Análise Acústica da Voz para Pré-diagnóstico de Patologias da Laringe

Marcelo de Oliveira Rosa

Dissertação apresentada à Escola de Engenharia
de São Carlos, da Universidade de São Paulo,
como parte dos requisitos para obtenção do título
de Mestre em Engenharia Elétrica.

ORIENTADOR: Prof. Dr. José Carlos Pereira

DEDALUS - Acervo - EESC

31100017207

São Carlos - SP
1998

FOLHA DE APROVAÇÃO

Candidato: Engenheiro **MARCELO DE OLIVEIRA ROSA**

Dissertação defendida e aprovada em 09-03-1998
pela Comissão Julgadora:

Prof. Doutor **JOSÉ CARLOS PEREIRA (Orientador)**
(Escola de Engenharia de São Carlos - Universidade de São Paulo)

Prof. Associado **MARCOS GRELET**
(Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo)

Prof. Doutor **MARCELO BASÍLIO JOAQUIM**
(Escola de Engenharia de São Carlos - Universidade de São Paulo)

Prof. Doutor **DENIS MINICIUS COURY**
Coordenador da Área de Engenharia Elétrica

JOSÉ CARLOS A. CINTRA
Presidente da Comissão de Pós-Graduação da EESC

Ao Senhor Deus e à minha família
(José, Josefa *in memoriam*, Paulina e Fernando),
por tudo que Ele e eles representam para mim.

Agradecimentos

Ao Prof. Dr. José Carlos Pereira pela amizade, pela orientação na área de processamento digital de sinais, e por ter me aceito como seu orientado.

Ao Prof. Dr. Marcos Grellet do departamento de Otorrinolaringologia, pela amizade e auxílio na realização desta pesquisa, no Hospital das Clínicas de Ribeirão Preto.

Aos professores Carlos Magno Corrêa Dias e José Eloir K., que me incentivaram à realizar este mestrado durante o período de graduação.

Ao Prof. Dr. Marcelo Basílio Joaquim, pelas conversas que tivemos, desde alguns e-mails perdidos até sugestões e ajudas em processamento digital de sinais.

À amiga e médica Simone Adad Araújo, pelas longas e prazerosas conversas, questionamentos sobre mestrado, além do grande auxílio em exames clínicos das larínges.

Aos amigos, médicos residentes e professores dos setores de Otorrinolaringologia, Fonoaudiologia e Cirurgia de Cabeça e Pescoço do H. C. de Ribeirão Preto pelo auxílio na realização dos exames dos pacientes, sem os quais não seria possível a realização desta pesquisa.

Aos meus amigos de Curitiba (Acyr, Luis de Bona, Marcia, Luis Cláudio, Hubert e família, Diogo, Andréia, Emile, Juliano, Bea, Alessandra, Emersom) por sempre me receberem durante os breves momentos em que estive nesta “minha” cidade durante os dois anos de mestrado.

Aos meus amigos de república (Rene, Ivan, Daniel, André e André “Bixo”) pela amizade e por terem me “aguentado” durante o segundo ano do mestrado.

Aos meus amigos de laboratório, mestrado e de outras circunstâncias (Adriana, Alessandra, Ana Elisa, André Murta, Arthur, Berenice, Cris, dona Maria, Deise, Denise, Denise, Donato, Fabiane, Fabiano, Fábio, Fátima, Guilherme, Hélio, Hérika, João Bosco, João Paulo, Joseane, Juliana, Lauro, Lúcia e seus tios e primos de Manaus, Luciana, Magaly, Marcelo Vieira, Marcia, Marcos, Maria Virgínia e família, Mário, Maristela, Miriam, Mori, Patrícia, Renata, Renato, Ricardo e Cristina, Roseli, Rudinei, Sérgio, Thais, Vagner, Yog) pela amizade que permitiu que os momentos difíceis desta pesquisa tenham sido completamente suplantados.

