,00	36,68	1345,37	2,89

Tabela 4.3: Frequência Fundamental (Hz) - Locutores masculinos com eletroglotógrafo.

<b>Frequência Fundamental Masculino - Eletroglotógrafo</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	140,64	133,00	31,75	1008,05	-0,17
15-49 anos (N1=5)	122,40	123,00	8,88	78,80	-2,50
50-100 anos (N2=6)	155,83	156,00	36,65	1343,37	-0,65

Tabela 4.4: Frequência Fundamental(Hz) - Locutores femininos com microfone.

<b>Frequência Fundamental Feminino - Microfone</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	174,56	180,00	35,61	1267,78	1,28
15-49 anos (N1=5)	177,80	181,00	46,01	2116,70	2,62
50-100 anos (N2=4)	170,50	170,00	22,78	519,00	-0,26

Tabela 4.5: Frequência Fundamental(Hz) - Locutores femininos com eletroglotógrafo.

Frequência Fundamental Feminino - Eletroglotógrafo					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	180,11	181,00	40,87	1670,61	0,55
15-49 anos (N1=5)	177,40	181,00	46,84	2194,30	2,70
50-100 anos (N2=4)	183,50	186,50	38,75	1501,67	-2,21

As Fig.4.2 ( $F_0$  Masculino) e Fig.4.3 ( $F_0$  Feminino) mostram as variações da Frequência Fundamental em relação aos grupos de locutores, divididos em suas respectivas faixas etárias. Os gráficos foram obtidos usando o recurso *boxplot* do programa Matlab<sup>®</sup>. A linha em vermelho indica a mediana do grupo de valores considerados.

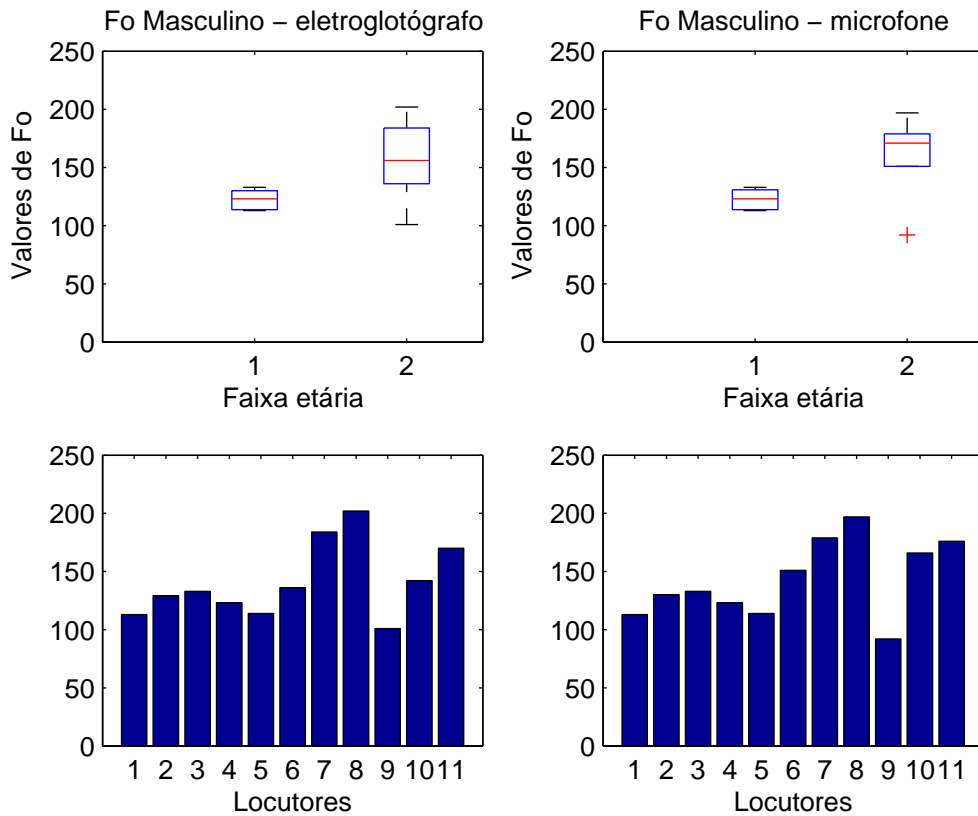


Figura 4.2: Frequência fundamental(Hz) - Masculino.

De acordo com a literatura, o valor da Frequência Fundamental tende a aumentar para os homens e diminuir, ou até manter-se, para as mulheres [65]. O resultado do sinal de voz e eletroglotográfico para os homens está dentro das expectativas. Para as mulheres, tivemos pouca variação da Frequência Fundamental na análise do sinal eletroglotográfico, e uma pequena queda para o sinal de voz.

Obtemos a mesma informação da Frequência Fundamental, utilizando tanto o eletroglotógrafo quanto o microfone.



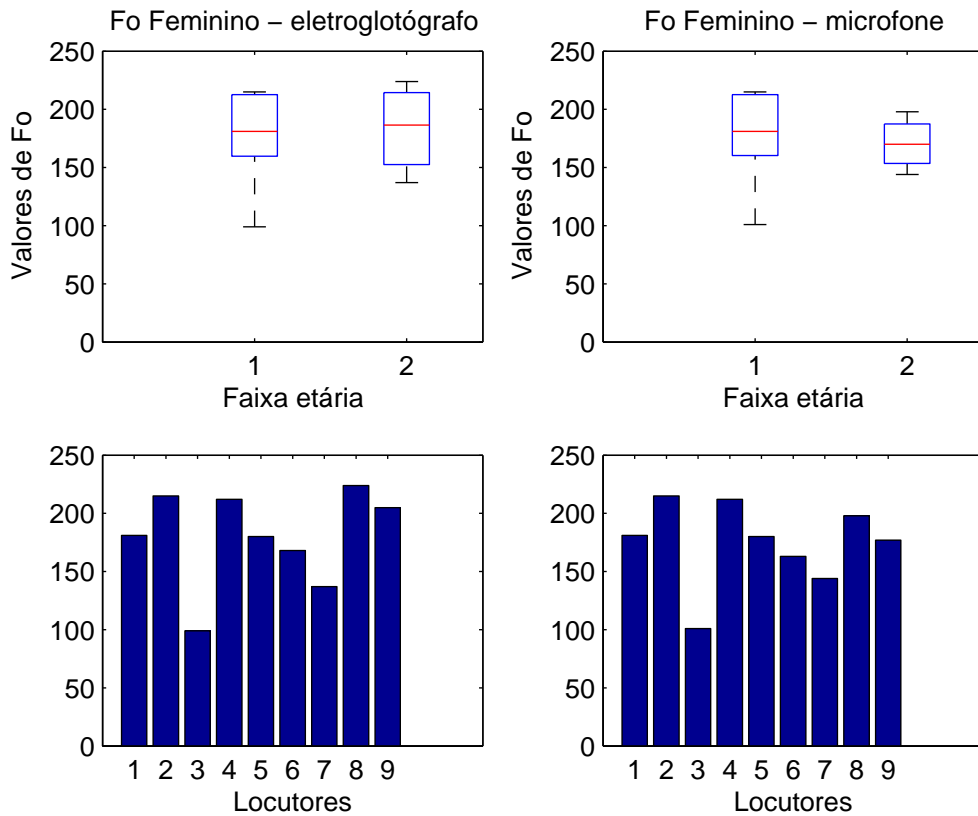


Figura 4.3: Frequência fundamental(Hz) - Feminino.

### *Jitter*

No programa *Praat* obtemos vários valores de *Jitter*, mas para a nossa análise escolhemos utilizar o *Jitter Local* (utiliza somente os dois períodos consecutivos). Na análise das vozes, a média do *Jitter* para os homens foi 0,65% com um desvio padrão de 0,64% e 1,44% com um desvio padrão de 1,74%, referentes à captação pelo microfone e o eletroglotógrafo respectivamente. Nas vozes das mulheres, para a captação em microfone, encontramos uma média de 0,62% com um desvio padrão de 0,50%, enquanto que com o eletroglotógrafo foi 1,16% com um desvio padrão de 0,90%.

As Tab.4.6, Tab.4.7, Tab.4.8 e Tab.4.9, mostram as estatísticas de *Jitter* obtidas para os homens e as mulheres referentes à vogal /e/:

Tabela 4.6: *Jitter* - Locutores masculinos com microfone.

<b><i>Jitter</i> Masculino - Microfone</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	0,65	0,50	0,64	0,41	6,32
15-49 anos (N1=5)	0,36	0,30	0,13	0,02	-2,41
50-100 anos (N2=6)	0,90	0,55	0,80	0,64	2,80

Tabela 4.7: *Jitter* - Locutores masculinos com eletroglotógrafo.

<b><i>Jitter</i> Masculino - Eletroglotógrafo</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	1,44	0,80	1,74	3,02	3,71
15-49 anos (N1=5)	0,38	0,30	0,13	0,02	2,66
50-100 anos (N2=6)	2,32	1,45	1,99	3,98	0,90

Tabela 4.8: *Jitter* - Locutores femininos com microfone.

<b><i>Jitter</i> Feminino - Microfone</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	0,62	0,40	0,50	0,25	6,80
15-49 anos (N1=5)	0,44	0,44	0,15	0,02	3,72
50-100 anos (N2=4)	0,85	0,60	0,72	0,52	3,01

Tabela 4.9: *Jitter* - Locutores femininos com eletroglotógrafo.

<i>Jitter</i> Feminino - Eletroglotógrafo					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	1,16	0,80	0,90	0,82	0,70
15-49 anos (N1=5)	0,56	0,50	0,15	0,02	1,46
50-100 anos (N2=4)	1,90	1,90	0,90	0,81	1,26

As Fig.4.4 (*Jitter* Masculino) e Fig.4.5 (*Jitter* Feminino) mostram as variações do *Jitter* em relação aos grupos de locutores, divididos em suas respectivas faixas etárias.

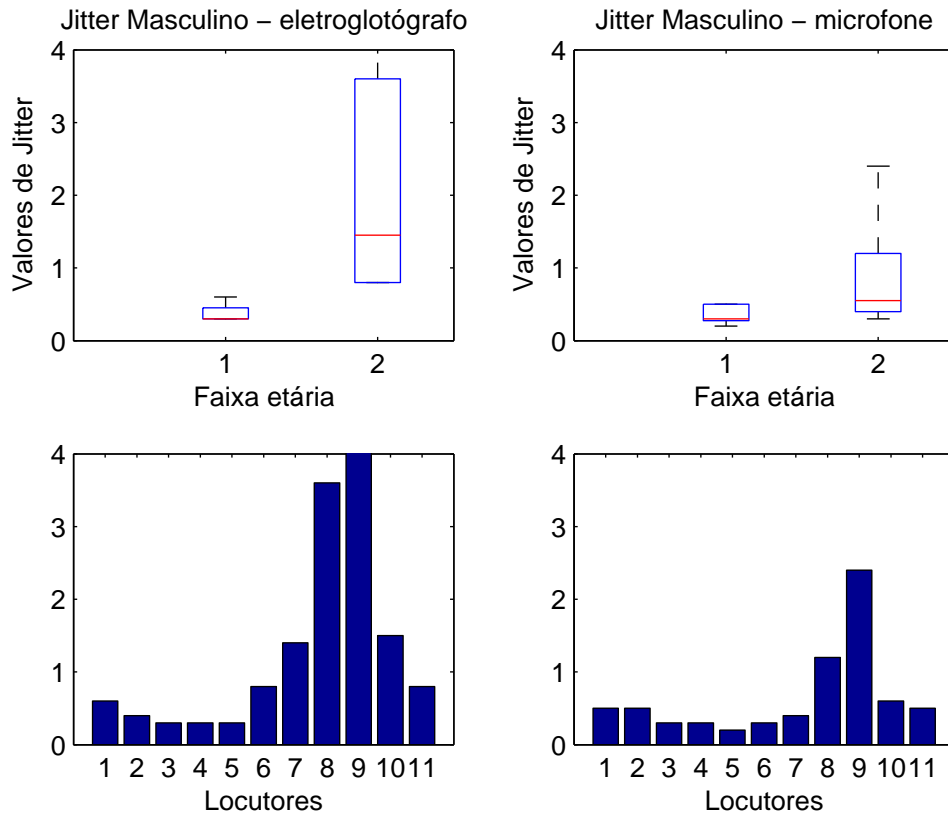


Figura 4.4: *Jitter* - Masculino.

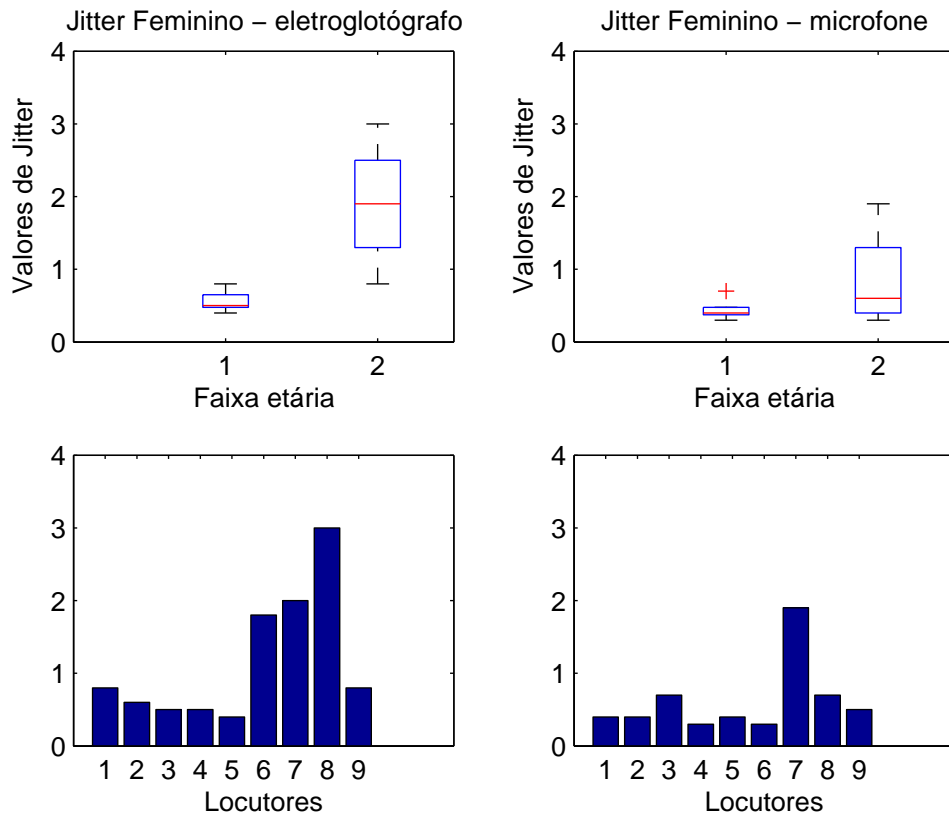


Figura 4.5: Jitter - Feminino.

Os valores de *Jitter* tendem a aumentar com o avanço da idade devido aos problemas de ossificação das cartilagens, degeneração dos músculos da laringe e mudanças respiratórias [26] [52] [65]. Esses fatores podem interferir na periodicidade do sinal. Das análises, os valores de *Jitter* para os homens e para mulheres sofreram um aumento da faixa etária 1 para a faixa etária 2, em ambos os sinais (voz e eletroglotógrafo). Observamos uma grande variação do valor de Jitter na faixa etária 2 em ambos os sexos. Para os homens, o valor aumentou passando de 0,90% para 2,32% e para as mulheres 0,85% para 1,90%. Apesar das variações, podemos concluir que o *Jitter* é um parâmetro que contribui para a identificação entre as faixas etárias, independente do tipo de gravação.

## *Shimmer*

Na análise das vozes, a média do *Shimmer* para os homens foi 11,00% com um desvio padrão de 7,54% e 11,27% com um desvio padrão de 9,88%, referentes à captação pelo microfone e o eletroglotógrafo, respectivamente. Nas vozes femininas, para a captação em microfone, encontramos uma média de 11,70% com um desvio padrão de 5,56%, enquanto que com o eletroglotógrafo foi 11,79% com um desvio padrão de 8,03%.

As Tab.4.10, Tab.4.11, Tab.4.12 e Tab.4.13 mostram as estatísticas do *Shimmer* obtidas para os homens e mulheres referentes à vogal /e/:

Tabela 4.10: *Shimmer* - Locutores masculinos com microfone.

<i>Shimmer</i> Masculino - Microfone					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	11,00	9,00	7,54	56,80	-1,42
15-49 anos (N1=5)	16,40	20,00	6,43	41,30	-2,94
50-100 anos (N2=6)	6,50	4,00	5,21	27,10	2,05

Tabela 4.11: *Shimmer* - Locutores masculinos com eletroglotógrafo.

<i>Shimmer</i> Masculino - Eletroglotógrafo					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	11,27	14,00	9,88	97,62	-1,44
15-49 anos (N1=5)	1,80	2,00	0,45	0,20	5,00
50-100 anos (N2=6)	19,17	18,50	5,53	30,57	-0,49

Tabela 4.12: *Shimmer* - Locutores femininos com microfone.

<b><i>Shimmer</i> Feminino - Microfone</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	11,70	11,30	5,56	30,92	-1,09
15-49 anos (N1=5)	15,78	15,40	3,56	12,67	0,56
50-100 anos (N2=4)	6,60	6,55	1,77	3,13	1,12

Tabela 4.13: *Shimmer* - Locutores femininos com eletroglotógrafo.

<b><i>Shimmer</i> Feminino - Eletroglotógrafo</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	11,79	9,40	8,03	64,54	-1,52
15-49 anos (N1=5)	6,09	5,30	2,67	7,13	-2,21
50-100 anos (N2=4)	18,98	21,15	6,22	38,65	2,97

As Fig.4.6 e Fig.4.7 mostram as variações do *Shimmer* em relação aos grupos de locutores, divididos em suas respectivas faixas etárias.

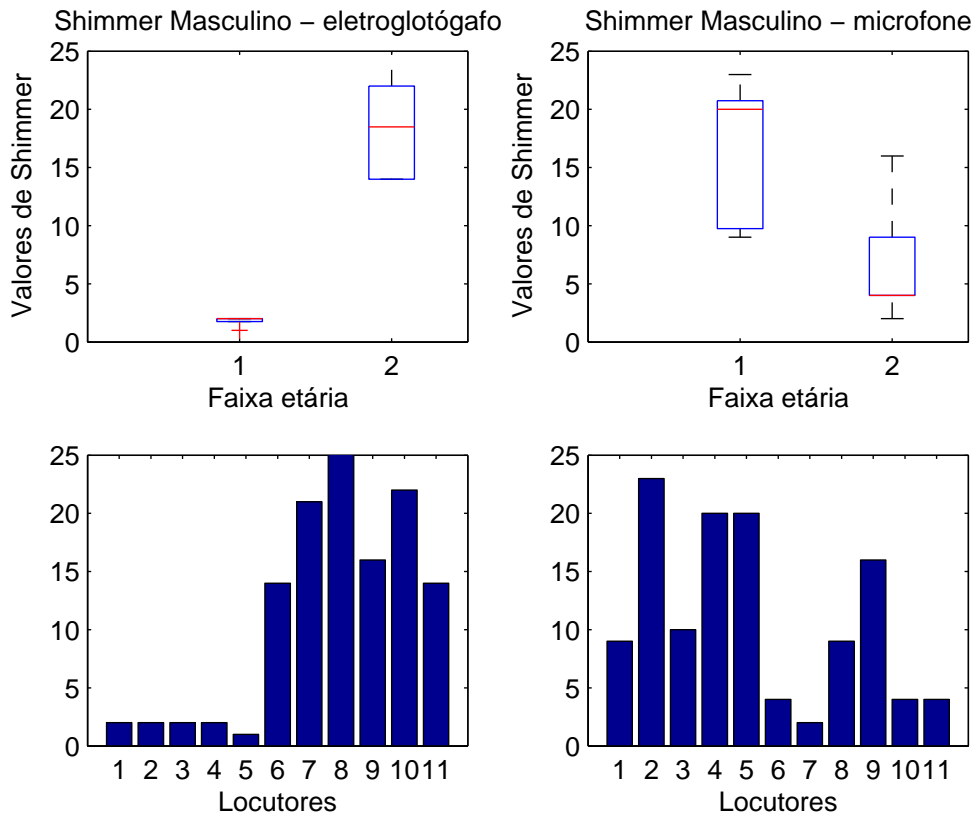


Figura 4.6: Shimmer - Masculino.

Segundo a literatura, os valores de *Shimmer* para os homens e para as mulheres tendem a aumentar na terceira idade. Fatores físicos podem contribuir para a mudança do *Shimmer* ao longo da idade [52] [65]. Para as mulheres, o valor de *Shimmer* tende a uma queda durante a fase hormonal, voltando a normalidade após esse período [11] [65]. O valor de *Shimmer* para os homens, na faixa etária 1, é maior em relação à faixa etária 2 captado pelo microfone, enquanto que utilizando o eletroglotógrafo ocorre o fenômeno contrário. O mesmo ocorreu para as mulheres. Os valores das faixas etárias 1 e 2 apresentaram uma enorme discrepância entre as análises do microfone e eletroglotógrafo. Esses resultados não parecem confiáveis.

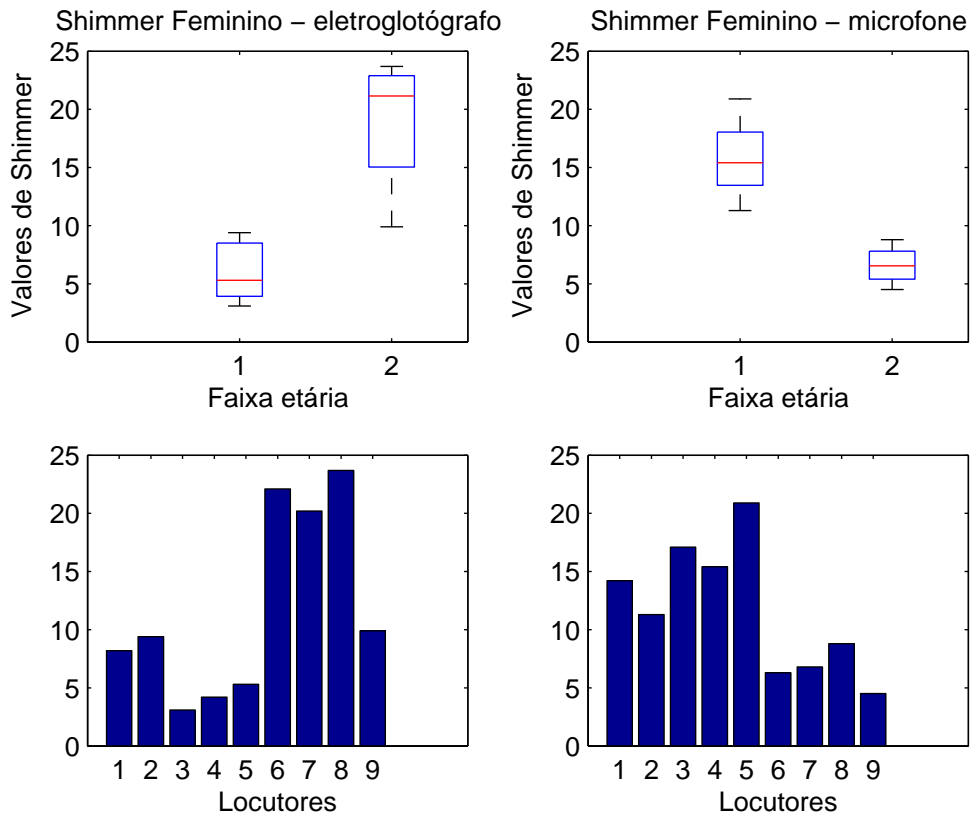


Figura 4.7: Shimmer - Feminino.