Aos meus queridos amigos adventistas de São Carlos (particularmente Roberto e Erly, e Wellington, Liliana, e suas filhas) que me receberam em seus lares e compartilharam comigo minhas alegrias e dificuldades.

À Capes pela concessão de auxílio financeiro para o desenvolvimento da dissertação.

E a tantos outros que não foram esquecidos, mas não citados por falta de espaço na folha.

Sumário

LISTA DE FIGURAS	I
LISTA DE TABELAS	II
LISTA DE SÍMBOLOS	VII
RESUMO	X
ABSTRACT	XI
1. INTRODUÇÃO	1
2. FUNDAMENTAÇÃO TEÓRICA	6
2.1 VOZ E TRATO VOCAL	6
2.1.1 RESPIRAÇÃO	7
2.1.2 VOCALIZAÇÃO	9
2.1.3 RESSONÂNCIA	11
2.2 PATOLOGIAS DO TRATO VOCAL	12
2.2.1 ÚLCERAS DE CONTATO	15
2.2.2 DIPLOFONIA	15
2.2.3 FALSETE	15
2.2.4 AFONIA FUNCIONAL	16
2.2.5 DISFONIA FUNCIONAL	16
2.2.6 QUEBRAS DE VOCALIZAÇÃO	16
2.2.7 QUEBRAS DE ALTURA	16
2.2.8 DISFONIA ESPASMÓDICA	17
2.2.9 ESPESSEAMENTO DE PREGAS VOCAIS	17

2.2.10 LARINGITE TRAUMÁTICA	17
2.2.11 DISFONIA VENTRICULAR	17
2.2.12 NÓDULOS	17
2.2.13 CISTOS	18
2.2.14 PÓLIPOS	18
2.2.15 EDEMA DE REINKE	18
2.2.16 CÂNCER	18
2.2.17 DISARTRIA	19
2.2.18 MUDANÇAS ENDÓCRINAS	19
2.2.19 GRANULOMA	19
2.2.20 HEMANGIOMA	19
2.2.21 HIPERCERATOSE	19
2.2.22 LARINGITE INFECCIOSA	20
2.2.23 LEUCOPLASIA	20
2.2.24 PAILOMA	20
2.2.25 PARALISIA DE PREGAS VOCAIS	20
2.2.26 MEMBRANA LARINGIANA	20
2.3 MODELO MATEMÁTICO DO TRATO VOCAL	22
2.3.1 EXCITAÇÃO GLOTTAL E CONSIDERAÇÕES FÍSICAS ASSOCIADAS	23
2.3.2 SISTEMA LINEAR VARIANTE NO TEMPO - SUPRAGLOTTAL	24
2.3.3 MODELO ACÚSTICO DA FALA	25
2.3.4 MODELO COMPLETO	30
2.4 ELEMENTOS MATEMÁTICOS DA ANÁLISE ACÚSTICA	33
2.4.1 JANELAS	33
2.4.2 ESPECTROGRAMAS	35
2.5 MEDIDAS DE PERTURBAÇÃO	38
2.5.1 MÉTODOS DE AQUISIÇÃO DO SINAL DE VOZ	38
2.5.2 ESTIMATIVA DO PERÍODO E AMPLITUDE DA FREQUÊNCIA FUNDAMENTAL	39
2.5.3 JITTER E SHIMMER	41
2.6 MEDIDAS DE RUÍDO ESPECTRAL	47
2.7 AVALIAÇÃO POR RESÍDUOS	57
2.8 OUTRAS TÉCNICAS ACÚSTICAS	60
3. DISCRIMINAÇÃO PATOLÓGICA	61
3.1 AQUISIÇÃO DE VOZES	61
3.1.1 GRAVAÇÃO DA VOZ	62
3.1.2 CÁLCULOS PRELIMINARES SOBRE O SINAL DE VOZ	64
3.2 MEDIDAS DE PERTURBAÇÃO	65
3.2.1 ESTIMADOR SÍNCRONO PARA PERÍODO E AMPLITUDE DE F_0	65
3.2.2 PFF (PERTURBATION FREQUENCY FACTOR)	69
3.2.3 DPF (DIRECTION PERTURBATION FACTOR)	69
3.2.4 JR (JITTER RATIO)	69

3.2.5 JF (JITTER FACTOR)	70
3.2.6 SÉRIES TEMPORAIS	70
3.2.7 AVI/PVI (AMPLITUDE/PERTURBATION VARIATION INDEX)	71
3.2.8 MÉDIAS MÓVEIS	71
3.2.9 PERTURBAÇÕES DE ORDEM SUPERIOR	72
3.3 MEDIDAS DE RUÍDO	75
3.3.1 HPFR	75
3.3.2 MEDIDAS DE ENERGIA ESPECTRAL	76
3.3.3 BR	76
3.3.4 SNL	77
3.3.5 HNR	77
3.3.6 HR	78
3.3.7 RA	79
3.3.8 NNE	80
3.3.9 N/S	81
3.3.10 HNR CEPSTRAL	82
3.3.11 PRIMEIRO E SEGUNDO HARMÔNICOS	83
3.3.12 CPP	84
3.3.13 P/A	84
3.3.14 RPK	84
3.4 MEDIDAS NO SINAL RESIDUAL	85
3.4.1 ESTIMAÇÃO PARAMÉTRICA NA ANÁLISE DE VOZ	85
3.4.2 FILTROS ADAPTATIVOS	88
3.4.3 ESTIMATIVA USANDO GRADIENTE DESCENDENTE	88
3.4.4 ESTIMATIVA RECURSIVA POR MÍNIMOS QUADRADOS - RLS	90
3.4.5 ESTIMATIVA POR FILTRO DE KALMAN	91
3.4.6 CONSIDERAÇÕES SOBRE CONVERGÊNCIA	92
3.4.7 MEDIDAS DO SINAL RESIDUAL	93
3.5 ANÁLISE ESTATÍSTICA EMPREGADA	97
3.5.1 PROCEDIMENTO PARA AVALIAÇÃO DA DISCRIMINAÇÃO	99
3.6 REDES NEURAIS ARTIFICIAIS	101
3.6.1 CONSIDERAÇÕES PRÁTICAS	104
4. RESULTADOS	108
4.1 MEDIDAS DE PERTURBAÇÃO	108
4.1.1 PERTURBAÇÕES DE PERÍODO DA FREQUÊNCIA FUNDAMENTAL	108
4.1.2 PERTURBAÇÕES DA AMPLITUDE DA FREQUÊNCIA FUNDAMENTAL	120
4.2 MEDIDAS DE RUÍDO	131
4.2.1 AVALIAÇÃO DAS MEDIDAS HARMÔNICAS	131
4.2.2 MEDIDAS DE ENERGIA ESPECTRAL	134
4.2.3 MEDIDAS DE ENERGIA ESPECTRAL - HFPR	142
4.2.4 MEDIDAS DE SOPROSIDADE - BR	143
4.2.5 MEDIDAS DE ASPEREZA - SNL	144