Confirmamos o que diz a literatura analisando o gráfico do eletroglotógrafo, sendo o resultado do gráfico do microfone, contrário. Será necessário fazer uma análise com um espaçamento maior entre as faixas etárias, e uma maior verificação da utilização do programa Praat para essas análises.

### ***HNR***

O *HNR* (*Harmonic to Noise Ratio*) é uma medida que indica a presença de ruído no sinal de voz, determinando a qualidade da voz. O fechamento inadequado das cordas vocais pode fazer com que ocorra um excessivo fluxo de ar que, ao passar pela glote, cria turbulência na produção do sinal. Esse ruído pode ser atribuído à não periodicidade do sinal.



Na análise das vozes, a média do *HNR* para os homens foi 14,66dB com um desvio padrão de 5,35dB e 20,29dB com um desvio padrão de 11,53dB, referentes à captação pelo microfone e o eletroglotógrafo, respectivamente. Nas vozes femininas, para a captação feita com o microfone, encontramos uma média de 14,55dB com um desvio padrão de 3,18dB, enquanto que com o eletroglotógrafo foi 16,66dB com um desvio padrão de 7,18dB.

As Tab.4.14, Tab.4.15, Tab.4.16 e Tab.4.17 mostram as estatísticas obtidas para os homens e mulheres do *HNR* referente à vogal /e/:

Tabela 4.14: *HNR* (dB) - Locutores masculinos com microfone.

<b><i>HNR</i> Masculino - Microfone</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	14,66	13,37	5,35	28,59	-0,63
15-49 anos (N1=5)	12,78	12,58	3,12	9,72	1,16
50-100 anos (N2=6)	16,23	17,96	6,55	42,89	-1,03

Tabela 4.15: *HNR* (dB) - Locutores masculinos com eletroglotógrafo.

<b><i>HNR</i> Masculino - Eletroglotógrafo</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=11)	20,29	14,97	11,53	132,92	-1,78
15-49 anos (N1=5)	31,74	31,32	3,41	11,64	2,11
50-100 anos (N2=6)	10,75	11,32	4,02	16,13	0,20

Tabela 4.16: *HNR* (dB) - Locutores femininos com microfone.

<b><i>HNR</i> Feminino - Microfone</b>					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	14,55	14,66	3,18	10,13	-0,73
15-49 anos (N1=5)	13,02	13,79	3,09	9,53	-1,91
50-100 anos (N2=4)	16,46	16,77	2,35	5,54	-3,31

Tabela 4.17: *HNR* (dB) - Locutores femininos com eletroglotógrafo.

<i>HNR</i> Feminino - Eletroglotógrafo					
Locutores	Média	Mediana	D.Padrão	Variância	Curtose
Todos (N1+N2=9)	16,66	17,85	7,18	51,59	0,13
15-49 anos (N1=5)	21,43	22,60	3,34	11,17	-2,99
50-100 anos (N2=4)	10,71	11,07	6,14	37,70	1,43

As Fig.4.8 e Fig.4.9 mostram as variações do *HNR* em relação aos grupos de locutores, divididos em suas respectivas faixas etárias.

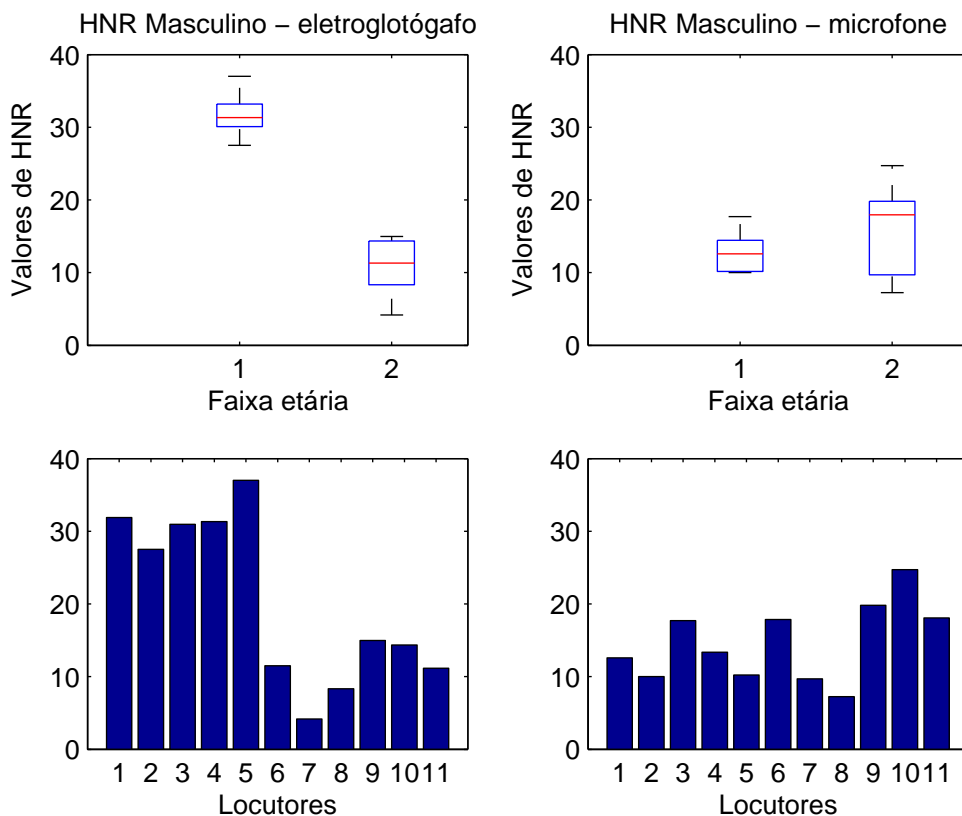


Figura 4.8: *HNR* (dB)- Masculino.

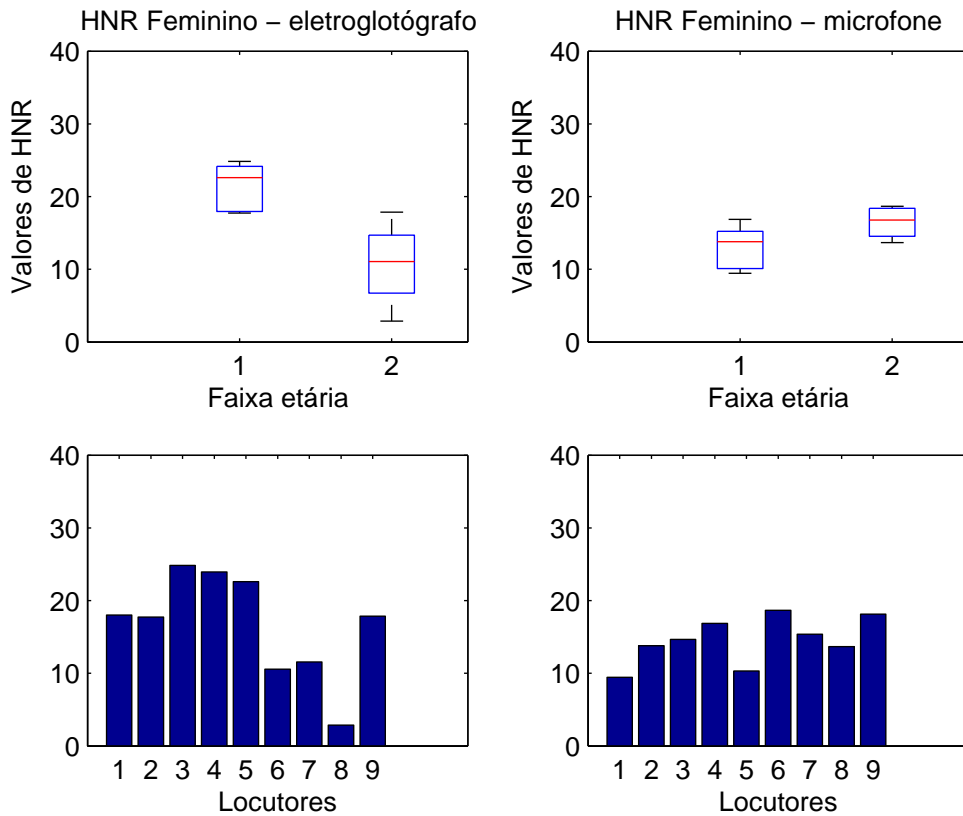


Figura 4.9: HNR (dB)- Feminino.

Com a idade, existe uma grande tendência de o nível de ruído ser alto, logo o *HNR* deveria ser baixo. Como vimos nos gráficos, os valores para *HNR* dos homens da faixa etária 1 é menor em relação à faixa etária 2, captado pelo microfone, enquanto que utilizando o eletroglotógrafo o valor na faixa etária 1 é consideravelmente maior em relação à faixa etária 2. O valor de *HNR* para as mulheres na faixa etária 1 aumenta em relação à faixa etária 2 captado pelo microfone, enquanto que utilizando o eletroglotógrafo o valor na faixa etária 1 é menor em relação à faixa etária 2. Sendo o eletroglotógrafo um instrumento usado para investigar a área de contato entre as cordas vocais em fonação, e o *HNR* uma medida que pode caracterizar problemas de fechamento da glote, acreditamos que o gráfico do eletroglotógrafo expressa melhor a medida.

Com relação ao gráfico do microfone, os valores com o microfone não parecem confiáveis, possivelmente o valor da faixa etária 2 não caracteriza um idoso propriamente dito. Será necessária uma análise com espaçamento maior entre as faixas etárias.

#### 4.1.2 Segundo caso: Sinais glotais x Sinais do Eletroglotógrafo

Apresentaremos, nesta subseção, uma comparação entre os parâmetros do sinal glotal, utilizando filtragem inversa dos sinais de voz e os parâmetros dos sinais eletroglotográficos. As medidas foram realizadas com o grupo I.

##### Frequência Fundamental:

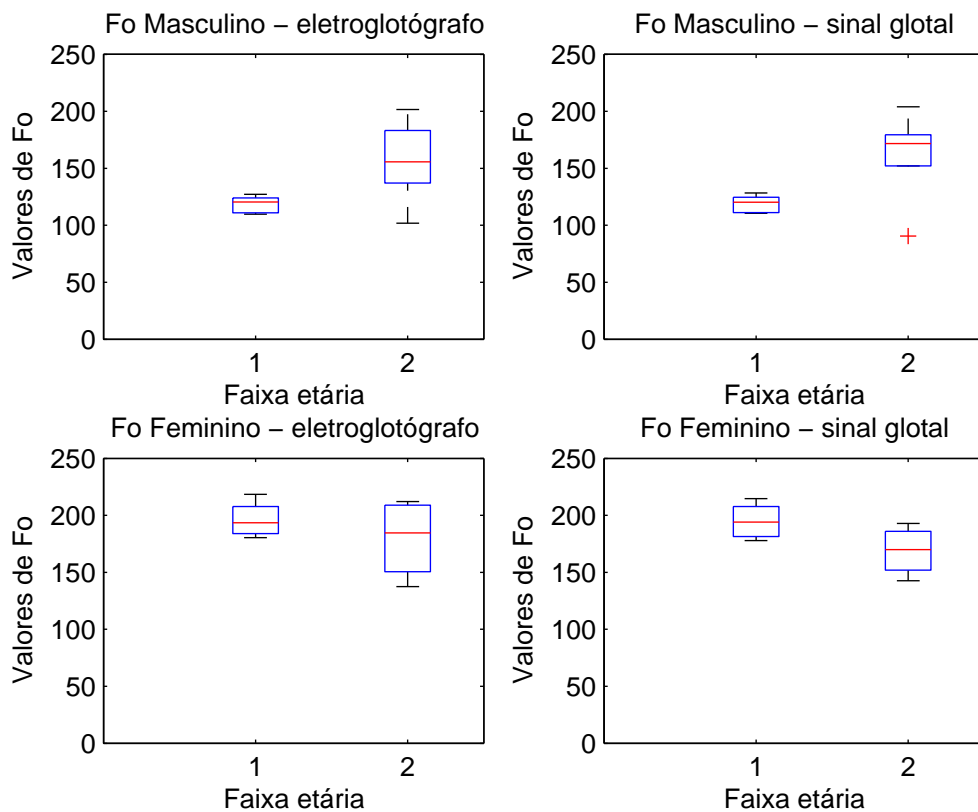


Figura 4.10: F0 (Hz)- Sinal Eletroglotógrafo - Sinal glotal.

A Fig. 4.10 mostra as variações da Frequência Fundamental. Para os homens manteve a resposta apresentada na seção anterior (sinal de voz x sinal do eletroglotógrafo), ou seja, com a idade ocorre um aumento da  $F_0$ . Já o gráfico (*boxplot*) das mulheres apresentou uma queda, diferindo da análise anterior, mas corroborando com a teoria.

***Jitter:***

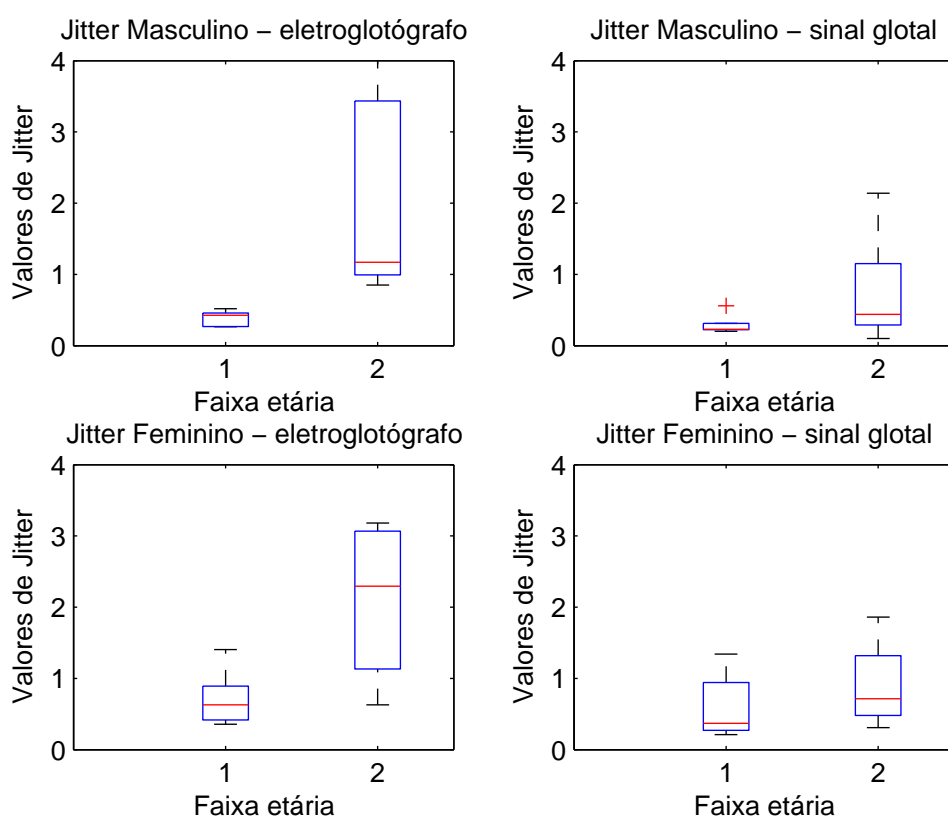


Figura 4.11: Jitter - Sinal Eletroglotógrafo - Sinal glotal.

A Fig.4.11 mostra o gráfico (*boxplot*) do *Jitter* entre as faixas etárias. Podemos ver que o *Jitter* aumenta em ambos os gráficos, mostrando o mesmo resultado obtido na seção anterior (sinal de voz x sinal do eletroglotógrafo).

*Shimmer:*

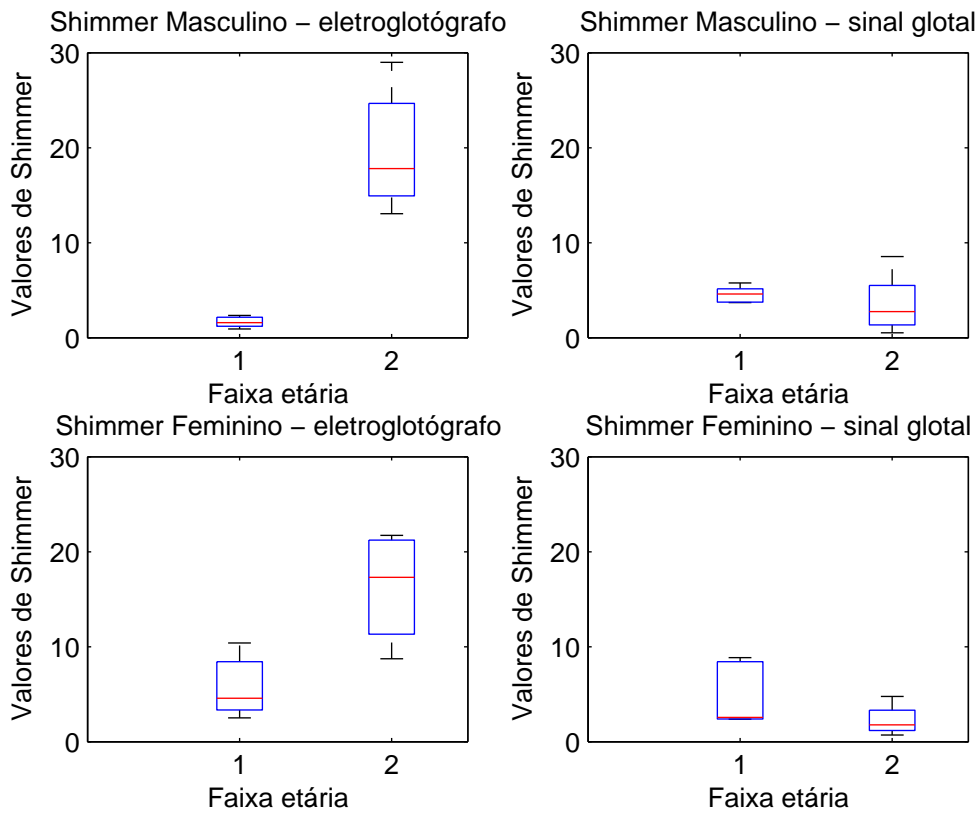


Figura 4.12: Shimmer - Sinal Eletroglotógrafo - Sinal glotal.

Na Fig.4.12 vemos que o *Shimmer* apresentou o mesmo resultado obtido na seção anterior (sinal de voz x sinal do eletroglotógrafo).

**HNR:**

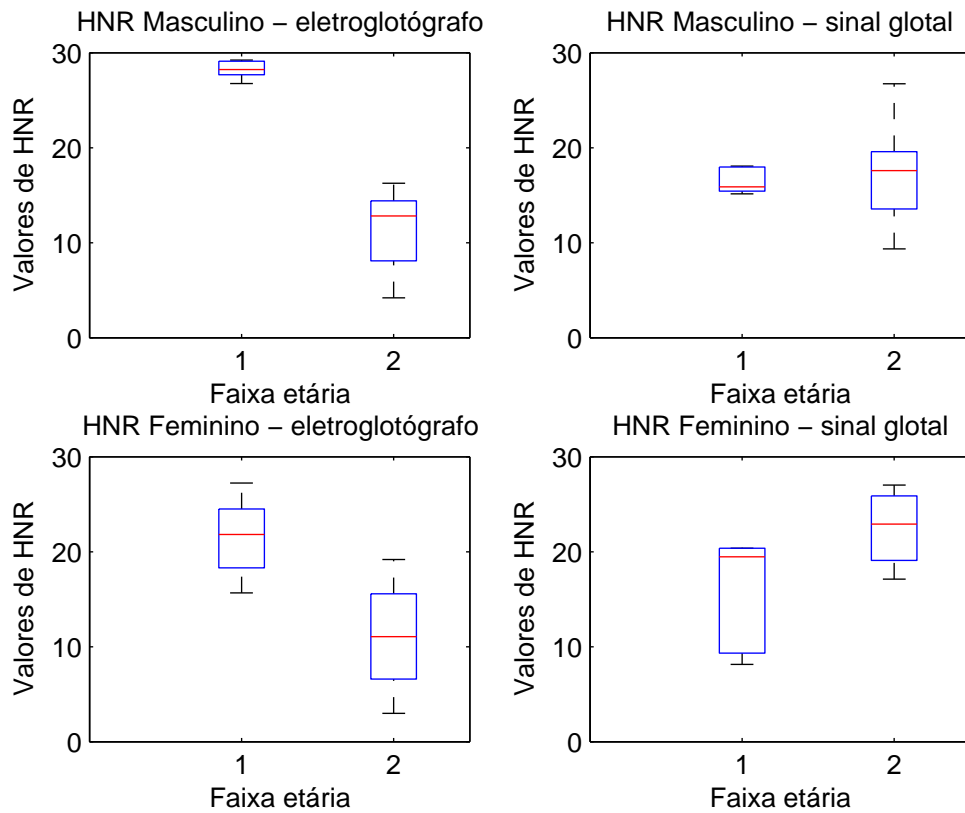


Figura 4.13: HNR (dB)- Sinal Eletroglotógrafo - Sinal glotal.

Vemos na Fig.4.13 o gráfico do *HNR*. Este apresentou, com o sinal glotal, o mesmo resultado obtido na seção anterior (sinal de voz x sinal do eletroglotógrafo), sendo a leitura do eletroglotógrafo, aparentemente mais precisa.

A partir de toda essa análise, consideramos como verdadeiro os valores de Frequência Fundamental e Jitter. Shimmer e HNR serão analisados posteriormente com um espaçamento maior entre as faixas etárias.

## 4.2 Parâmetros obtidos do sinal glotal com três faixas etárias

O objetivo deste estudo foi levantar os parâmetros entre três faixas etárias: jovens (15 a 30 anos), adultos (31 a 59 anos) e idosos (60 a 100 anos), e compará-los. Participaram 130 pessoas, sendo 53 homens e 77 mulheres, divididos em três grupos com faixas etárias diferentes (grupo II).

Neste caso só consideramos sinais de vozes e sinais glotais.