4.2.6 MEDIDAS BASEADAS EM CEPSTRUM	153
4.2.7 MEDIDA DE RELAÇÃO HARMÔNICA-RUÍDO HNR	164
4.2.8 MEDIDA DE RELAÇÃO HARMÔNICA - HR	166
4.2.9 MEDIDAS DE RELAÇÃO HARMÔNICA - RA	168
4.2.10 MEDIDAS DE RUÍDO NO SINAL DE VOZ - NS	172
4.2.11 MEDIDA DE RUÍDO NO SINAL DE VOZ - NNE	174
4.2.12 OUTRAS MEDIDAS	175
4.3 MEDIDAS RESIDUAIS	178
4.3.1 COEFICIENTE DE EXCESSO	179
4.3.2 PICO DA AUTOCORRELAÇÃO	181
4.3.3 MEDIDA DE PERTURBAÇÃO - <i>JITTER</i>	183
4.3.4 MEDIDA DE PICO RESIDUAL	185
4.3.5 MEDIDA DE NIVELAMENTO ESPECTRAL DO FILTRO INVERSO - SFF	187
4.3.6 MEDIDA DE NIVELAMENTO ESPECTRAL DO SINAL RESIDUAL - SFR	189
4.3.7 MEDIDA DE RELAÇÃO SINAL-RUÍDO	191
4.4 RESULTADOS DA IDENTIFICAÇÃO PATOLÓGICA EMPREGANDO REDES NEURAIAS ARTIFICIAIS	193
4.4.1 CONFIGURAÇÕES EMPREGANDO PERTURBAÇÃO NA AMPLITUDE E PERÍODO DA FUNDAMENTAL.	195
4.4.2 CONFIGURAÇÕES EMPREGANDO ESTIMATIVAS DE RUÍDO TURBULENTO	199
4.4.3 CONFIGURAÇÕES EMPREGANDO RESÍDUOS DA FILTRAGEM INVERSA	202
4.4.4 CONFIGURAÇÃO MISTA	204
5. CONCLUSÕES	207
 BIBLIOGRAFIA	 211
 GLOSSÁRIO	 218
 APÊNDICE	 219

Lista de Figuras

Figura 1 - Trato Vocal.....	7
Figura 2 - Representação Simplificada do Sistema Produtor de Fala	23
Figura 3 - Espectro densidade de potência para o modelo de uma configuração específica da estrutura supraglotal.....	29
Figura 4 - Espectro densidade de potência considerando as perdas devido a radiação nos lábios.	30
Figura 5 - Modelo Representativo do Trato Vocal.....	30
Figura 6 - Trecho periódico do sinal glotal	31
Figura 7 - Espectrograma para o fonema /a/ masculino digitalizado à 22050 Hz, obtido por filtros passa-faixa de 45 Hz e 300 Hz, respectivamente	36
Figura 8 - Estimativas de período e amplitude ciclo-a-ciclo no fonema /a/	39
Figura 9 - Espectro densidade de potência para fonema /a/ em uma voz normal.....	47
Figura 10 - Espectro densidade de potência para fonema /a/ em uma voz áspera.....	48
Figura 11 - Espectro densidade de potência para fonema /a/ em uma voz soprosa	48
Figura 12 - Espectro densidade de potência para fonema /a/ em uma voz rouca	49
Figura 13 - Exemplo de cepstrum real para fonema /a/	53
Figura 14 - Sinal residual para um fonema /a/ de uma voz normal.....	58