### 4.2.1 Parâmetros usados nas comparações

Para as análises, utilizamos os seguintes parâmetros:

Parâmetros do sinal de voz - utilizando o programa Praat: *F<sub>0</sub>*, *jitter*, *Shimmer* e *HNR*

Parâmetros do sinal glotal - utilizando o programa Aparat: *pp*, *K0*, *DH12*, *HRF*, *PSP*, *NAQ*, *Aq*, *CIQ*, *OQ1*, *OQ2*, *Oqa*, *Qoq*, *SQ1* e *SQ2*

Os parâmetros *DH12* e *HRF* são extraídos diretamente do programa Aparat, enquanto que *pp* e *K0* necessitam de cálculos adicionais. Nesses cálculos utilizamos parâmetros fornecidos pelo programa Aparat [63].



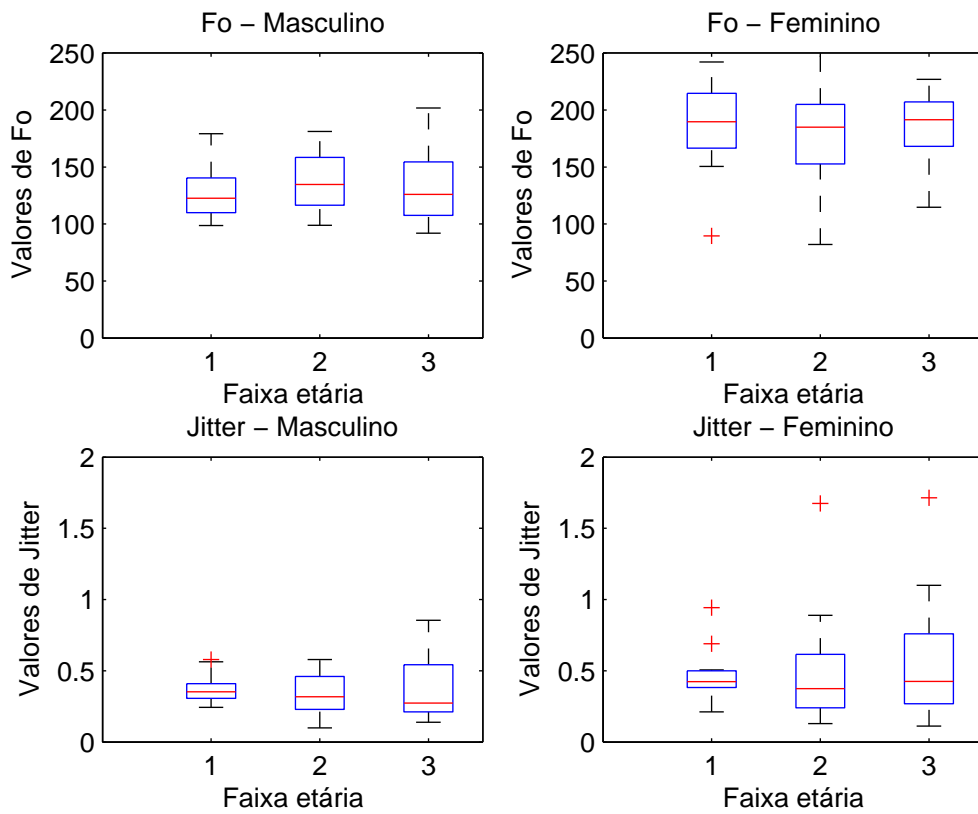


Figura 4.14: F0 e Jitter - Masculino e Feminino.

A Fig.4.14 apresenta os gráficos (*boxplot*) para a **Frequência Fundamental** e **Jitter**. Para os homens ocorreu um aumento da média da **Frequência Fundamental** entre faixas etárias 1 e 3, e para as mulheres houve uma tendência a manter a média de **Fo** entre as faixas etárias. Nos gráficos de **Jitter**, para os homens, ocorreu uma diminuição na mediana mas com grande variação na fase idosa, e para as mulheres um aumento na faixa etária 3.

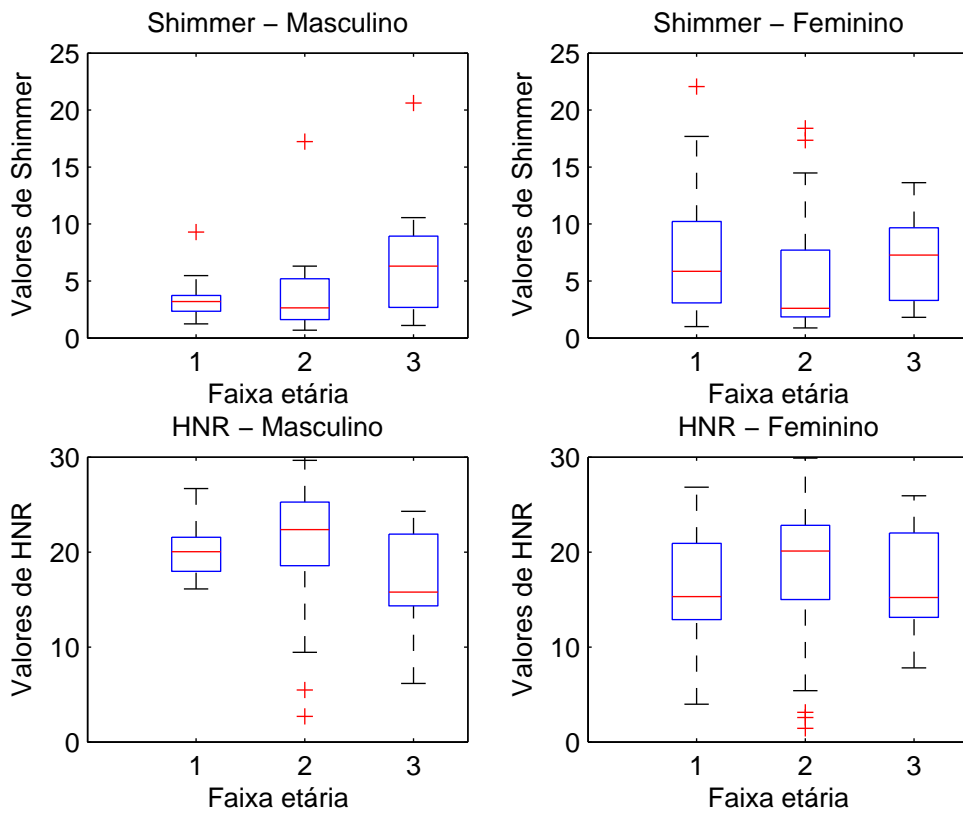


Figura 4.15: Shimmer e HNR - Masculino e Feminino.

A Fig.4.15 apresenta as análises para *Shimmer* e *HNR* para homens e mulheres. Dividindo em 3 faixas etárias, conseguimos observar melhor a variação dos parâmetros. O valor de *Shimmer* aumentou com a idade para os homens e para as mulheres, confirmando a literatura. Os resultados confirmaram, para ambos os sexos, que o *HNR* deve ser menor na fase idosa, uma vez que o nível de ruído tende a ser maior.

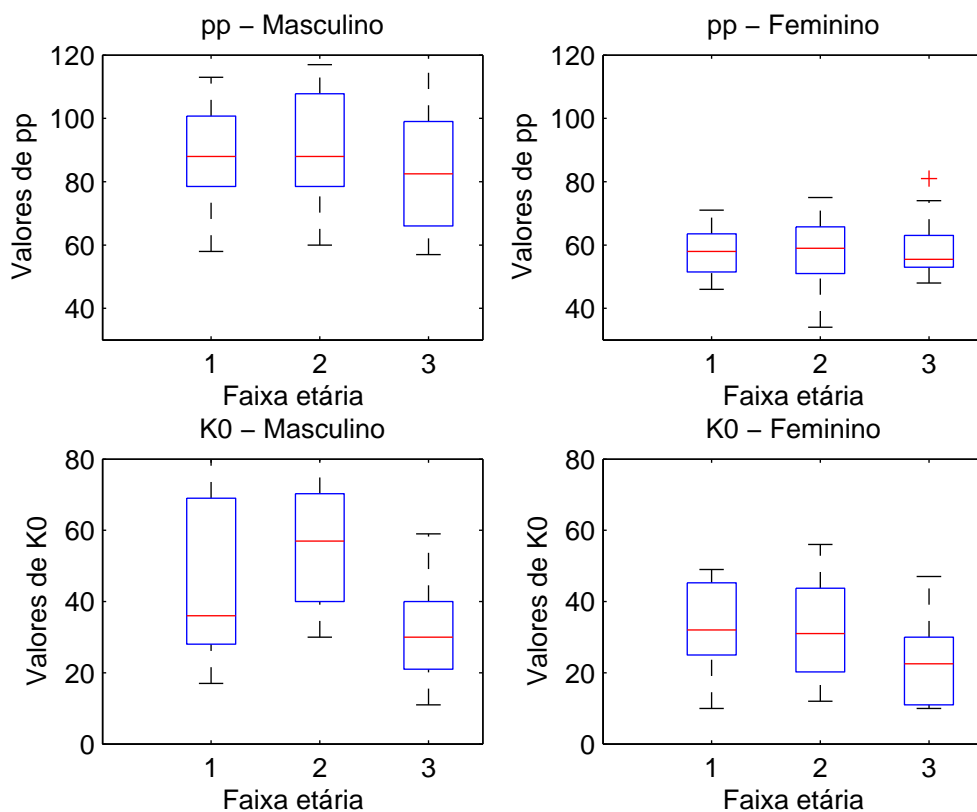


Figura 4.16: pp e K0 - Masculino e Feminino.

A Fig.4.16 apresenta as análises para **pp** (distância entre máximos de abertura glotal) e **K0** (duração da fase de fechamento) para homens e mulheres. Os valores de **pp** mostraram uma diminuição na fase idosa de ambos os sexos, o que pode ser devido às mudanças físicas do idoso. O **K0** (duração da fase de fechamento) mostrou uma queda no valor na fase idosa, o que se deve a uma diminuição no tempo de fechamento das cordas vocais.

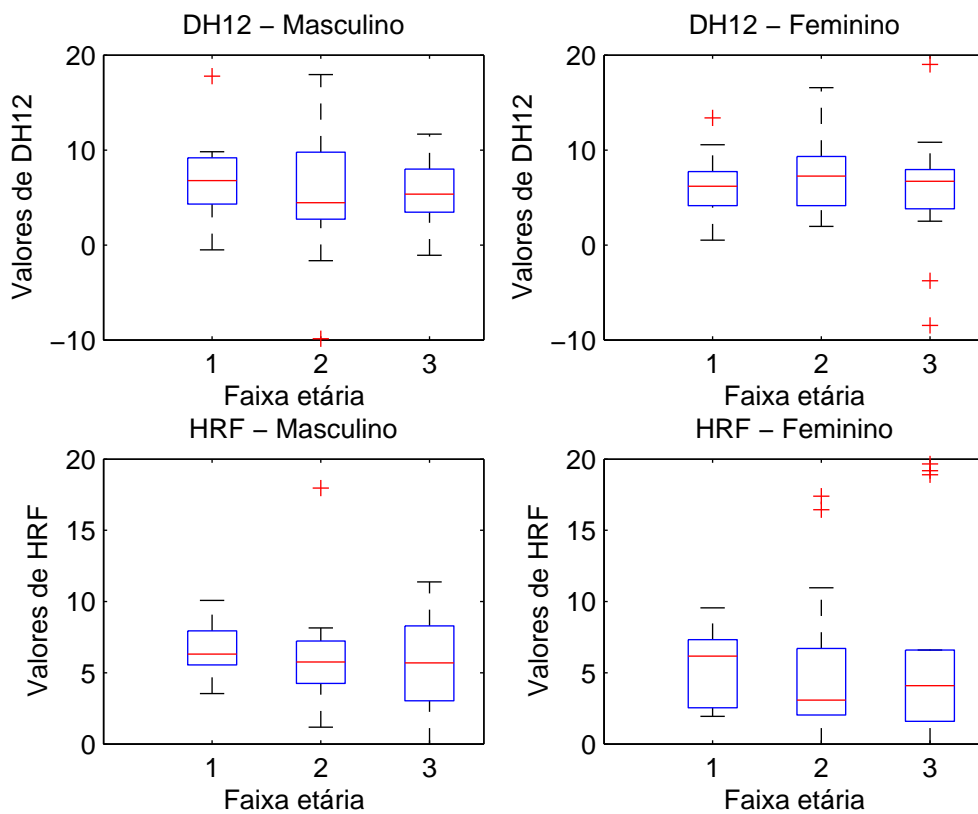


Figura 4.17: DH12 e HRF - Masculino e Feminino.

A Fig.4.17 apresenta as análises para **DH12** (diferença entre os 2 primeiros harmônicos) e **HRF** (riqueza harmônica do sinal) para homens e mulheres. Os gráficos de **DH12** mostraram uma pequena diminuição na fase idosa para os homens e tendência a manter-se para as mulheres. O **HRF** mostrou uma maior variação na fase idosa para os homens, e uma queda na mediana entre as faixas etárias 1 e 3 para as mulheres.

Os homens podem apresentar mais modificações físicas na fase idosa do que as mulheres; talvez essa seja a explicação para os valores encontrados [52] [65].

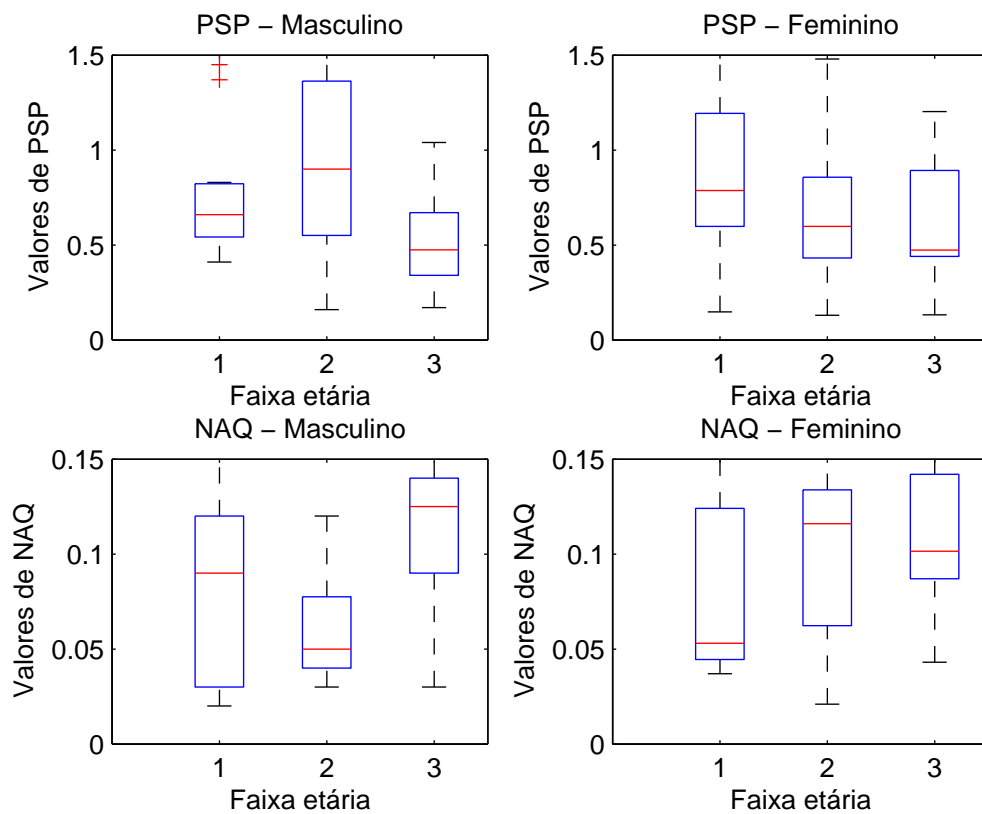


Figura 4.18: PSP e NAQ - Masculino e Feminino.

A Fig.4.18 apresenta as análises para **PSP** e **NAQ** para homens e mulheres. Os gráficos de **PSP** mostraram uma queda na fase idosa para os homens e mulheres. Sendo o PSP um parâmetro que mede a estimativa do declive espectral nos harmônicos, essa queda pode indicar uma variação na qualidade vocal na fase idosa. O **NAQ**, que pode ser usado para definir os parâmetros da fase de fechamento glotal, mostrou um aumento entre as faixas etárias na fase idosa para os homens e, para as mulheres, um aumento entre as faixas etárias 1 e 3.

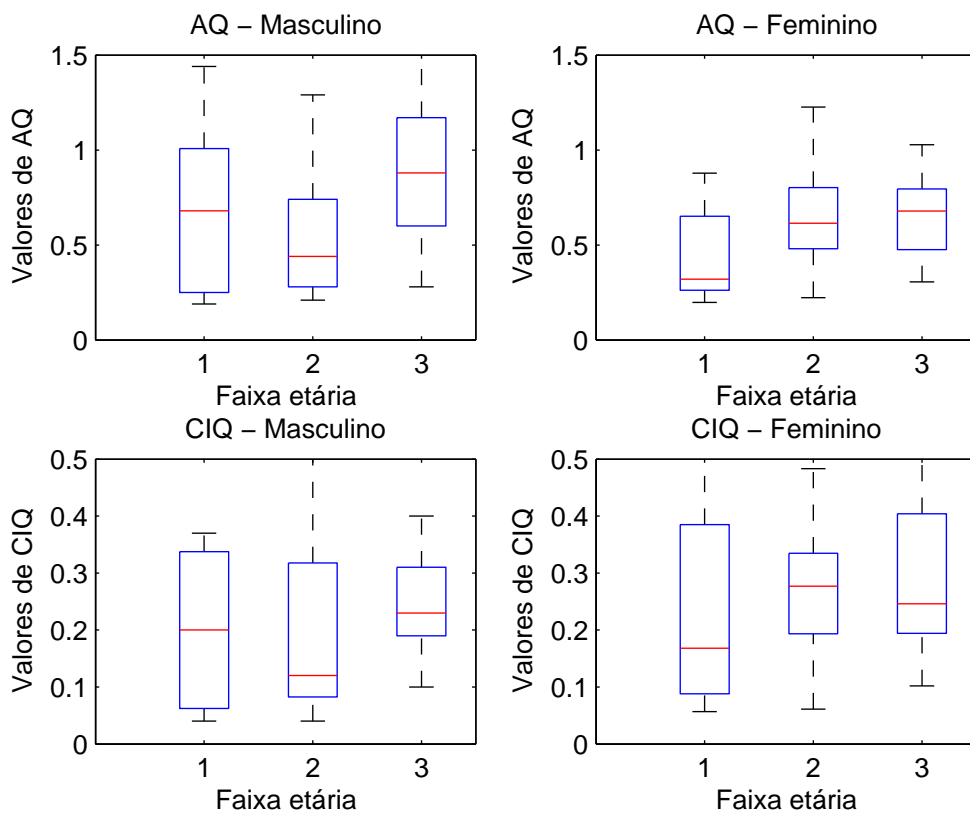


Figura 4.19: Aq e CIQ - Masculino e Feminino.

Na Fig.4.19 vemos os gráficos de **AQ** (quociente da amplitude) apresentarem um aumento para homens e mulheres na fase idosa. O gráfico do **CIQ** (quociente de fechamento), para os homens, mostrou um aumento na fase idosa e para as mulheres um aumento em relação a faixa etária 1, com grande variação na fase idosa.

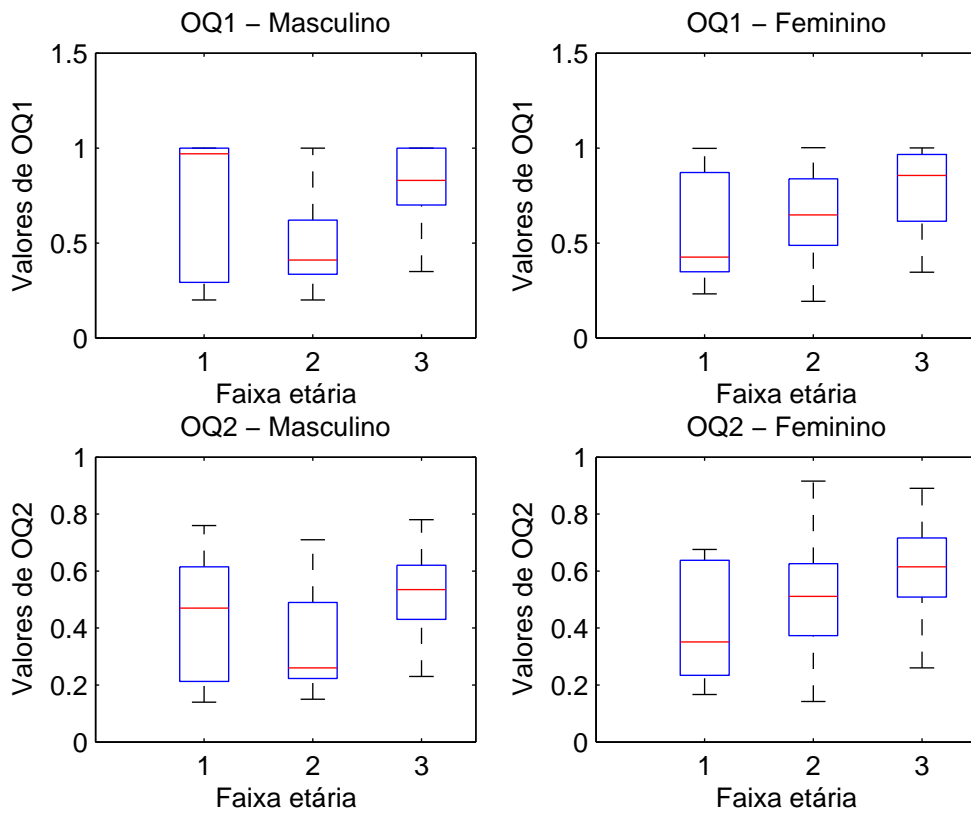


Figura 4.20: OQ1 e OQ2 - Masculino e Feminino.

A Fig.4.20 mostra os gráficos de **OQ1** e **OQ2** (quociente da abertura) e ambos apresentam um aumento na fase idosa para ambos os sexos. A nomenclatura **OQ1** e **OQ2** é porque podemos dividir o parâmetro em duas partes. **OQ1** pode ser definido como se abertura fosse somente no momento que começa o sinal glotal, **OQ2** é definido no momento da abertura da glote.

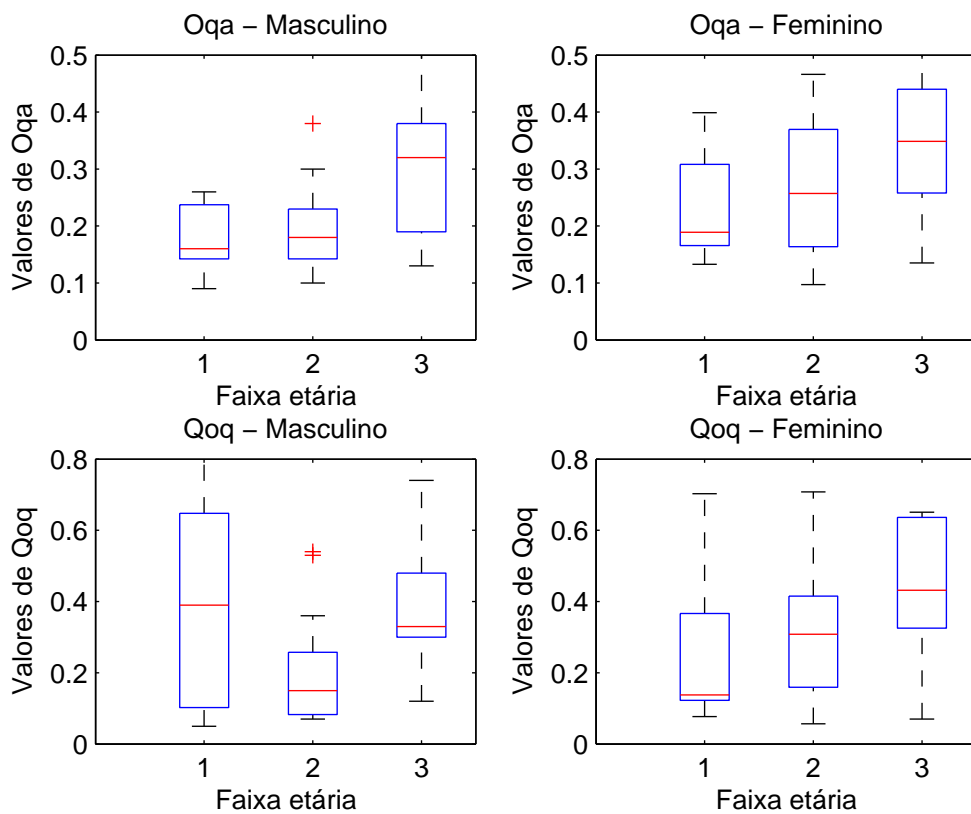


Figura 4.21: Oqa e Qoq - Masculino e Feminino.

A Fig.4.21 mostra os gráficos de **Oqa** (quociente de abertura definido pelo modelo *Liljencrants-Fant(LF)*) [51] e **Qoq** (abertura da glote como sendo o início do sinal glotal), ambos apresentam um aumento na fase idosa para ambos os sexos.



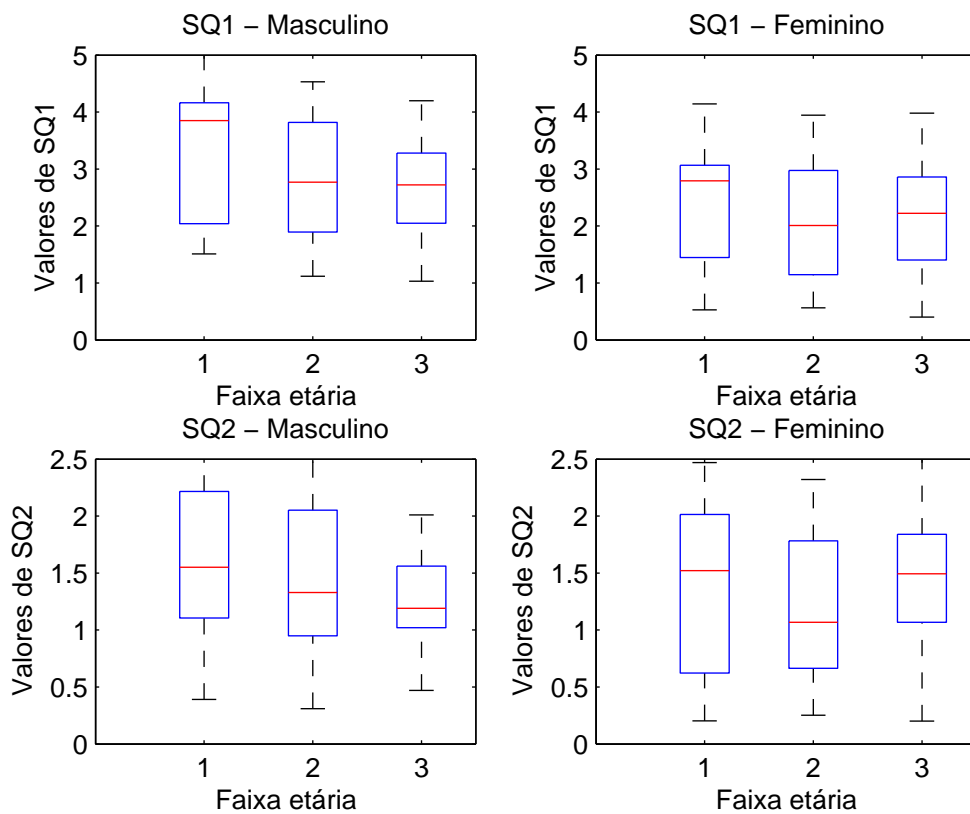


Figura 4.22: SQ1 e SQ2 - Masculino e Feminino.

A Fig.4.22 mostra os gráficos de **SQ1** e **SQ2** (quociente de velocidade - fase de abertura e fechamento glotal). Para os homens houve uma queda de **SQ1** e **SQ2**, e para as mulheres apresentou um aumento na fase idosa em relação a faixa etária 2. O parâmetro **SQ** pode ser dividido em dois. **SQ1** pode ser definido no momento que começa o sinal glotal, e **SQ2** é definido no momento da abertura da glote.

## 4.2.2 Gráficos das análises modificados

Após a análise dos gráficos (*boxplot*), pensamos na possibilidade de retirar alguns pontos que estão muito fora dos padrões (*outliers*), com o intuito de melhorar a visualização de alguns gráficos (*boxplot*). Fizemos isso com os parâmetros: **Frequência Fundamental**, *Jitter*, *Shimmer*, *HNR*, *pp*, *K0*, *DH12*, *HRF*.

Começamos com a Frequência Fundamental (**Fo**) e o *Jitter*.

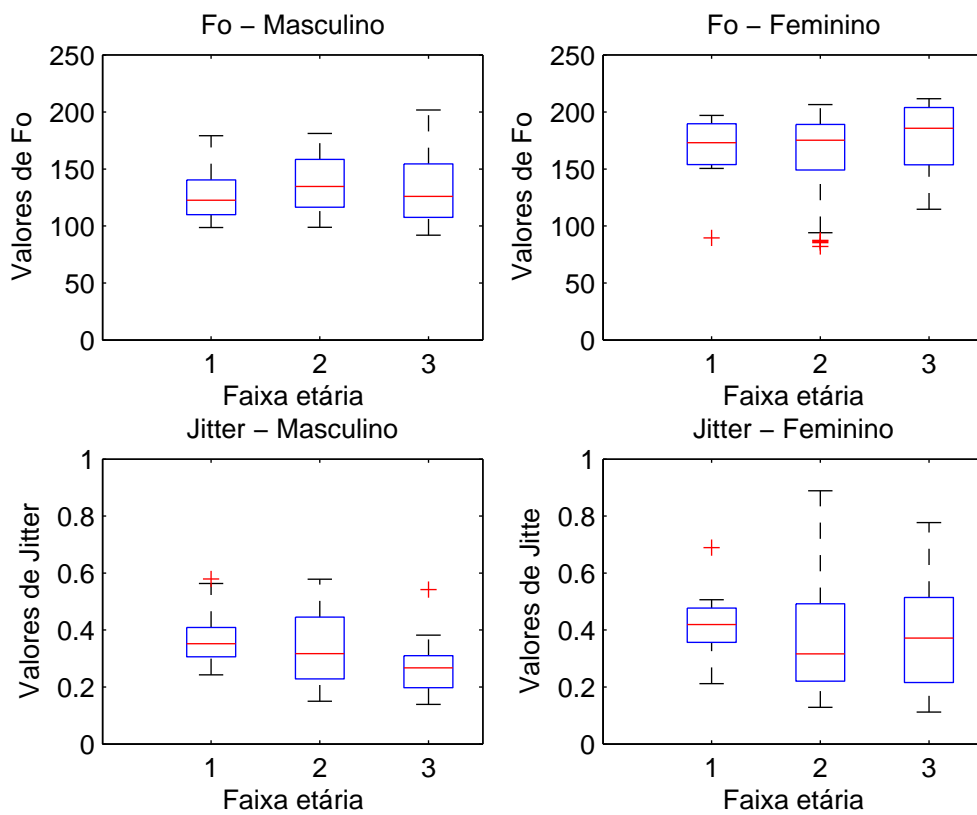


Figura 4.23: F0 e Jitter - Masculino e Feminino.

Na Fig.4.23 vemos que no gráfico (*boxplot*) de **Fo** para os homens não houve alteração nos valores depois dos pontos descartados. Nas mulheres, a média aumentou em relação às faixas etárias anteriores, com uma maior variação na fase idosa. O **Jitter** deveria ter aumentado para a fase idosa, conforme mostra a literatura, mas isso não ocorreu para os homens. No caso das mulheres, vemos um aumento entre as faixas etárias 2 e 3, o que não era possível ver no gráfico (*boxplot*) anterior.

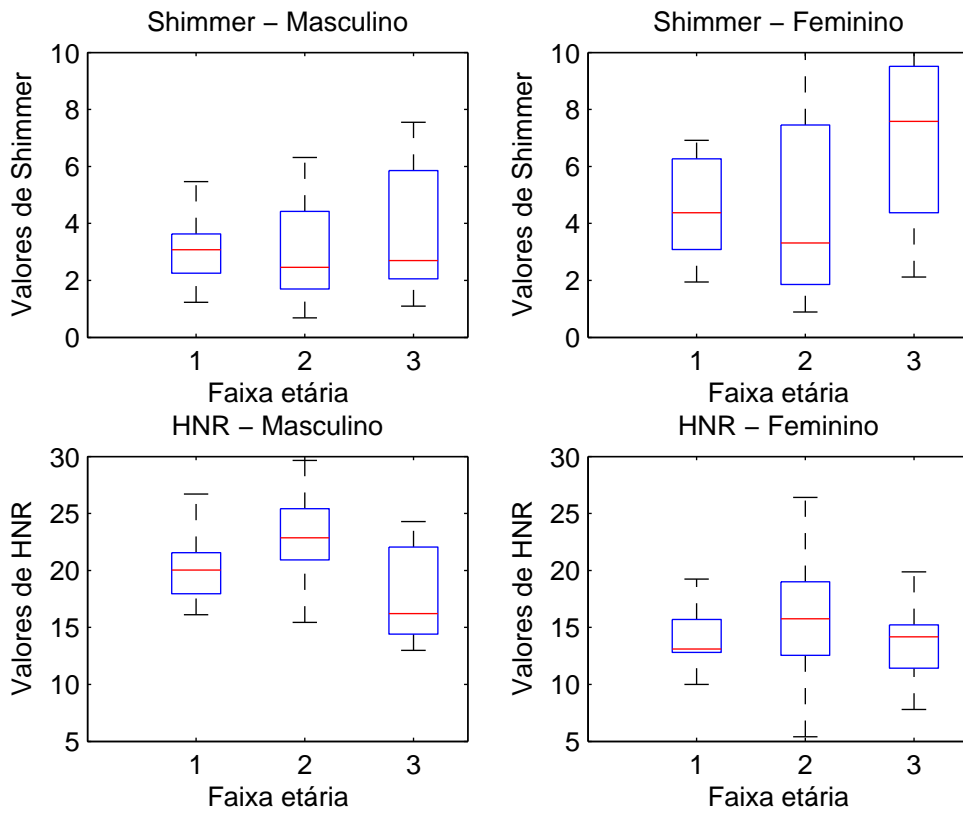


Figura 4.24: Shimmer e HNR - Masculino e Feminino.

Na Fig.4.24, vemos os gráficos (*boxplot*) de *Shimmer* confirmando o que diz a literatura. Os valores para os homens e para as mulheres aumentaram na fase idosa. A retirada dos pontos fora do padrão (*outliers*) beneficiou a análise. O *HNR* praticamente manteve a resposta anterior. A média do *HNR* nas mulheres teve uma queda e uma maior variação na fase idosa.

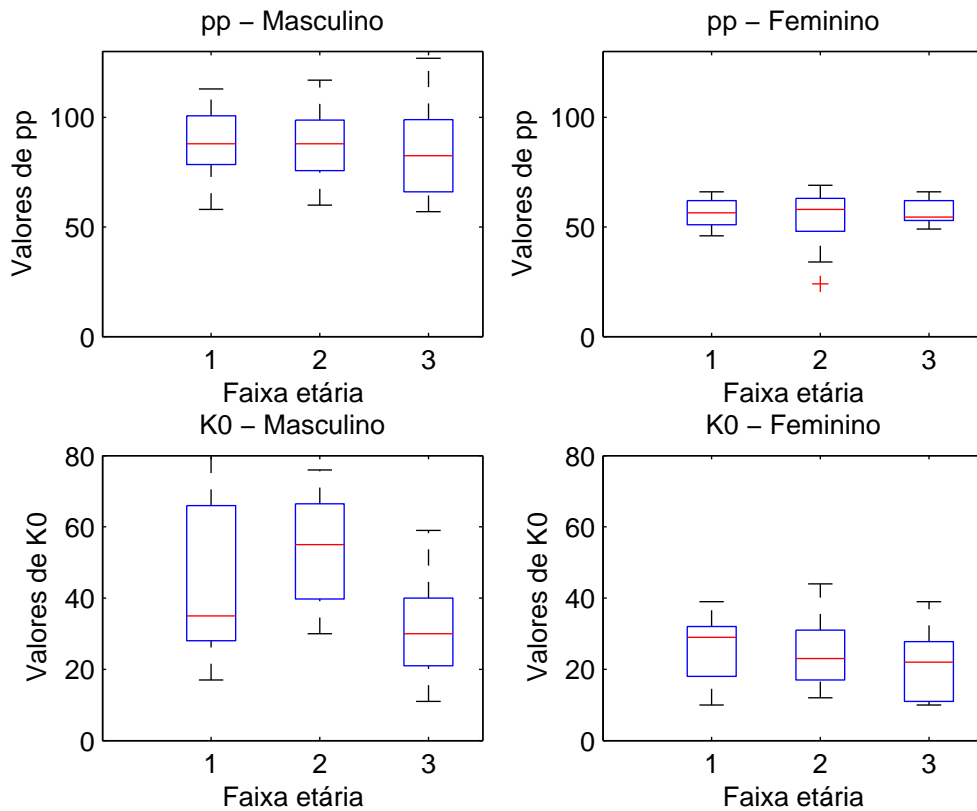


Figura 4.25: pp e K0 - Masculino e Feminino.

Na Fig.4.25, os valores dos gráficos (*boxplot*) se mantiveram e as análises anteriores estão mantidas. Os valores de **pp** mostraram uma pequena diminuição na fase idosa de ambos os sexos, que pode ser devida às mudanças físicas do idoso. Quanto ao parâmetro **K0**, tivemos uma queda no valor na fase idosa, o que se deve a uma diminuição no tempo de fechamento das cordas vocais.

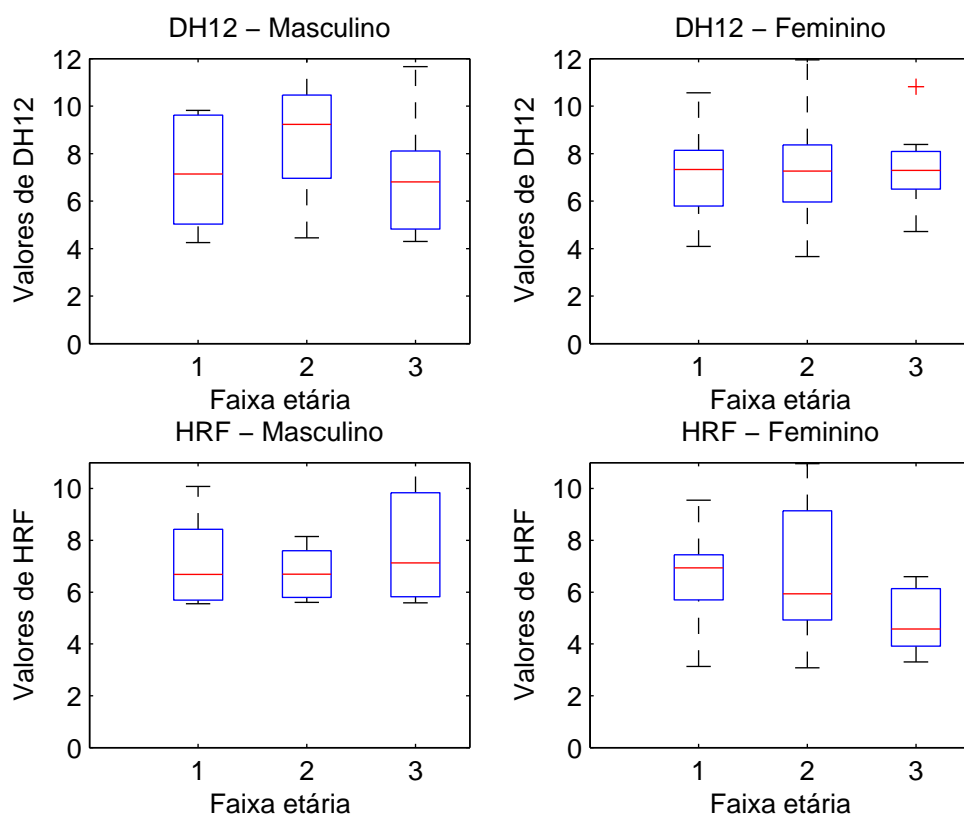


Figura 4.26: DH12 e HRF - Masculino e Feminino.

Na Fig.4.26, vemos os gráficos (*boxplot*) de **DH12** e **HRF**. Os gráficos (*boxplot*) **DH12** confirmam a queda na fase idosa para os homens, principalmente em relação à segunda faixa, e uma tendência a manter-se o valor para as mulheres. Da faixa etária 2 para a 3 observamos uma queda do **DH12**, possivelmente devido a problemas físicos ocorridos na terceira idade. O **HRF**, nos homens, que tinha sofrido uma queda agora, sem os pontos fora de padrão, tiveram uma resposta contrária. Possivelmente foram tirados pontos em excesso. Nas mulheres ficou mais visível a queda no valor de **HRF** na fase idosa, o que acreditamos que seja o correto.

### 4.3 Dispersão dos parâmetros

Com o objetivo de identificar parâmetros que possam separar vozes envelhecidas, fizemos uma análise de variação dos parâmetros ao longo do tempo em que as vogais eram pronunciadas. Para isso, escolhemos um locutor masculino e um locutor feminino das faixas etárias 1 e 3.

Para cada sinal, fizemos uma divisão em quatro partes, procedemos também a filtragem inversa, e levantamos características/parâmetros de cada uma dessas partes. Comparamos os resultados de uma parte com a outra. Dessa forma, tentamos verificar se o parâmetro variava muito ou pouco no decorrer do sinal, tentando identificar se o cálculo da média seria realmente significativo.

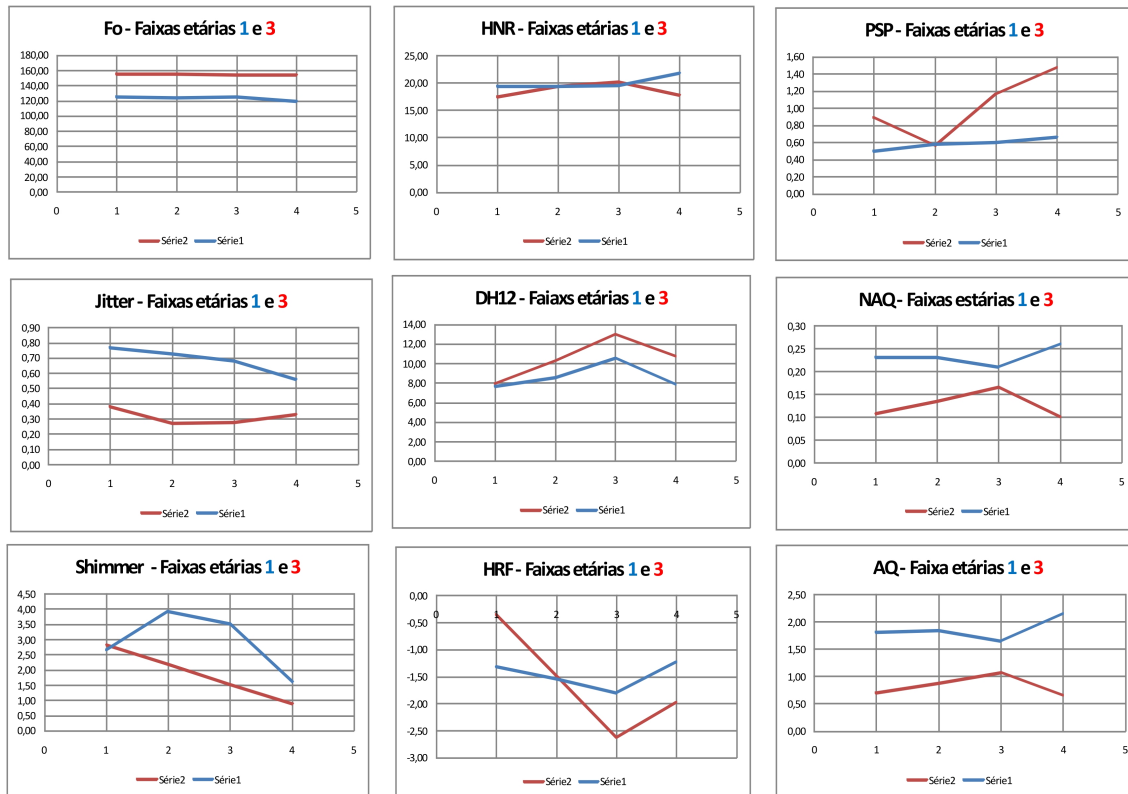


Figura 4.27: Locutor Masculino Faixa etária 1 e 3 (eixo x-fases do sinal no tempo/eixo y-valores dos parâmetros).

Na Fig. 4.27 temos os parâmetros **Fo**, **Jitter**, **HNR**, **DH12**, **NAQ** e **AQ** que apresentaram pouca variação ao longo do sinal, enquanto que os outros: **PSP**, **Shimmer** e **HRF** mostraram uma grande variação.

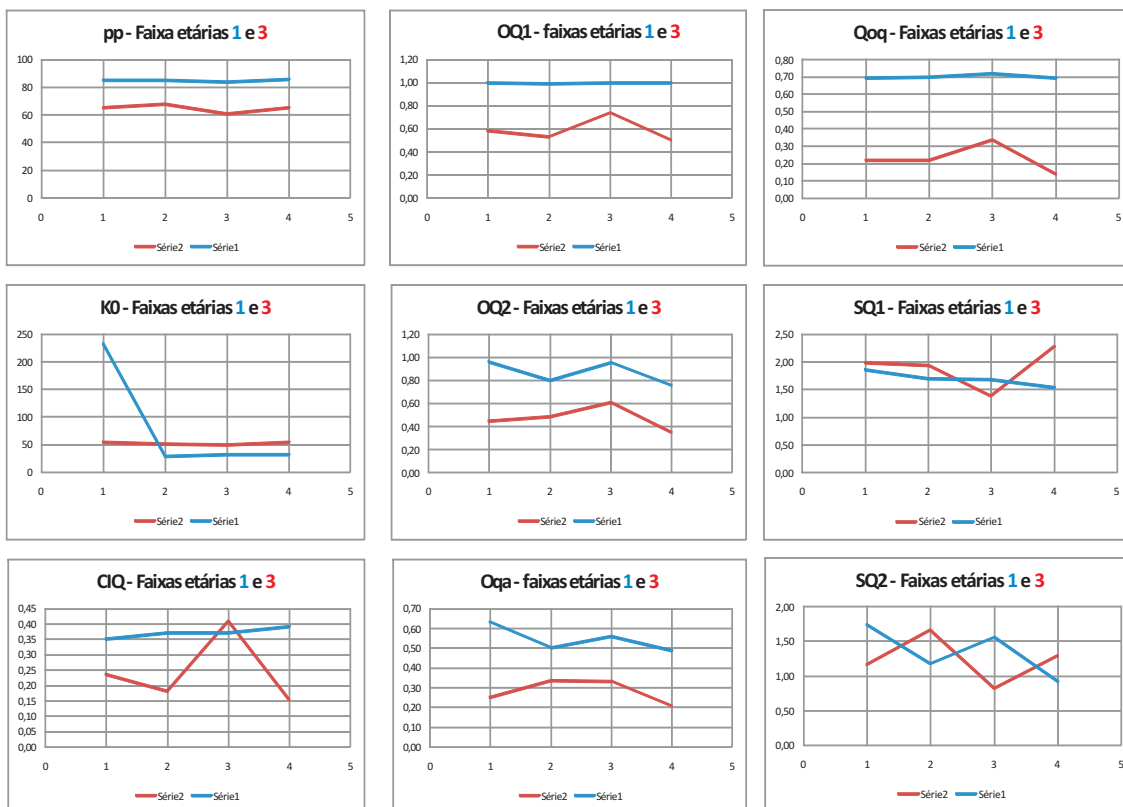


Figura 4.28: Locutor Masculino Faixa etária 1 e 3.

A Fig. 4.28 mostra **pp**, **OQ1**, **OQ2**, **Qoq**, **Oqa** com pouca variação, os gráficos **SQ1** e **SQ2** com variações consideráveis e o **K0** e **CIQ** tiveram grandes variações. **K0** apresenta uma queda considerável na primeira fase do sinal na faixa etária 1, mantendo-se constante até o fim do sinal e o **CIQ** um aumento na fase 3 da faixa etária 3.



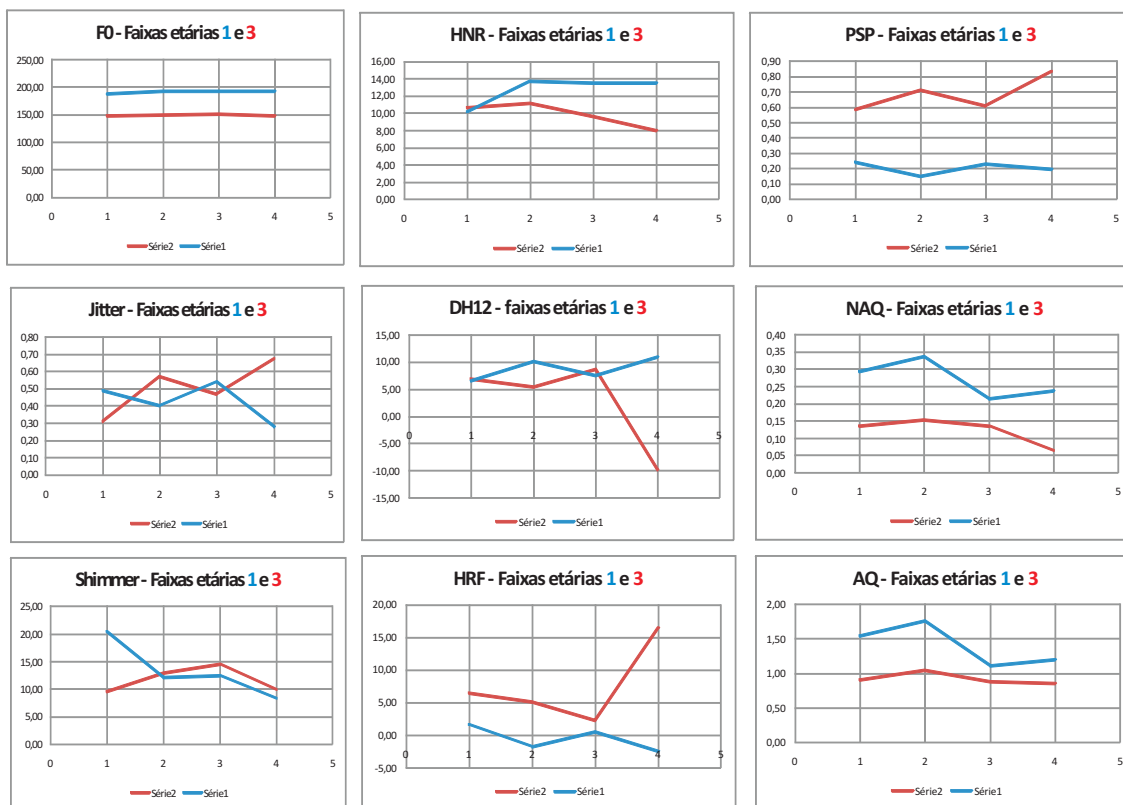


Figura 4.29: Locutor Feminino Faixa etária 1 e 3.

A Fig.4.29 mostra os gráficos para locutores femininos. Temos **FO**, **AQ** e **HNR** com pouca variação, **Shimmer**, **HRF** e **NAQ** com variação nos extremos do sinal, enquanto que **Jitter**, **DH12** e **PSP** com grandes variações.

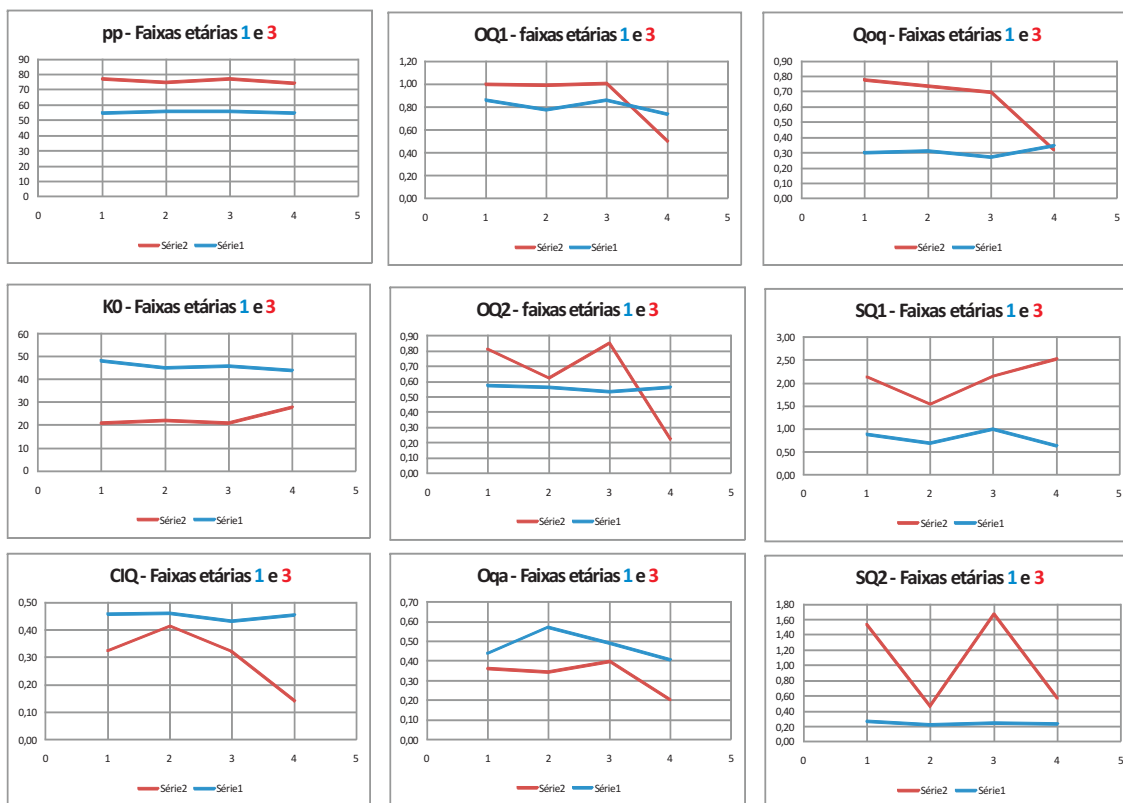


Figura 4.30: Locutor Feminino Faixa etária 1 e 3.

Na Fig.4.30 os gráficos de pp e K0 tiveram pouca variação. Alguns gráficos como OQ1, Qoq, SQ1, CIQ tiveram grandes variações na parte final do sinal. Os gráficos, OQ2, Oqa e SQ2 mostraram grande dispersão entre as faixas etárias.

## **Análise dos Parâmetros**

O estudo dos parâmetros realizado neste capítulo leva a algumas discussões que serão apresentadas a seguir. Antes, porém, gostaríamos de lembrar as comparações que foram realizadas:

Com duas faixas etárias (grupo I), fizemos dois tipos de comparações: Sinal de voz com sinal eletroglotográfico e sinal de voz com sinal glotal. As faixas etárias foram divididas em: faixa etária 1 (jovens), faixa etária 2 (idosos).

Com três faixas etárias (grupo II), usamos apenas o glotal. As faixas etárias foram divididas em: faixa etária 1 (jovens), faixa etária 2 (adultos), faixa etária 3 (idosos).

A **Frequência Fundamental** e o **Jitter** apresentaram valores dentro das expectativas em todos os casos e podemos dizer que apresentam-se como bons discriminantes na identificação da faixa etária correspondente aos idosos.

O **Shimmer** e o **HNR** não apresentaram valores confiáveis quando utilizada a comparação com apenas duas faixas etárias. Porém, ao serem utilizadas três faixas etárias, esses parâmetros mostraram-se como bons discriminantes para a faixa etária correspondente ao idosos, em ambos os sexos.

Os parâmetros relacionados com a intensidade, abertura e fechamento glotal e também com a qualidade da voz, extraídos a partir do sinal glotal, apresentaram importantes informações sobre o envelhecimento da voz. São eles: **DH12**, **HRF**, **PSP**, **NAQ**, **OQ** (**OQ1** e **OQ2**), **Oqa**, **SQ** (**SQ1** e **SQ2**).

Esses sinais só foram analisados usando três faixas etárias.

O **DH12** apresentou valores mais baixos para os homens, da faixa etária 3, e manteve-se constante para as mulheres também da faixa etária 3.

O **HRF** e o **PSP** apresentaram, na faixa etária 3, valores mais baixos para ambos os sexos.

O **NAQ**, **CIQ** e o **OQ** apresentaram valores mais altos para a faixa etária 3, para ambos os sexos, em relação à faixa etária 1. Esses parâmetros têm significativo efeito sobre as emoções e sexo dos locutores, e têm relação com mudanças de intensidade da voz [1].

O parâmetro **SQ** avalia a assimetria do pulso glotal durante a fonação, auxiliando na identificação da queda da qualidade vocal.

Os parâmetros **AQ** e **PP** podem ser relacionados às mudanças físicas do aparelho fonador para os idosos. Verificamos que o parâmetro **AQ** apresentou um valor maior para ambos os sexos na faixa etária 3. E o **PP** apresentou valores mais baixos para os homens na faixa etária 3 e manteve-se constante para as mulheres, comparando todas as faixas etárias. Os valores de **AQ** e **PP** para as mulheres mantiveram-se constantes, comparando as três faixas etárias. As mulheres idosas podem apresentar pouca modificação física em seu aparelho fonador, segundo [74].

Os parâmetros **K0** e **QoQ** descrevem fases de abertura e fechamento glotal [63], e podem ser bons discriminantes para a identificação de uma voz idosa pois, com o passar do tempo, diversas alterações anatômicas e fisiológicas ocorrem e podem comprometer a abertura da glote. Nas análises, **K0** apresentou valores mais baixos na faixa etária 3. E, **QoQ** manteve-se constante para os homens e apresentou um valor maior para as mulheres.

Apresentamos, a seguir, duas tabelas com um resumo dos parâmetros calculados usando cada tipo de sinal discutido neste trabalho (sinal de voz, sinal eletroglotográfico e sinal glotal), além de classificar como Confiável e não Confiável dependendo da compatibilidade das características dos parâmetros levantados e, também, de acordo com as discussões apresentadas na literatura.

Tabela 4.18: Parâmetros Grupo I.

	<b>Grupo I</b>					
	<b>Masculino</b>			<b>Feminino</b>		
	<b>Faixas Etárias 1 e 2</b>			<b>Faixas Etárias 1 e 2</b>		
<b>Sinais</b>	<b>Voz</b>	<b>Eletro</b>	<b>Glotal</b>	<b>Voz</b>	<b>Eletro</b>	<b>Glotal</b>
<i>Freq.Fund.</i>	Sim/Confiável	Sim/Confiável	Sim/Confiável	Sim/Confiável	Sim/Confiável	Sim/Confiável
<i>Jitter</i>	Sim/Confiável	Sim/Confiável	Sim/Confiável	Sim/Confiável	Sim/Confiável	Sim/Confiável
<i>Shimmer</i>	Não/Confiável	Sim/Confiável	Não/Confiável	Não/Confiável	Sim/Confiável	Não/Confiável
<i>HNR</i>	Não/Confiável	Sim/Confiável	Não/Confiável	Não/Confiável	Sim/Confiável	Não/Confiável

Tabela 4.19: Parâmetros Grupo II.

	<b>Grupo II</b>	
	<b>Masculino</b>	<b>Feminino</b>
<b>Sinal Glotal</b>	<b>Faixas Etárias 1 e 3</b>	<b>Faixas Etárias 1 e 3</b>
Freq.Fund.	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Baixa
<i>Jitter</i>	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Alta
<i>Shimmer</i>	Sim/Confiável/Disp.Alta	Sim/Confiável/Disp.Baixa
<i>HNR</i>	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Baixa
pp	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Baixa
K0	Sim/confiável/Disp.Alta	Sim/Confiável/disp.Baixa
DH12	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Alta
HRF	Sim/Confiável/Disp.Alta	Sim/Confiável/Disp.Baixa
PSP	Sim/Confiável/Disp.Alta	Sim/Confiável/Disp.Baixa
NAQ	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Baixa
AQ	Sim/Confiável/Disp.Baixa	Sim/Confiável/Baixa
CIQ	Sim/Confiável/Disp.Alta	Sim/Confiável/Disp.Alta
OQ1	Sim/Confiável/disp.Baixa	Sim/Confiável/Disp.Baixa
OQ2	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Alta
Oqa	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Baixa
Qoq	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Baixa
SQ1	Sim/Confiável/Disp.Baixa	Sim/Confiável/Disp.Baixa
SQ2	Sim/Confiável/Disp.Alta	Sim/Confiável/Disp.Alta

# Capítulo 5

## Aplicação

Como um exemplo de aplicação dos parâmetros e análises estudados no capítulo anterior, implementamos uma rede neural com os parâmetros extraídos de um sinal de voz, de modo a identificar automaticamente casos de vozes envelhecidas.

### 5.1 Redes Neurais Artificiais - RNA

Uma rede neural é um sistema computacional constituído por unidades conhecidas como neurônios. Os neurônios são elementos processadores interligados, trabalhando em paralelo para desempenhar uma determinada tarefa. Os modelos RNA's constituem uma importante técnica estatística não-linear capaz de resolver uma gama de problemas de grande complexidade. Por isso, são modelos úteis em situações em que não é possível definir explicitamente uma lista de regras. Em geral, isso acontece quando o ambiente gerador dos dados muda constantemente. As principais áreas de atuação são para classificação de padrões e previsão [35]. Para começar a falar de redes neurais, o ponto de partida é definir o que são e como se constituem as suas unidades básicas. Redes Neurais Artificiais são técnicas com-

putacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência. O conhecimento é passado para a rede por um algoritmo de treinamento, e o aprendizado é transformado e armazenado em densidades de conexões chamadas “pesos” [77]. O aprendizado resulta em muitas apresentações de determinado conjunto de treinamento. Destacamos o conceito de época, sendo uma apresentação completa de todo o conjunto de treinamento.

As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento inteligente de uma rede neural artificial vem das interações entre as unidades de processamento da rede.

Usualmente os neurônios estão dispostos em camadas que são classificadas em três grupos [59]:

- Camada de Entrada: onde os padrões são apresentados à rede;
- Camadas Intermediárias ou Escondidas: onde é feita a maior parte do processamento, através das conexões ponderadas; podem ser consideradas como extratoras de características;
- Camada de Saída: onde o resultado final é concluído e apresentado.

A rede neural Multi-Layer Perceptron (MLP) foi utilizada neste estudo por ser um aproximador universal e que representa qualquer função. O algoritmo *Back-Propagation* é o mais utilizado em aplicações práticas de previsão, classificação e reconhecimento de padrões em geral [77].



Para utilizar uma Rede Neural artificial (RNA) na tarefa de reconhecimento de locutor deve-se levar em conta:

- Parâmetros de uma Rede Multilayer Perceptron;
- Normalização dos pesos;
- Critério de parada do treinamento de RNA;
- Variações no treinamento de rede MLP;

A normalização dos pesos é de extrema importância para o sistema, porque é nela que ocorre o tratamento dos valores conectados. A partir dos dados adquiridos no módulo de captura, cabe ao módulo de normalização o preparo para que sejam entregues à análise da Rede Neural [77].

O treinamento, após iniciado, pode ser interrompido em três circunstâncias: quando é atingido o número máximo de épocas do treinamento, quando é atingido o erro mínimo desejado, ou quando uma determinada porcentagem dos dados de treinamento é corretamente classificada [77].

## 5.2 Seleção de atributos e Método *Wrapper*

A seleção de atributos tem como objetivo descobrir subconjuntos de atributos relevantes para uma tarefa, considerando os originais. Atributos redundantes podem confundir o algoritmo, prejudicando o desempenho de aprendizagem na velocidade e na taxa de acerto [67].

Em relação aos algoritmos de busca, podemos implementar dois tipos de abordagens para seleção de atributos: filtro e *wrapper*. Essas abordagens independem do método ou algoritmo utilizado na seleção, e são caracterizadas por seu grau de dependência em relação ao algoritmo de classificação.

Para a seleção de atributos utilizando abordagem *wrapper*, um subconjunto de soluções é definido de acordo com uma base de dados e algoritmo de indução particulares, levando em conta a tendência indutiva do algoritmo e sua interação com o conjunto de treinamento. A abordagem utilizando filtro, ao contrário da abordagem *wrapper*, escolhe um subconjunto de atributos independente do algoritmo de classificação, estimando a qualidade dos atributos apenas em relação aos dados [61].

A Fig. 5.1 mostra o esquema de um algoritmo de seleção de atributos utilizando a abordagem *wrapper*. A metodologia *wrapper* é muito utilizada quando se deseja selecionar atributos em problemas de aprendizado supervisionado, tendo como forte característica o emprego de algoritmos de aprendizado de máquina para avaliar o custo dos subconjuntos de atributos selecionados durante o processo de busca [28]. Após a busca de um subconjunto de atributos, este é avaliado e recebe um custo que, para ser válido, deve ser classificado por algoritmo de indução, resultando em uma estimativa de desempenho [67].

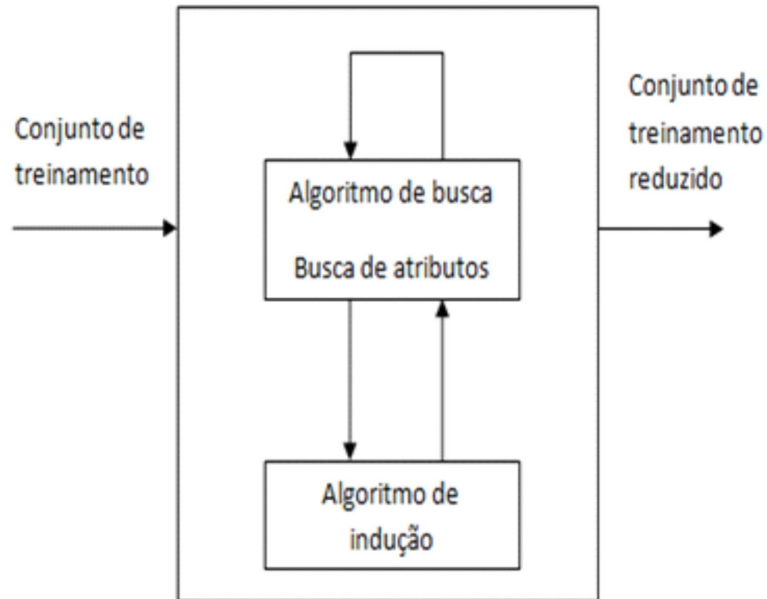


Figura 5.1: Abordagem Wrapper [28]

Os melhores subconjuntos serão selecionados e deverão passar por um segundo algoritmo de indução para extrair os atributos mais relevantes, para que estes demonstrem seu potencial [28].

Utilizamos a técnica de seleção de atributos com abordagem *wrapper*, e como algoritmo de busca um algoritmo genético [61].

### 5.3 Montagem da Rede Neural utilizando parâmetros do sinal de voz e sinal glotal

Para a montagem da Rede Neural, utilizaram-se 18 parâmetros, sendo 4 extraídos do sinal de voz e 14 do sinal glotal obtidos mediante filtragem inversa do sinal de voz.

Abaixo os parâmetros descritos:

- Frequência Fundamental, *Jitter*, *Shimmer* e *HNR*;
- DH12, HRF, PSP, K0, PP, NAQ, AQ, CIQ, OQ1, OQ2, OQA, QOQ, SQ1, SQ2.

### 5.3.1 Construção da Rede Neural

A rede neural foi implementada junto com Leonardo Forero Mendoza e faz parte do seu exame de qualificação de Doutorado na PUC-Rio - Departamento de Engenharia Elétrica. Agradecemos a sua contribuição.

Com o objetivo de ter a mesma quantidade de amostras em ambos os sexos, foram adicionadas 7 vozes masculinas e descartadas 17 vozes femininas, obtendo uma base de 60 vozes masculinas e 60 vozes femininas, divididas em três faixas etárias: 15-30 anos, 31-60 anos e 61-100 anos para cada sexo. Foi um total de 20 locutores por faixa etária.

A técnica de seleção de atributos com abordagem *wrapper* foi constituída por: crossover de 0,6, a mutação de 0,03 e uma população de 40 indivíduos por 60 gerações, utilizando como cromossomo 18 parâmetros de entrada. Para a classificação utilizou-se uma rede Neural artificial MLP, com 3 neurônios de entrada e 500 épocas de treinamento [28]. O número de épocas é determinado pelo conjunto de validação, que analisa a variação do erro durante o treinamento.

Uma vez determinado os parâmetros de maior relevância, através do *wrapper*, podemos montar a rede neural.

Os resultados dos atributos mais relevantes para a identificação do envelhecimento da voz foram [28]:

- Base feminina: Shimmer, AQ;
- Base masculina: K0, DH12, NAQ, OQ1, OQ2, Oqa, SQ2;
- Base mista: Jitter, HRF, NAQ, AQ, CIQ, OQ1, Oqa, SQ2.

Com os subconjuntos escolhidos na seleção de parâmetros, utilizamos 65% da base de dados para treinamento, 15 % como conjunto de validação e 20% para o teste [28].

As Fig.5.2 e Fig.5.3 mostram as redes neurais correspondentes aos locutores femininos e masculinos:

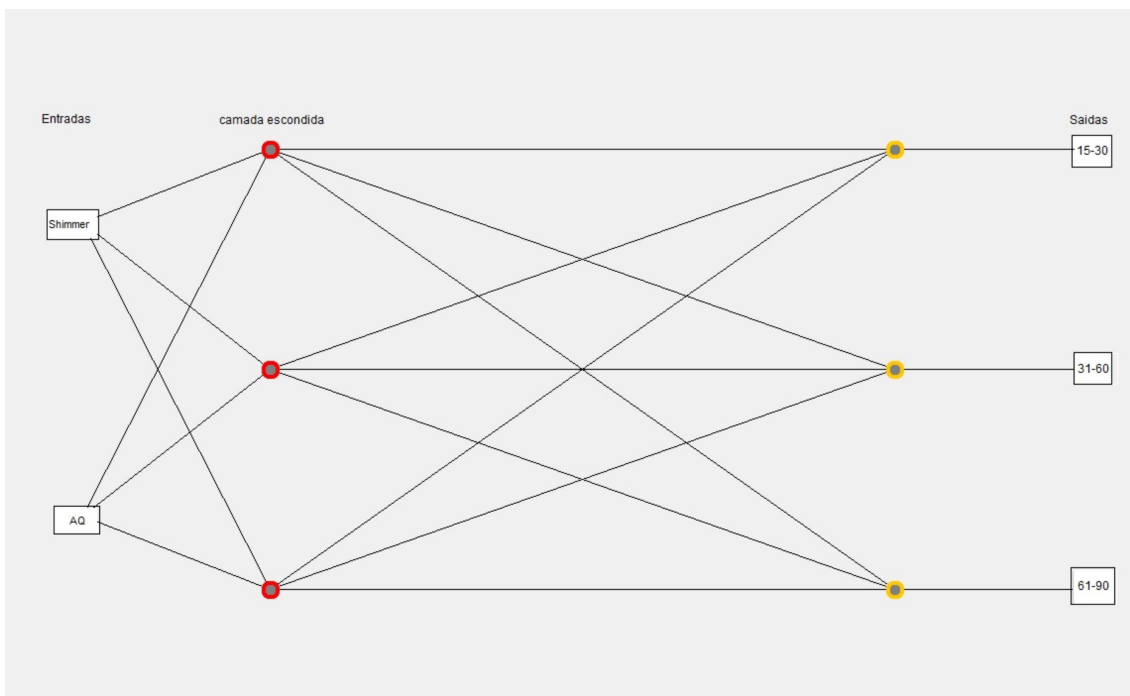


Figura 5.2: Rede Neural MLP para o grupo feminino

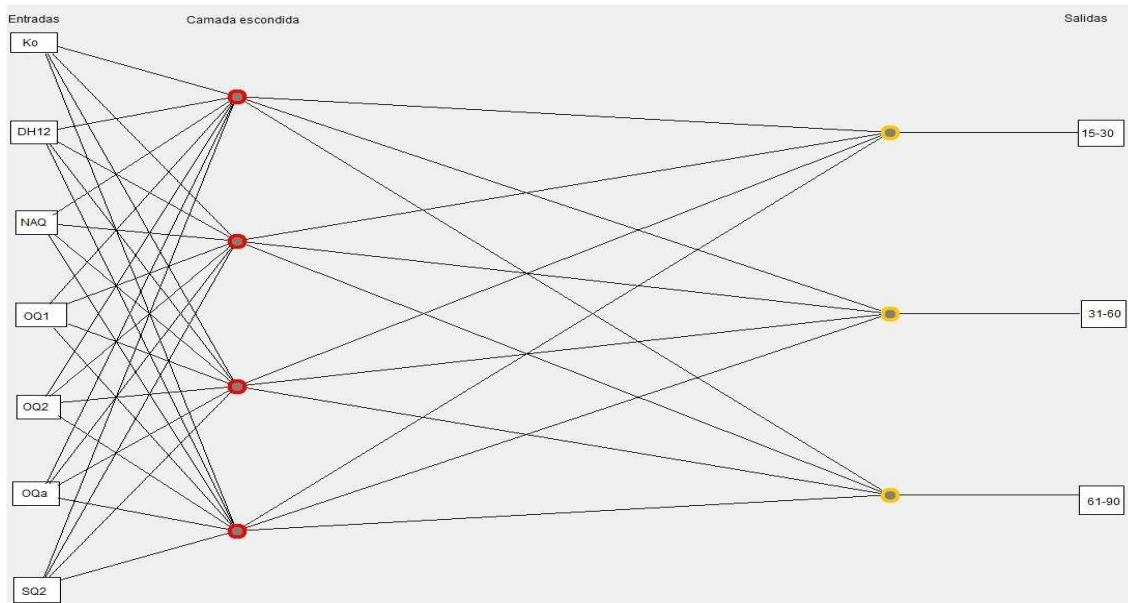


Figura 5.3: Rede Neural MLP para o grupo masculino

### 5.3.2 Resultado

Para a base feminina a melhor configuração foi uma camada com 3 neurônios, com uma única camada intermediária, e 3000 épocas de treinamento. Desta maneira classificamos 10 locutores corretamente e 2 locutores incorretamente. Tivemos um acerto de 83,33% [28]. O Shimmer teve pouca variação depois dos 50 anos nos locutores femininos, e esse pode ser a causa de a rede neural não classificar em um 100% [28]. Para as mulheres tivemos o *Shimmer* e AQ como os atributos mais significativos. O *Shimmer* foi identificado como um bom parâmetro para identificar mudanças de faixa etária, e o AQ, que é uma medida que se relaciona com a pressão e a intensidade da voz [28] [63], funciona como um bom discriminante. Pelos gráficos o valor de AQ cresce à medida que aumenta a idade do locutor.

Para a base masculina a melhor configuração foi uma camada com 4 neurônios, com uma única camada intermediária, e 5000 épocas de treinamento. Desta maneira classificamos 11 locutores corretamente e 1 locutor incorretamente. Tivemos também um acerto de 91,66% para a base masculina [28]. Os atributos mais significativos para identificar idade entre homens forem Ko, DH12, NAQ, OQ1, OQ2, Oqa, SQ2.

Para a base mista (homens e mulheres) a configuração ideal foi uma camada de 5 neurônios, com uma camada intermediária, e 4000 épocas de treinamento. Conseguimos classificar 12 locutores tendo uma porcentagem de acerto de 83,3%.

Os resultados obtidos foram satisfatórios, a abordagem wrapper se mostrou eficaz, e todos os parâmetros escolhidos provaram ser bons discriminantes na mudanças das faixas etárias. Estudamos os valores de cada parâmetro para cada faixa etária separadamente, e todos apresentaram mudanças com o avanço da idade. Usando a rede neural vimos que, uma vez combinados os parâmetros, podem trazer grande contribuição na identificação de uma voz envelhecida.

# Capítulo 6

## Conclusões e Trabalhos Futuros

### 6.1 Conclusões

1. Procuramos através deste trabalho descobrir parâmetros que permitissem classificar uma voz como envelhecida. Na análise do sinal de voz com o sinal eletroglotográfico, extraímos os parâmetros: Frequência Fundamental, *Jitter*, *Shimmer* e o *HNR*. Chamamos de grupo I, e dividimos os locutores, masculinos e femininos, em duas faixas etárias. A Frequência Fundamental e o *Jitter* se apresentaram como bons discriminantes utilizando tanto o sinal de voz como o sinal eletroglotográfico. As variações desses parâmetros foram idênticas nos dois métodos, apresentando uma variância maior quando utilizado o eletroglotógrafo. As análises do *Shimmer* e *HNR* apresentaram resultados contrários em relação aos métodos. O método do eletroglotógrafo apresentou a variação concordante com a literatura, e o método do microfone, discordante. Com relação ao gráfico do microfone, possivelmente o valor da faixa etária 2 não caracteriza um idoso propriamente dito. Será necessária uma análise com espaçamento maior entre as faixas etárias.



2. Executamos análises entre o sinal glotal e o sinal eletroglotográfico, com os locutores do grupo I. Utilizamos o programa Aparat para a filtragem inversa do sinal de voz de todos os locutores, e comparamos com o sinal eletroglotográfico correspondente. Utilizamos os mesmos parâmetros da análise anterior: Frequência Fundamental, *Jitter*, *Shimmer* e *HNR*, e obtivemos os mesmos resultados, em termos qualitativos do estudo, do sinal de voz comparado com o sinal do eletroglotógrafo. Podemos confirmar que o eletroglotógrafo fornece parâmetros confiáveis para as análises.
  
3. Com um número superior de gravações, somente com o sinal de voz, utilizamos os programas Praat e Aparat para a extração de diversos parâmetros acústicos, que podem ajudar na caracterização de uma voz envelhecida. Neste caso só trabalhamos com sinal de voz e sinal glotal, chamamos de grupo II e dividimos em três faixas etárias. A análise dos parâmetros extraídos com o programa Praat (*F<sub>0</sub>*, *Jitter*, *Shimmer* e *HNR*) tiveram o resultado concordante com a literatura. Um maior espaçamento entre as faixas etárias contribuiu para a análise, e concluímos que todos são bons discriminantes para identificar vozes envelhecidas. Os parâmetros obtidos pelo programa Aparat, um total de 14 ao todo, foram significantes para as análises de envelhecimento da voz. Alguns estão relacionados com a intensidade da voz, abertura ou fechamento glotal e a qualidade da voz [1] [63]. Problemas físicos, que ocorrem com o avanço da idade, são fatores que auxiliam na identificação do idoso, e podemos ver isso diretamente nos gráficos. Alguns parâmetros apresentaram pontos fora do padrão, que foram retirados da análise e os gráficos refeitos para uma melhor visualização. Os gráficos de dispersão mostram parâmetros que têm

uma variação muito grande ao longo do sinal mas, quando utilizados na média, podem contribuir para a classificação. Os parâmetros DH12, HRF, PSP, NAQ, OQ (OQ1 e OQ2), Oqa, SQ (SQ1 e SQ2), que estão ligados à qualidade e à intensidade vocal [63], mostraram-se concordantes com a literatura. Os parâmetros DH12, HRF, PSP (qualidade do sinal) apresentaram uma queda na fase idosa. Essa queda pode expressar problemas que ocorrem com a idade. O NAQ, CIQ e o OQ apresentaram um aumento para ambos os sexos na terceira idade. Esses parâmetros tem significativo efeito sobre as emoções e sexo dos locutores, e tem correlação com mudanças de intensidade da voz [1]. O parâmetro CIQ é inversamente proporcional à intensidade da voz, seu aumento caracteriza uma queda na intensidade do sinal. O parâmetro OQ relaciona a largura do sinal no momento da vibração, o valor alto caracteriza uma queda na amplitude do sinal glotal. Os parâmetros NAQ e AQ (amplitude do sinal) tiveram aumento na fase idosa. Com a idade, o sinal de voz apresenta um aumento de ruído na amplitude (*Shimmer*). O parâmetro SQ avalia a assimetria da glote durante a fonação, e apresentou um queda para ambos os sexos. Isso é concordante com a literatura. O PP (distância entre máximos de abertura glotal) teve uma queda para os homens e se manteve para as mulheres. Esse parâmetro não apresentou correlação com as faixas etárias. O K0 é um parâmetro pouco usado em análises de sinais, e se mostrou um bom discriminante para locutores masculinos e femininos.

4. A identificação das faixas etárias utilizando uma rede neural com abordagem wrapper comprovou a importância da extração dos parâmetros. Na classificação das vozes das mulheres, os parâmetros mais significativos foram o

Shimmer e o AQ. O Shimmer teve uma grande variação na fase idosa mas o AQ não. Como o AQ é uma medida que se relaciona com a pressão da voz e com o tipo de fonação dos locutores [63], pode ser o motivo da escolha como atributo significativo junto com o Shimmer. Os atributos mais significativos para identificar idade entre homens forem K0, DH12, NAQ, OQ1, OQ2, Oqa, SQ2. Analisando os gráficos vemos todos esses parâmetros como bons discriminantes. Os atributos DH12, NAQ, OQ1, OQ2, Oqa, SQ2 e HRF estão relacionados com a qualidade e a intensidade da voz, e K0 é um parâmetro que cresce significativamente no grupo dos idosos masculinos. Os parâmetros K0 e SQ2, escolhidos como bons discriminantes, mostraram uma grande dispersão ao longo do sinal para os homens. Temos outros parâmetros que, analisando os gráficos dos homens, mostram grandes variações na fase idosa mas não foram escolhidos como atributos mais significativos.

5. Os resultados foram satisfatórios e melhores que os obtidos em outros trabalhos, uma vez que utilizamos uma maior quantidade de parâmetros para a classificação das vozes. A utilização de parâmetros do sinal glótico, que sempre foram utilizados para identificar patologias diversas, mostraram-se bons discriminantes para caracterizar vozes envelhecidas.

## 6.2 Trabalhos Futuros

- Utilizar os parâmetros para a identificação de sexo dos locutores.
- Fazer análises utilizando outras vogais sustentadas.
- Proceder a mesma análise deste estudo, mas com vogais concatenadas.
- Aplicar os resultados com outras vogais sustentadas e também concatenadas em uma rede neural para identificar faixas etárias de locutores.

# Apêndice A

## Programa *Praat*

Ao iniciar o Praat duas janelas principais são mostradas no monitor - objetos e figuras. Informações importantes:

- Praat objetos (ver Fig.A.2) - encontram-se análises de sinais e sons. Todas as funções de gravação, reprodução, audição, desenho e edição também estão nesta janela.
- Praat figura - encontra-se desenho para a edição e adição de textos. É o único lugar que permite a impressão do material.

Para iniciar uma gravação (ver Fig.A.3) utilizamos a janela de objetos onde encontra-se a janela de gravação. Nesta, temos a opção de gravar, ouvir o material gravado, salvar canal esquerdo e direito, e escolher a taxa de amostragem do sinal. Após a escolha do tipo de áudio a ser gravado (mono ou estéreo), é necessário escolher a taxa de amostragem e efetuar a gravação através do botão de gravação. O nível do sinal gravado é acompanhado através do medidor de nível do sinal, evitando que haja distorção na gravação [32]. Para guardar o sinal, no menu *File* seleciona-se o formato desejado (*AIFC*, *NeXT/Sun*, *NIST*, *WAV*) e o diretório de gravação.

Utilizando o botão *Save to List*, coloca-se o sinal na lista de objetos atribuindo um nome ao ficheiro, e depois fecha-se o arquivo. O som vai aparecer na lista de objetos. Para imprimir ou salvar um arquivo de imagem, deve-se transferi-lo para a janela de figuras, onde todos os dados importantes (como, por exemplo, linhas, legendas, escalas, *grids*) ficam armazenados para análises futuras.

Através do menu *Edit*, o Praat oferece uma ferramenta muito flexível para visualizar, reproduzir e extrair informação de um arquivo de som [60]. No topo deste menu estão disponíveis as seguintes opções:

- *File* - permite a extração de partes de sinal para ficheiros individuais;
- *Edit* - permite fazer copy/paste de partes do sinal;
- *Query* - permite obter informação sobre a posição do cursor;
- *View* - permite selecionar os conteúdos da janela, controlar o zoom e a reprodução do ficheiro;
- *Select* - controla a posição do cursor;
- *Spectrum* - permite controlar as definições do espectrograma. O valor da frequência no ponto onde está colocado o cursor é indicado na parte esquerda da janela, na cor vermelha;
- *Pitch* - permite o controle das definições do *pitch*. Por definição, o controle do pitch é assinalado através de uma linha azul e o valor na posição do cursor é indicado na parte direita da janela, na cor azul;
- *Intensity* - permite o controle das definições da intensidade. Por definição, o contorno da intensidade é assinalado através de uma linha amarela e o valor

na posição do cursor é indicado na parte direita da janela, na cor verde.

- *Formant* - permite o controle das definições das formantes. Por definição, os formantes são assinaladas através das linhas vermelhas;
- *Pulses* - permite a extração de informação específica relacionada com a qualidade de voz (jitter e shimmer). Os ciclos de vibração das cordas são assinalados no sinal, com linhas azuis verticais.

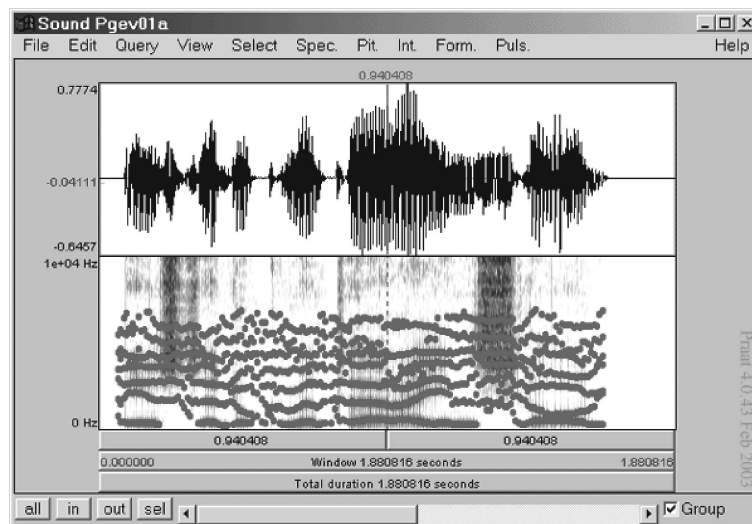


Figura A.1: Pulsos do sinal

Para iniciar o Praat basta clicar no ícone ou procurar nos executáveis (praat.exe), e para finalizar basta fechar o programa. Sempre que o programa for finalizado é necessário salvar para não perder dados (sempre ao fechar o programa pergunta se deseja salvar). Quando uma sessão termina, as janelas do programa figuras (*pictures*) e objetos (*objects*) são fechadas automaticamente. Quando a janela de objetos é fechada, o programa é finalizado, por isso é importante prestar atenção durante a execução. A finalização do programa não acontece quando fechada a janela de figuras [60].

Para uma análise da qualidade da voz, utiliza-se o menu *Pulse* para realizar um conjunto de medidas para quantificar irregularidades na frequência (*jitter*) e na amplitude (*shimmer*) de cada um os ciclos de vibração das cordas vocais. No menu *Pitch*, seleciona-se o *Extract visible pitch contour* para se obter informações sobre o sinal. Um novo ficheiro é apresentado na janela Praat Objetos. Na opção *info*, um nova janela é aberta com informações sobre valores médios e variação do som em análise: a variação da mediana, a distribuição dos valores à volta da mediana, etc [60]. Outra maneira de se obter uma descrição detalhada dos parâmetros de *jitter* e *shimmer* é selecionar *Voice Report* no menu *Edit-Pulse*, uma janela individual é exibida.

Ao iniciar o Praat duas janelas principais são mostradas no monitor - objetos e figuras. Informações importantes:

- Praat objetos (ver Fig.A.2) - encontram-se análises de sinais e sons. Todas as funções de gravação, reprodução, audição, desenho e edição também estão nesta janela.
- Praat figura - encontra-se desenho para a edição e adição de textos. É o único lugar que permite a impressão do material.

Para imprimir ou salvar um arquivo de imagem, deve-se transferi-lo para a janela de figuras, onde todos os dados importantes (como, por exemplo, linhas, legendas, escalas, *grids*) ficam armazenados para análises futuras.

### **Funções da janela de objetos:**

- *Sounds help* - ajuda com os sons capturados;



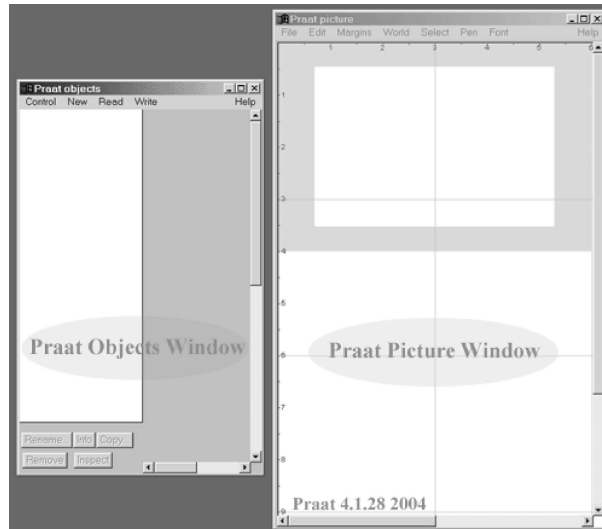


Figura A.2: Janela Objeto e Figura

- *Edit* - edição do sinal. Nesta função abre-se uma janela com o sinal e várias análises acústicas (Colar, cortar, deletar, etc...);
- *Play* - reprodução do sinal, para parar basta pressionar *Esc*;
- *Draw* - transfere a imagem para a janela de figuras para editar e imprimir;
- *Query* - abre um menu com várias opções, como por exemplo obter a duração do sinal;
- *Annotate* - cria um objeto para anotação;
- *Analyse* - funções diversas para análise do sinal: período, espectro, formantes, pontos do sinal e intensidade;
- *Manipulation* - cria um objeto para manipulação que contem o som, *pitch*, etc...;
- *Synthesize* - funções de conversão, filtros, “*enhance*” e combinações de sons;

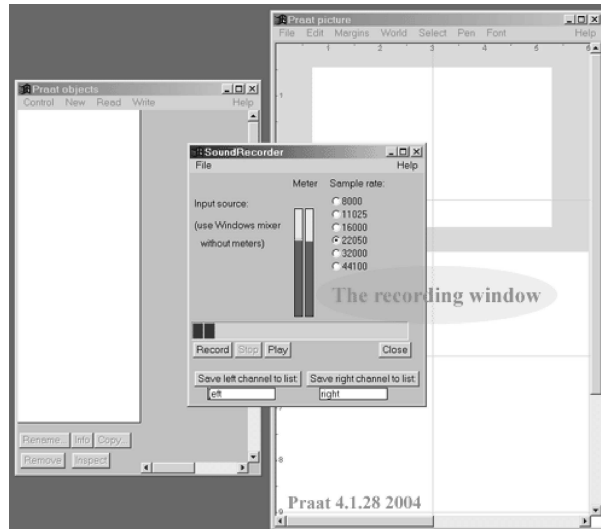


Figura A.3: Janela gravação

- *Rename* - seleciona-se o objeto desejado e altera-se o nome. Você não tem a opção de *undo*, mas pode fazer o procedimento novamente;
- *Copy* - seleciona-se o objeto a ser duplicado e executa-se a função *copy*, colocando um outro nome. O Praat permite que você tenha dois objetos com o mesmo nome. Pode-se renomear depois, caso haja necessidade;
- *Remove* - seleciona-se o objeto a ser apagado e executa-se a função *remove*. Não existe *undo* para esta função. Caso necessite-se deste objeto posteriormente, salva-se no *hard disk* antes de executar a função *remove*.

O sub-menu *Analyse* é um grupo de botões dinâmicos, que cria novos ficheiros, diferentes dos ficheiros de som:

- *Periodicity* - lida com o contorno de pitch: *Pitch analysis*, *Harmonicity* e *Glottal pulses*;
- *Spectrum* - relaciona todos os tipos de análises espectrais;

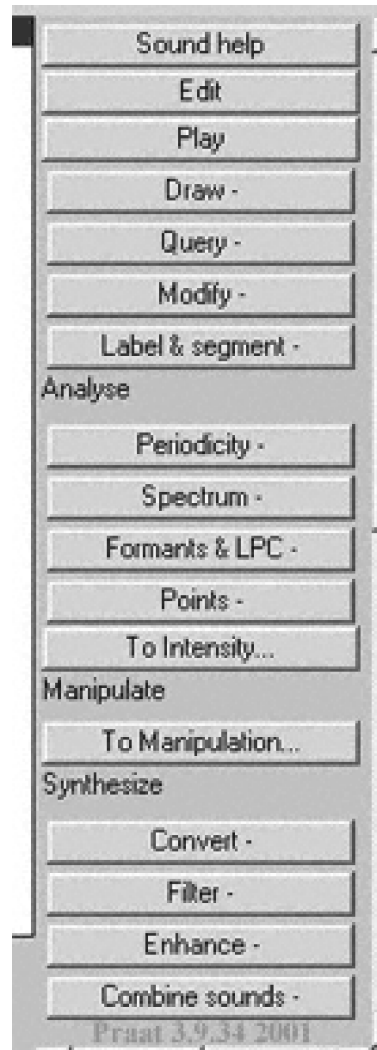


Figura A.4: Janela Itens Objetos

- *Formant/LPC* - cria o contorno das formantes e coeficientes de predição linear;
- *Intensity* - calcula o contorno da intensidade.

No topo deste menu estão disponíveis as seguintes opções:

- *File* - permite a extração de partes de sinal para ficheiros individuais;
- *Edit* - permite fazer copy/paste de partes do sinal;
- *Query* - permite obter informação sobre a posição do cursor;

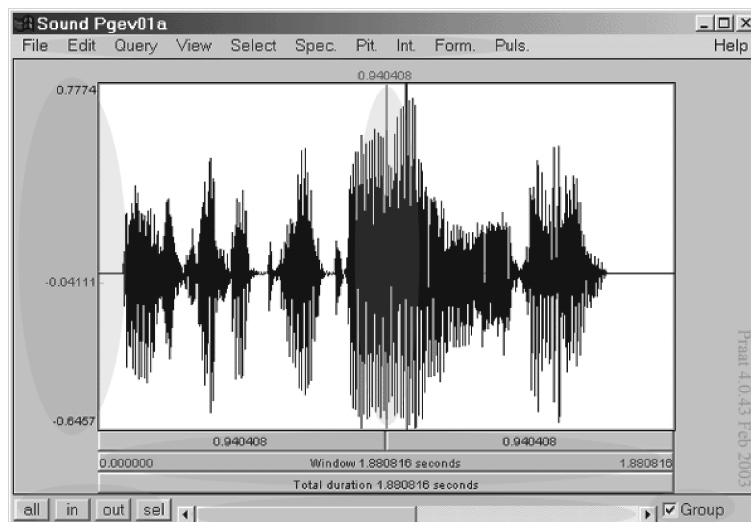


Figura A.5: Janela de Edição

- *View* - permite seleccionar os conteúdos da janela, controlar o zoom e a reprodução do ficheiro;
- *Select* - controla a posição do cursor;
- *Spectrum* - permite controlar as definições do espectrograma. O valor da frequência no ponto onde está colocado o cursor é indicado na parte esquerda da janela, na cor vermelha;
- *Pitch* - permite o controle das definições do *pitch*. Por definição, o controle do pitch é assinalado através de uma linha azul e o valor na posição do cursor é indicado na parte direita da janela, na cor azul;
- *Intensity* - permite o controle das definições da intensidade. Por definição, o contorno da intensidade é assinalado através de uma linha amarela e o valor na posição do cursor é indicado na parte direita da janela, na cor verde.
- *Formant* - permite o controle das definições das formantes. Por definição, os formantes são assinaladas através das linhas vermelhas;

- *Pulses* - permite a extração de informação específica relacionada com a qualidade de voz (jitter e shimmer). Os ciclos de vibração das cordas são assinalados no sinal, com linhas azuis verticais.
- Menu *Spectrum*;

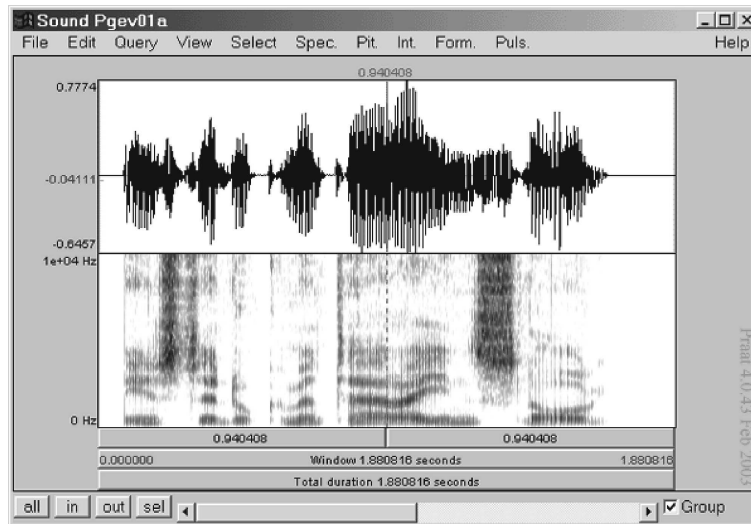


Figura A.6: Espectrograma do sinal

- Menu *Pitch*;

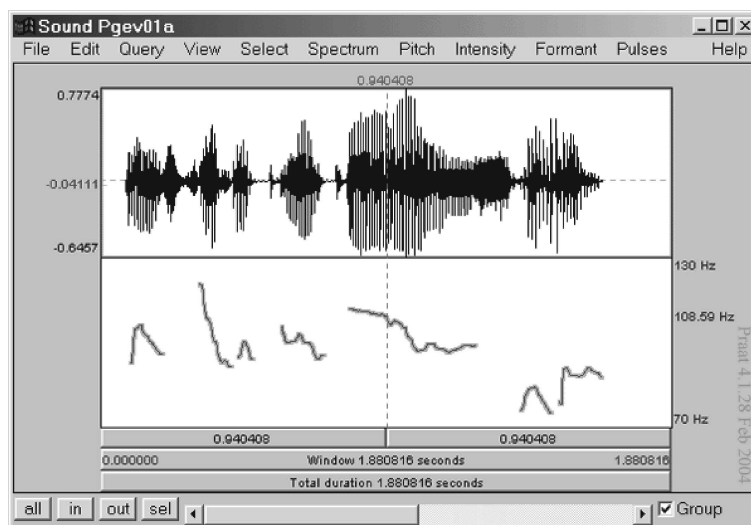


Figura A.7: Frequências do sinal

- Menu *Intensity*;

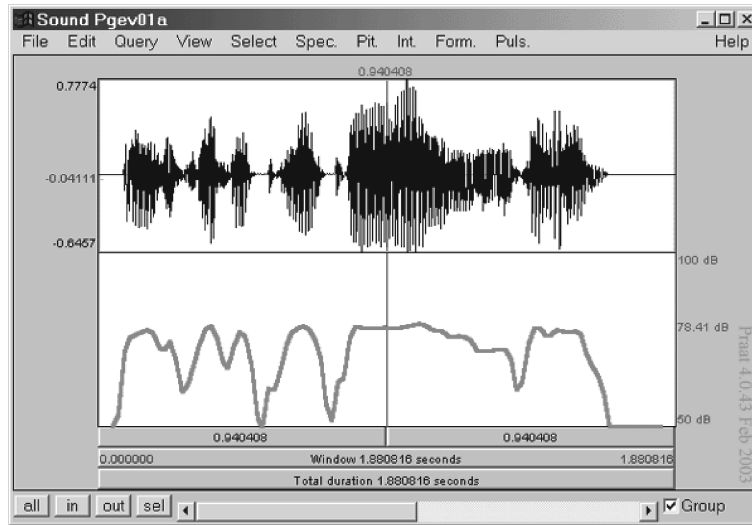


Figura A.8: Intensidade do sinal

- Menu *Formants*;

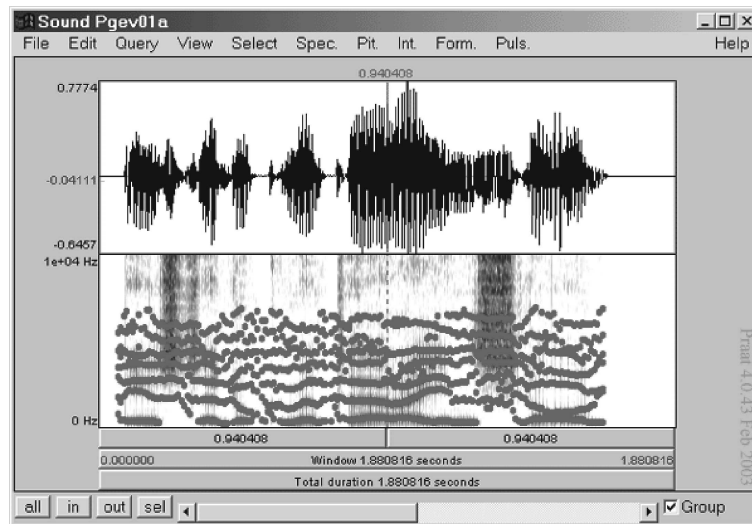


Figura A.9: Formantes do sinal

- Menu *Pulse*;

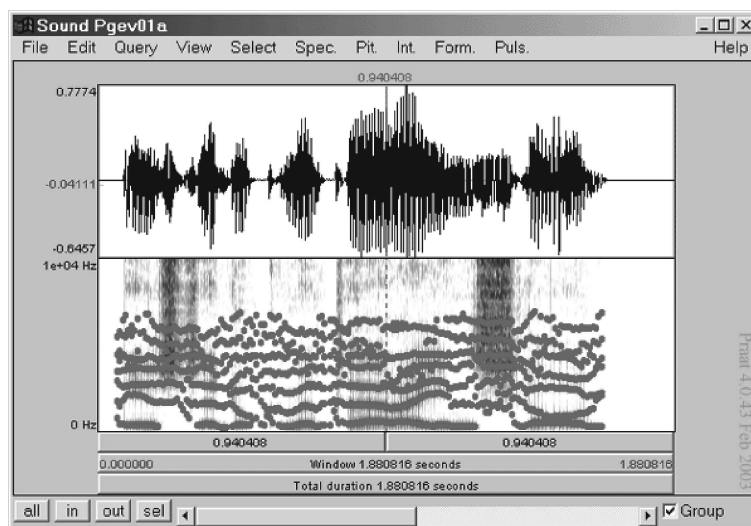


Figura A.10: Pulsos do sinal

# Apêndice B

## Programa *Aparat*

A plataforma principal de desenvolvimento é o *Linux* e necessita de pouco espaço do disco rígido [48]. A instalação em um ambiente *Matlab* necessita do *Matsig*, uma biblioteca de processamento de sinais que tem que ser instalado separado. A partir da versão 0.3.0 o pacote de instalação do *Aparat* já contém o *Matsig*. O *Aparat* é então executado através do comando ***aparat***. Pode-se ainda utiliza-lo fora do ambiente do *Matlab*, basta ter uma pasta com todos os arquivos do pacote de instalação e criar um atalho ***aparat.exe***.

Esta janela mostra diferentes parâmetros (ver Fig. 3.8) na base de tempo e uma vez selecionados, são demonstrados no gráfico do sinal na janela *Signal view*, parâmetros na base de frequência e modelo LF. Os parâmetros de tempo e frequência serão descritos de forma mais detalhada nas seções *glottaltimparams* e *glottalfreqparams* [48]. O LF-model simula a forma de onda ideal dentro do espectro selecionado. Essa função tem uma demanda computacional grande e o tempo de execução depende do computador utilizado.



## *Spectrum view* - Espectro do sinal

Esta janela mostra o espectro do sinal. Temos a possibilidade de escolher qual espectro desejamos visualizar, e cada um é designado com uma curva e cor diferente.

Segue o espectro do locutor 1 (Fig.3.10):

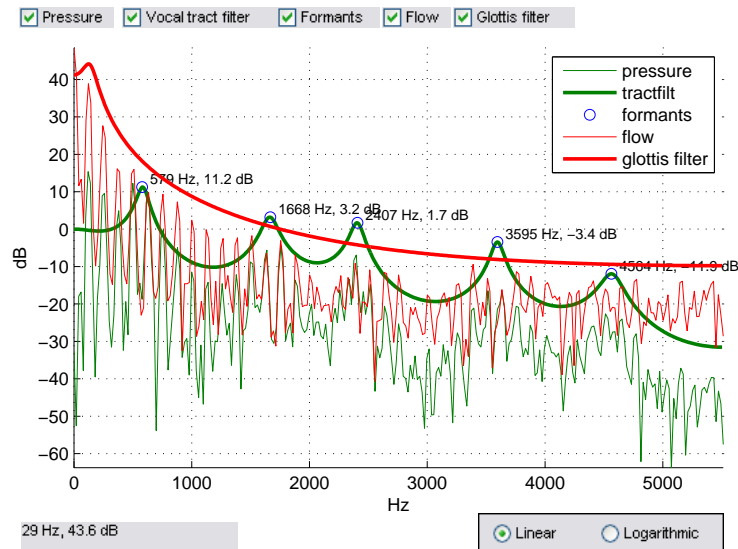


Figura B.1: Espectro Locutor 1

Podemos ainda visualizar as curvas linear ou logarítmica. As linhas fina verde e vermelha visualiza a pressão do sinal e o fluxo glotal, as linhas grossas verde e vermelha visualizam o filtro do trato vocal e o filtro da glote. Os picos do trato vocal também são mostrados como estimativas de formantes.

Aquisição de dados ou gravação dos sinais devem ser feitos separadamente.

Certos cuidados que deve-se ter na captação é:

- Utilizar microfones omnidirecional, evitando os dymanicos e os cardióides que tem características não -lineares de fase;
- Qualquer filtro no pré amplificador tem que ser desligado;

- Verificar que o dispositivo de gravação tem uma grandeza linear e resposta de frequência de fase ao longo da banda de voz;
- Não utilizar gravadores de MP3 ou minidiscos para a captação de voz;
- Se possível realizar as captações em câmara anecóica ou em ambientes com pouca reverberação. Ambientes com um alto tempo de reverberação podem influenciar na qualidade do sinal;
- Calibrar os equipamentos antes da gravação e não mexer até a finalização da captação;
- Manter as mesmas condições de gravação durante todo o processo de captação;

### Outras formas de visualização

Outras formas de visualização é através de gráficos denominados [48]: Plano-z, Trato vocal, Fases do sinal. Abaixo as figuras dos gráficos obtidos do locutor 1:

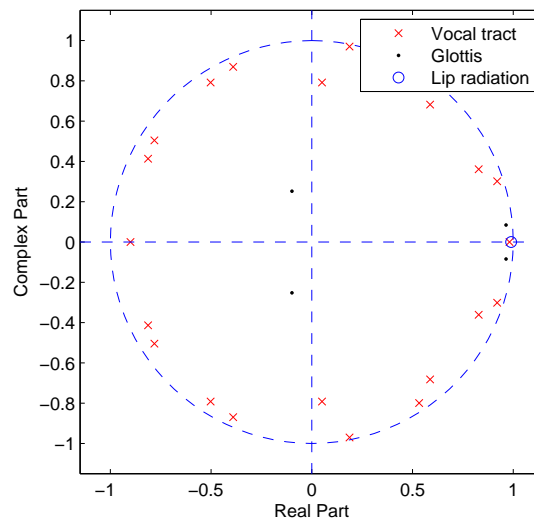


Figura B.2: Plano-z Locutor 1

O Plano-z visualiza os pólos do trato vocal e os filtros da glote no plano-z. Não deve ter mascasos haja pólos no eixo horizontal, estes podem ser retirados através do menu da janela principal [48].

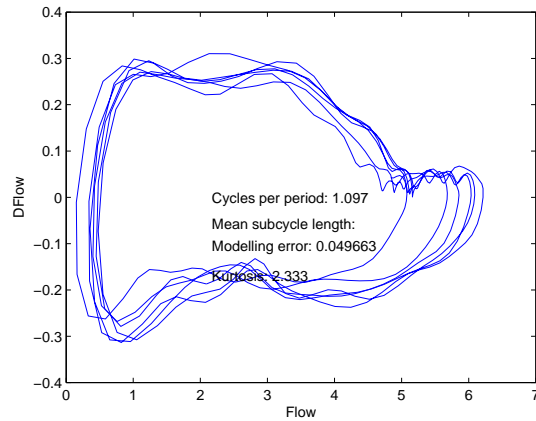


Figura B.3: Fase Locutor 1

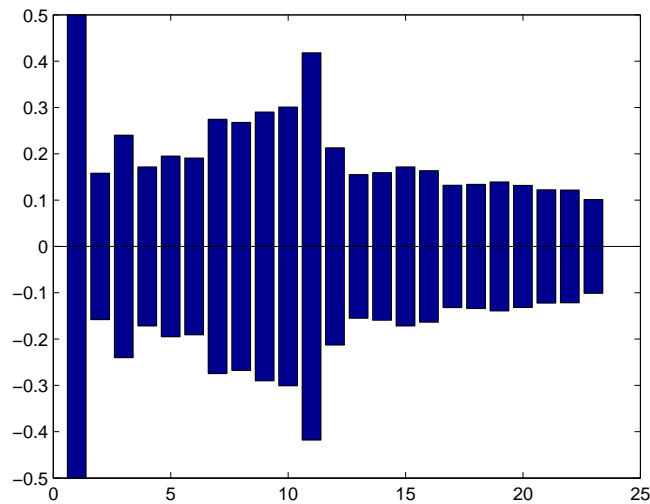


Figura B.4: Trato vocal Locutor 1

O gráfico do trato vocal visualiza a seção transversal do tubo do trato vocal. Neste gráfico obtemos informações para avaliar a qualidade da filtragem inversa. Qualquer ressonância no fluxo glotal é mostrado no plano de fase assim como

*loops* na trajetória, são mostrados à direita da figura. Se os *loops* foram pequenos ou não existiram, a filtragem inversa pode ser considerada bem sucedida.

### **Iaif.m**

Utilizando o IAIF extraímos diversos parâmetros relacionando a filtragem inversa do sinal. A fórmula para aquisição dos dados é [48]:

$$[G,HVT2,E,HG2,IAIFOPT]=IAIF(S1,FS,varargin)$$

onde:

- S1 é um vetor com o sinal de entrada;
- FS é a frequência de amostragem;
- Varargin é uma estrutura com opções que podem ser utilizadas para a obtenção do resultado;

obtem-se na saída:

- G é um objeto de sinal com o fluxo glotal;
- HVT2 é o modelo do trato vocal;
- E é o erro para o modelo final;
- HG2 é o modelo intermediário do fluxo glotal aproximado;
- IAIFOPT é uma estrutura com os parâmetros do modelo IAIF;

### **Glottalfreqparams.m**

Os parâmetros referentes à base de frequências, podem ser vistos neste arquivo *.mat*.

A análise consta de múltiplos parâmetros relacionados ao fluxo glotal. A fórmula para aquisição desses dados é [48]:

**Q=GLOTTALFREQPARAMS(G,F0)**

onde:

- G é um objeto de sinal com o fluxo glotal;
- F0 é a frequência fundamental do sinal;

obtem-se na saída Q que é uma estrutura com os seguintes dados:

- DH12 representa a diferença entre os dois primeiros formantes em decibéis;
- PSP é o parâmetro correspondente a um espectro parabólico polinomial de segunda ordem;
- HRF representa o fator de riqueza harmônica, o que significa a relação dos harmônicos superiores com o primeiro harmônico;

### **Glottaltimeparams.m**

Múltiplos parâmetros referentes à base do tempo são vistos neste arquivo *.mat*, e todos relacionados ao fluxo glotal. A fórmula para obter-se esses dados é [48]:

**Q=GLOTTALFREQPARAMS(G,F0)**

onde:

- G é um objeto de sinal com o fluxo glotal;
- F0 é a frequência fundamental do sinal;

obtem-se na saída Q que é uma estrutura com os seguintes dados:

- OQ1 e OQ2 representam os quocientes calculados a partir dos dois instantes de abertura do primeiro e segundo pulso glotal respectivamente;

- NAQ é o quociente da amplitude normalizada. É definido como a relação entre a amplitude pico a pico entre dois pulsos e normalizados em relação ao período de tempo. Está relacionada com a adução glotal;
- AQ é o quociente da amplitude;
- CLQ é o quociente de fechamento da glote;
- Oqa é a variação do quociente de abertura derivada do *LF-model*;
- **QQQ *Quasi-Open Quotient***,
- SQ1 e SQ2 - quocientes de velocidade entre a fase de abertura e fechamento da glote;
- Tpo - o tempo no instante de abertura do primeiro ciclo;
- tso - o tempo no instante de abertura do segundo ciclo;
- TC - o tempo no instante de fechamento;
- tmax - o tempo no instante do fluxo máximo;
- tmin - o tempo no instante do fluxo mínimo;
- tadmin - o tempo no instante mínimo da derivada;
- tadmin - o tempo no instante máximo da derivada;
- tqo - o tempo no "quasi" instante de abertura;
- tqc - o tempo no "quasi" instante do fechamento;

# Apêndice C

## Tabelas diversas

Tabela C.1: Informações cadastrais- Locutores

Informações cadastrais					
	Idade	Sexo	Altura	Cor	Fumantes
Locutor	61anos	Masc	1,60	Branca	Sim
Locutor	63anos	Masc	1,70	Branca	Sim(20 anos)
Locutora	60anos	Fem	1,60	Branca	Sim
Locutora	60anos	Fem	1,60	Branca	Não
Locutora	63anos	Fem	1,70	Branca	Não
Locutora	63anos	Fem	1,65	Branca	Não
Locutora	81anos	Fem	1,60	Branca	Não
Locutora	84anos	Fem	1,60	Branca	Não
Locutora	84anos	Fem	1,60	Branca	S/dados
Locutora	85anos	Fem	1,65	Branca	S/dados
Locutora	85anos	Fem	1,60	Branca	S/dados
Locutora	85anos	Fem	1,55	Branca	Não
Locutora	85anos	Fem	1,50	Branca	Não
Locutora	93anos	Fem	1,50	Branca	Não
Locutora	95anos	Fem	1,60	Negra	S/dados
Locutora	99anos	Fem	1,60	Branca	S/dados

Tabela C.2: Tabela de *Jitter*

<b>Jitter</b>					
Vogal A	local	local-abs	rap	ppq5	ddp
Locutora(60 anos)	0.729%	52.4E-6 seg	0.163%	0.257%	0.488%
Locutora(60 anos)	0.886%	68.9E-6 seg	0.420%	0.458%	1.259%
Locutora(63 anos)	2.385%	120.3E-6 seg	1.368%	1.511%	4.103%
Locutora(63 anos)	0.139%	7.74E-6 seg	0.070%	0.081%	0.210%
Locutora(81 anos)	0.609%	30.6E-6 seg	0.322%	0.330%	0.965%
Locutora(84 anos)	3.893%	274.8E-6 seg	2.151%	1.380%	6.453%
Locutora(84 anos)	0.438%	19.4E-6 seg	0.225%	0.243%	0.676%
Locutora(85 anos)	0.905%	55.3E-6 seg	0.511%	0.399%	1.532%
Locutora(85 anos)	3.517%	224.9E-6 seg	1.987%	2.324%	5.962%
Locutora(85 anos)	0.173%	8.22E-6 seg	0.080%	0.089%	0.240%
Locutora(85 anos)	0.803%	37.6E-6 seg	0.442%	0.427%	1.325%
Locutora(93 anos)	0.766%	36.9E-6 seg	0.301%	0.290%	0.903%
Locutora(95 anos)	1.728%	102.2E-6 seg	0.863%	0.901%	2.589%
Locutora(99 anos)	0.744%	33.3E-6 seg	0.300%	0.512%	0.899%



Tabela C.3: Tabela de *Shimmer*

Tabela 5 - Shimmer						
Vogal A	local	local-db	apq3	apq5	apq11	dda
Locutora(60 anos)	2.282%	0.213db	0.734%	0.840%	1.846%	2.201%
Locutora(60 anos)	3.911%	0.382db	0.710%	0.843%	1.134%	2.130%
Locutora(63 anos)	7.401%	0.814db	3.630%	4.147%	4.092%	10.890%
Locutora(63 anos)	0.824%	0.073db	0.392%	0.528%	0.635%	1.175%
Locutora(81 anos)	2.687%	0.237db	1.500%	1.592%	2.014%	4.501%
Locutora(84 anos)	6.109%	0.517db	3.314%	3.683%	5.392%	9.942%
Locutora(84 anos)	3.939%	0.379db	2.162%	2.441%	2.825%	6.486%
Locutora(85 anos)	4.158%	0.372db	2.437%	2.417%	2.653%	7.311%
Locutora(85 anos)	12.238%	1.186db	7.044%	8.423%	11.351%	21.131%
Locutora(85 anos)	2.785%	0.250db	1.504%	1.623%	2.297%	4.512%
Locutora(85 anos)	7.965%	0.708db	4.444%	3.948%	4.612%	13.332%
Locutora(93 anos)	12.833%	1.115db	3.890%	8.882%	23.695%	11.670%
Locutora(95 anos)	6.947%	0.691db	2.994%	3.661%	6.466%	8.982%
Locutora(99 anos)	6.223%	0.555db	2.858%	4.112%	6.074%	8.575%

Tabela C.4: Tabela de *HNR*

<b>Tabela 6 - HNR</b>	
Vogal A	HNR
Locutora(60 anos)	18.837dB
Locutora(60 anos)	21.780dB
Locutora(63 anos)	15.013dB
Locutora(63 anos)	30.261dB
Locutora(81 anos)	18.090dB
Locutora(84 anos)	11.511dB
Locutora(84 anos)	18.743dB
Locutora(85 anos)	19.378dB
Locutora(85 anos)	6.052dB
Locutora(85 anos)	22.847dB
Locutora(85 anos)	13.845dB
Locutora(93 anos)	14.406dB
Locutora(95 anos)	13.471dB
Locutora(99 anos)	15.336dB

Tabela C.5: Tabela de Frequência Fundamental

<b>Tabela 7 - Frequência Fundamental</b>					
Vogal A	Median	Mean	Stand.Devia.	Minimum	Maximum
Locutora(60 anos)	139.6Hz	138.8Hz	5.118Hz	120.4Hz	143.6Hz
Locutora(60 anos)	128.9Hz	128.9Hz	1.531Hz	123.9Hz	134.2Hz
Locutora(63 anos)	200.8Hz	197.5Hz	17.041Hz	90.6Hz	212.2Hz
Locutora(63 anos)	179.0Hz	178.9Hz	1.217Hz	175.2Hz	189.4Hz
Locutora(81 anos)	200.1Hz	198.0Hz	13.365Hz	101.9Hz	215.7Hz
Locutora(84 anos)	142.5Hz	142.8Hz	5.260Hz	136.0Hz	150.7Hz
Locutora(84 anos)	226.9Hz	225.9Hz	8.038Hz	202.3Hz	245.7Hz
Locutora(85 anos)	162.7Hz	163.5Hz	5.553Hz	155.5Hz	179.8Hz
Locutora(85 anos)	126.6Hz	158.8Hz	68.610Hz	72.6Hz	304.1Hz
Locutora(85 anos)	128.0Hz	147.8Hz	45.700Hz	70.9Hz	254.0Hz
Locutora(85 anos)	210.4Hz	210.5Hz	2.131Hz	205.7Hz	215.8Hz
Locutora(93 anos)	209.7Hz	208.0Hz	7.420Hz	198.7Hz	216.9Hz
Locutora(95 anos)	171.6Hz	169.4Hz	22.824Hz	70.9Hz	206.4Hz
Locutora(99 anos)	229.7Hz	223.4Hz	14.600Hz	188.1Hz	237.4Hz

Tabela C.6: Parâmetros Locutora 99 anos

Tabela 1 - Locutora 99 anos					
Vogais	A	E	I	O	U
Med. Pitch	229.710Hz	188.533Hz	220.261Hz	204.431Hz	223.738Hz
Mean Pitch	223.442Hz	172.296Hz	212.638Hz	198.957Hz	215.223Hz
Stand. Deviat.	14.600Hz	39.717Hz	24.406Hz	30.188Hz	26.014Hz
Minimum Pitch	188.084Hz	71.220Hz	124.996Hz	113.630Hz	153.237Hz
Maximum Pitch	237.386Hz	209.725Hz	239.287Hz	237.987Hz	245.124Hz
Number/Pulses	228	318	341	393	341
Number/Periods	227	316	340	392	340
Mean Period	4.4745E-3s	5.7912E-3s	4.7125E-3s	5.0255E-3s	4.6418E-3s
Standard Dev.Period	0.3032E-3s	1.7550E-3s	0.6317E-3s	0.8320E-3s	0.5968E-3s
Unvoiced Frames	1.905%	0	0.617%	0.500%	0
Number Voice/Break	0	0	0	0	0
Degree Voice/Break	227	0	0	0	0

# Bibliografia

- [1] Airas M., “ Methods and Studies fo Laryngeal Voice Quality Analysis in Speech Production” ,Helsinki University of Technology Department of Signal Processing and Acoustics, Espoo 2008 - Dissertation for the degree of Doctor of Science in Technology
- [2] Ahlfinger R., Cheeseman B., Doody P., “Harmonic Detection”, Version 1.5: Aug 12, GMT-5, produced by The Connexions Project, 2005
- [3] Alipour F., Scherer R., Knowles J., “Perceived Pitch of Whispered Vowels”, Journal of Voice, vol.10, No.2, pp.115-158, 1996.
- [4] Alipour F., Scherer R., Knowles J., “Velocity Distributions in Glottal Models”, Journal of Voice, vol.10, No.1, pp.50-58, 1996.
- [5] Albers M.J., “Signal to Noise Ratio”, Department of English, University of Memphis, Memphis, TN 38152, 901-678-4776, 2004.
- [6] Akande O., Murphy P., “Evaluation of Cepstrally derived Harmonics-to-Noise Ratios in Voice Signal”, Department of Electronics and Computer Engineering - University of Limerick, Limerick, Ireland.

- [7] Araujo S.A., Grellet M., Pereira J.C., Rosa M.O., “Normatização de Medidas Acústicas da Voz Normal”, Revista Brasileira Otorrinolaringologia, vol.68, No.4, pp.540-4, jul/ago, 2002.
- [8] Austin S.F., Titze I.R., “The Effect of Subglottal Resonance upon Vocal Fold Vibration”, Journal of Voice, vol.11, No.4, pp.391-402, 1997.
- [9] Beauvoir S., “A Velhice”, Editora Nova Fronteira - Rio de Janeiro, 1990.
- [10] Beauvoir S., “A Velhice Realidade Incômoda”, Difusão Européia do Livro - Volume 1 - São Paulo, 1976.
- [11] Behlau M.S., Tosi O., Lima P.A., “Determinação da frequência fundamental e suas variações em altura e intensidade, para falantes do português brasileiro”, Biblioteca Virtual de Saúde - Acta AWHO. n.4, p. 5-9, 1985.
- [12] Behrman A., Agresti C.J., Blumstein E., Lee, N., “Microfone and Eletroglottographic Data from Dysphonic Patients”, Journal of Voice, vol.12, No.2, pp.249-260, 1998.
- [13] Budant T.C. “Estudo da Variação dos parâmetros Acústicos da Voz para o Estabelecimento de padrão de Normalidade em Adolescentes”, Dissertação de Mestrado - CEFET-PR - Novembro 2002.
- [14] Burnett T.A., Senner J.E., Larson C.R., “Voice F0 Response to Pitch-Shifted Auditory Feedback”, Journal of Voice, vol.11, No.2, pp.202-211, 1997.
- [15] Carole F.T., “Harmonics-to-Noise Ratio: An Index of Vocal Aging”, Journal of Voice, vol.16, No.4, pp.480-487, 2002.

- [16] Donald V.P., Pretty B.R., “Effects of Microphone Type on Acoustic Measures of Voice”, *Journal of Voice*, vol.15, No.3, pp.331-343, 2001.
- [17] Cataldo E., “Modelagem de incertezas presentes em um modelo biomecânico para a produção de vogais” - DEM PUC-Rio, 2007.
- [18] Cataldo E., Sampaio R., Nicolato L., “Uma discussão sobre modelos mecânicos de laringe para síntese de vogais”, *Engevista*, 2004.
- [19] Cataldo E., Sampaio R., Soize C., Delceliers C. “Probabilistic Modeling of a nonlinear dynamical system used for producing voice”, *Comput Mech* - May 2008.
- [20] Cataldo E., Sampaio R., Soize C., Lucero J. “Modeling Random Uncertainties in Voice Production using a Parametric Approach”, *Mechanics Research Communications* - Jun 2008.
- [21] Dias R.S., “Implementação de um sistema para análise e síntese das vogais”, Universidade Federal do Pará, Tese de Mestrado, 1997.
- [22] Eadie T.L., Doyle P.C., Hansen K., Beaudin P.G., “Influence of Speaker Gender on Listener Judgments of Tracheoesophageal Speech”, *Journal of Voice*, vol.22, No.1, 2008.
- [23] Estrela F., “Análise Acústica da Voz de Idosos”, Monografia ao Centro de Estudos da Voz - CEV (Centro de Especialização de Voz) - São Paulo, 2000.
- [24] Fant G., Liljencrants J., Lin Q., “A four parameter model of glottal flow”, *Speech Transmission Laboratory Quarterly Progress and Status Report*, Royal Institute of Technology, Stockholm - Sweden, vol. 4, pp. 1-13, 1985.

- [25] Fiqueredo L.C., Goncalves M.I.R., Pontes A., Pontes P., “Estudo do comportamento Vocal no Ciclo Mestruar”, Revista Brasileira de Otorrinolaringologia, vol.70, Ed.3, Maio-Junho de 2004.
- [26] Filho L.F., “O processo de envelhecimento e o comportamento vocal”, CEFAC (Centro de Especialização em Fonoaudiologia Clínica) - Monografia, São Paulo 1999.
- [27] Forero L., “Redes Neurais e Máquinas de Vetores de Suporte no Reconhecimento de Locutor usando coeficiente MFC e características do sinal glotal”, Dissertação de Mestrado - Universidade Federal Fluminense - UFF, 2009.
- [28] Forero L., “Identificação de faixas etárias através características do sinal glotal com redes neurais e seleção de atributos wrapper híbrido RNA e AG”, Universidade Católica Rio de Janeiro - PUC-Rio, Artigo não publicado, 2009.
- [29] Gallivan G.J., Gallivan H.K., Eitnier C.M., “Dual Intracordal Unilateral Voceal Folds Cysts”, Journal of Voice, vol.22, No.1, pp.119-124, 2008.
- [30] Gislene L.N.M., Vicente L.C.C., “Os efeitos do climatério na frequência fundamental”, CEFAC (Centro de Especialização em Fonoaudiologia clínica), Monografia, São Paulo 1998.
- [31] Gottliebson R.O., Lee L., Weinrich B., Sanders J., “Voice Problems of Future Speech-Language Pathologist”, Journal of Voice, vol.21, No.6, pp.669-704, 2007.
- [32] Gut U., “Praat Manual” web page - <http://www.fon.hum.uva.nl/praat>, 2008 .
- [33] Hancock J., “Jitter Fundamentals”, High Frequency Electronics, Copyright © Summit Technical Media, LLC - 2004.



- [34] Harnsberger J.D., Shrivastav R., Junior W.S.B., Rothman H., Hollien H., “Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age”, *Journal of Voice*, vol.22, No.1, pp.58-69, 2008.
- [35] Haykin S., “Redes Neurais - Princípios e prática”, Artmed Editora - 1999.
- [36] Henrich N., Bailly L., Webb M., Muller F., Licht A.K., Hess M., “Exploration of vocal-folds and ventricular-bands interaction in singing using high-speed cinematography and electroglottography”, *International Congress on Acoustics*, 2007.
- [37] Henrich N., Lamesch S., Castellengo M., “Vocal registers and laryngeal mechanisms”, *International Congress on Acoustics*, 2007.
- [38] Henrich N., Bailly L., Ruty N., Van Hirtum A., Cisonni J., Pelorson X., “Aerodynamic interaction between the vocal folds and the ventricular bands”, *International Congress on Acoustics*, 2001.
- [39] Henrich N., d’Alessandro C., Doval B., “The voice source as a causal/anticausal linear filter”, *Voqual’03 - Geneva*, August 27-29 - 2003.
- [40] Henrich N., Sundin G., Ambroise D., d’Alessandro C., Castellengo M., Doval B., “Just noticeable differences of open quotient and asymmetry coefficient in singing voice”, *Journal of Voice*, vol.17, No.4, pp.481-494, 2002.
- [41] Henrich N., Barney A., De Stefano A., “The effect of glottal opening on the acoustic response of the vocal tract”, *Forum Acusticum*, 2005.
- [42] Henrich N., d’Alessandro C., Castellengo M., Doval B., “Glottal open quotient in singing”, *J. Acoust. Soc. Am.*, vol. 117, Pt. 1, March 2005.

- [43] Henrich N., Roubeau B., Castellengo M., “On the use of electroglottography for characterisation of the laryngeal mechanisms”, Proceedings of the Stockholm Music Acoustics Conference, August 6-9, (SMAC 03), Stockholm, Sweden - 2003
- [44] Henrich N., d’Alessandro C., Castellengo M., Doval B., “On the use of the derivative of eletroglottographic signals for characterization of nonpathological phonation”, Acoustical Society of America, 2004.
- [45] Henrich N., d’Alessandro C., Doval B., “Glottal flow models: waveforms, spectra and physical measurements”, LIMSI - CNRS, BP 133, Université Paris XI, F-91403 Orsay, France - 2004.
- [46] Higashikawa M., Nakai K., Takahashi H., “Perceived Pitch of Whispered Vowels”, Journal of Voice, vol.10, No.2, pp.155-158, 1996.
- [47] Hirano M., Bless D.M., “Exame Videostroboscópio da Laringe”, Editora Artes Médicas, Porto Alegre 1997.
- [48] HUT-Helsinki University of Technology - web page - <http://www.acoustics.hut.fi/software/>.
- [49] Jennings J.J., Kuehn D.P., “Loudness Level and Gender on Nasalance”, Journal of Voice, vol.22, No.1, pp.75-89, 2008.
- [50] Junior D.B., Heuer R.J., Satloff R.T., Hills J.R., Cater J.R., “Intrasubject Variability of Objective Voice Measures”, Journal of Voice, vol.10, No.2, pp.166-174, 1996.

- [51] Latsch V.L., “Um sistema de conversão texto-fala para Windows” - UFRJ - Escola de Engenharia - Departamento de Eletrônica e Computação - Agosto 2002.
- [52] Linville S.E., “The Sounds of Senescence”, *Journal of Voice*, vol. 10, No. 2, pp. 190-200, Lippincott-Raven Publishers, Philadelphia 1996.
- [53] Lopes J., Freitas S., Sousa R., Matos J., Abreu F., Ferreira A., “A medida HNR: sua relevancia na análise acústica da voz e sua estimação precisa”, Faculdade de Engenharia - Universidade do Porto - Artigo não publicado.
- [54] Kaplan H.I., Sadock B.J., “Compêndio de Psiquiatria”, Editora Artes Médicas - Porto Alegre, 1990.
- [55] Kent R.D., Read C., “The Acoustic Analysis of Speech”, Published by Singular Publishing Group Inc., SanDiego - California,USA - 1992.
- [56] Kolokolov A.S., “Measuring the Fundamental Tone of Voice Signal”, *Automation and Remote Control*, Vol. 64, No. 8, pp. 1310-1320. Translated from *Avtomatika i Telemekhanika*, No. 8, pp. 122-134 - 2003.
- [57] Mattos J. S., “Um estudo comparativo entre o sinal electroglotográfico e o sinal de voz”, Dissertação de mestrado em Engenharia de Telecomunicações, UFF 2008.
- [58] Netto S.L., Latsch V.L., “Obtenção de marcas de pitch em sinais de voz para síntese por concatenação Temporals”, IX Convenção Nacional da Sociedade de Engenharia de Áudio, São Paulo, 11-13 de abril de 2005.

- [59] Noriega L., “Multilayer Perceptron Tutorial”, School of Computing Staffordshire University - November 17, 2005.
- [60] Oliveira C., Teixeira A., “Praat”, Escola Superior de Saúde da Universidade de Aveiro - Licenciatura em Terapia da Fala, Artigo não publicado.
- [61] Pappa G.L., “Seleção de atributos utilizando algoritmos genéticos múltiplos objetivos”. Tese de mestrado - PUC Paraná - 2005
- [62] Perello J., Miguel J.A.S., “Alteraciones de la Voz”, Editora Científico Médica, 1980.
- [63] Pulaka H., “ Analysis of Human Voice Production Using Inverse Filtering, High-Speed Imaging, and Electrolottography”, Helsinki University of Technology, Dept. of Computer Science and Engineering - Master’s Thesis, 2005.
- [64] Rabiner L., Juang B.H. “Fundamentals of Speech Recognition”, Prentice-Hall, 1993.
- [65] Santos I., “Análise acústica da voz de indivíduos na terceira idade”, Dissertação de Mestrado - Universidade de São Carlos - USP, 2005.
- [66] Santos I., Vanzella T.P., Pereira J.C., “Comparação dos espectros vocálicos de crianças, adultos e idosos”, XV Congreso Argentino de Bioingeniería.
- [67] Santoro D.M., “Sobre o processo de seleção de subconjuntos de atributos - As abordagens filtro e Wrapper”, Dissertação de Mestrado - Universidade Federal de São Carlos - Abril 2005.

- [68] Shama K., Krishna A., Cholayya N.U., “Study of HNR and Critical Band Energy”, Hindawi Publishing Corporation - EURASIP Journal on Advances in Signal Processing - Volume 2007, Article ID 85286, 9 pages - 2007.
- [69] Shaw H.S., Deliyski D.D., “Mucosal Wave”, Journal of Voice, vol.22, No.1, pp.23-33, 2008.
- [70] Sonninem A., Hurme P., Laukkanen A., “The External Frame Function in the Control of Pitch”, Journal of Voice, vol.13, No.3, pp.319-340, 1999.
- [71] Steinhauer K., Grayhack J.P., Smiley-Oyen A.L., Shaiman S., McNeil M.R., “Perceived Pitch of Whispered Vowels”, Journal of Voice, vol.10, No.2, pp.155-158, 2004.
- [72] Steinhauer K., Grayhack J.P., Smiley-Oyen A.L., Shaiman S., McNeil M.R., “The Relationship Among Voice Onset”, Journal of Voice, vol.18, No.4, pp.432-442, 2004.
- [73] Titze I.R., “Principles of Voice Production”, Prentice-Hall, Englewood Cliffs, New Jersey - 1994.
- [74] Titze I.R., “Comments on the myoelastic-aerodynamic theory of phonation”, J. Acoust. Soc. Am., vol.23, pp.495-510, 1980.
- [75] Titze I.R., “Workshop on Acoustic Voice Analysis”, National Center for Voice and Speech, 1994.
- [76] Tonisi G., “Os efeitos do Climatério na Frequência Fundamental”, CEFAC (Centro de Especialização em Fonoaudiologia clínica) - Monografia, 1998.

- [77] Vellasco M., “Redes Neurais Artificiais”- Laboratório de Inteligência Computacional Aplicada - PUC-Rio - Artigo não publicado - 2007.
- [78] Winholtz W.S., Titze I.R., “Suitability of Minidisc Recordings”, *Journal of Voice*, vol.12, No.2, pp.138-142, 1998.
- [79] Zhang Y., Jiang J.J., “Acoustic Analyses of Sustained from Patients with Laryngeal Pathologies”, *Journal of Voice*, vol.22, No.1, pp.1-9, 2008.