Lista de Tabelas

Tabela 1 - Patologias e suas origens.....	13
Tabela 2 - Relação de patologias e quantidade de pessoas	62
Tabela 3 - Exemplo para teste estatístico de Mann-Withney	99
Tabela 4 - Exemplo de graduação (<i>ranking</i>) para teste estatístico.....	99
Tabela 5 - Relação entre os tipos de voz e o vetor de saída esperado	106
Tabela 6 - Análise estatística das perturbações de ordem zero (<i>jitter</i>) para fonema /a/.....	109
Tabela 7 - Análise estatística das perturbações de ordem 1 (<i>jitter</i>) para fonema /a/	109
Tabela 8 - Análise estatística das perturbações de ordem 2 (<i>jitter</i>) para fonema /a/	110
Tabela 9 - Análise estatística das perturbações de ordem 3 (<i>jitter</i>) para fonema /a/	111
Tabela 10 - Análise estatística das perturbações de ordem 4 (<i>jitter</i>) para fonema /a/	111
Tabela 11 - Análise estatística das medidas de perturbação (<i>jitter</i>) para fonema /a/.....	112
Tabela 12 - Análise estatística das perturbações de ordem 0 (<i>jitter</i>) para fonema /e/.....	113
Tabela 13 - Análise estatística das perturbações de ordem 1 (<i>jitter</i>) para fonema /e/.....	113
Tabela 14 - Análise estatística das perturbações de ordem 2 (<i>jitter</i>) para fonema /e/.....	114
Tabela 15 - Análise estatística das perturbações de ordem 3 (<i>jitter</i>) para fonema /e/.....	114
Tabela 16 - Análise estatística das perturbações de ordem 4 (<i>jitter</i>) para fonema /e/.....	115
Tabela 17 - Análise estatística das medidas de perturbação (<i>jitter</i>) para fonema /e/.....	115
Tabela 18 - Análise estatística das medidas de perturbação de ordem 0 (<i>jitter</i>) para fonema /i/	116
Tabela 19 - Análise estatística das medidas de perturbação de ordem 1 (<i>jitter</i>) para fonema /i/	117
Tabela 20 - Análise estatística das medidas de perturbação de ordem 2 (<i>jitter</i>) para fonema /i/	117
Tabela 21 - Análise estatística das medidas de perturbação de ordem 3 (<i>jitter</i>) para fonema /i/	118
Tabela 22 - Análise estatística das medidas de perturbação de ordem 4 (<i>jitter</i>) para fonema /i/	118
Tabela 23 - Análise estatística das medidas de perturbação (<i>jitter</i>) para fonema /i/	119
Tabela 24 - Análise estatística das perturbações de ordem 0 (<i>shimmer</i>) para fonema /a/	120
Tabela 25 - Análise estatística das perturbações de ordem 1 (<i>shimmer</i>) para fonema /a/	121
Tabela 26 - Análise estatística das perturbações de ordem 2 (<i>shimmer</i>) para fonema /a/	121
Tabela 27 - Análise estatística das perturbações de ordem 3 (<i>shimmer</i>) para fonema /a/	122
Tabela 28 - Análise estatística das perturbações de ordem 4 (<i>shimmer</i>) para fonema /a/	122
Tabela 29 - Análise estatística das medidas de perturbação (<i>shimmer</i>) para fonema /a/	123

Tabela 30 - Análise estatística das perturbações de ordem 0 (<i>shimmer</i>) para fonema /e/	123
Tabela 31 - Análise estatística das perturbações de ordem 1 (<i>shimmer</i>) para fonema /e/	124
Tabela 32 - Análise estatística das perturbações de ordem 2 (<i>shimmer</i>) para fonema /e/	124
Tabela 33 - Análise estatística das perturbações de ordem 3 (<i>shimmer</i>) para fonema /e/	125
Tabela 34 - Análise estatística das perturbações de ordem 4 (<i>shimmer</i>) para fonema /e/	126
Tabela 35 - Análise estatística das medidas de perturbação (<i>shimmer</i>) para fonema /e/	126
Tabela 36 - Análise estatística das perturbações de ordem 0 (<i>shimmer</i>) para fonema /i/	127
Tabela 37 - Análise estatística das perturbações de ordem 1 (<i>shimmer</i>) para fonema /i/	127
Tabela 38 - Análise estatística das perturbações de ordem 2 (<i>shimmer</i>) para fonema /i/	128
Tabela 39 - Análise estatística das perturbações de ordem 3 (<i>shimmer</i>) para fonema /i/	129
Tabela 40 - Análise estatística das perturbações de ordem 4 (<i>shimmer</i>) para fonema /i/	129
Tabela 41 - Análise estatística das medidas de perturbação (<i>shimmer</i>) para fonema /i/	130
Tabela 42 - Análise estatística do primeiro harmônico para fonema /a/	131
Tabela 43 - Análise estatística do segundo harmônico para fonema /a/	132
Tabela 44 - Análise estatística do primeiro harmônico para fonema /e/	132
Tabela 45 - Análise estatística do segundo harmônico para fonema /e/	133
Tabela 46 - Análise estatística do primeiro harmônico para fonema /i/	133
Tabela 47 - Análise estatística do segundo harmônico para fonema /i/	134
Tabela 48 - Análise estatística da energia espectral (60 à 400 Hz) para fonema /a/	135
Tabela 49 - Análise estatística da energia espectral (400 à 2000 Hz) para fonema /a/	135
Tabela 50 - Análise estatística da energia espectral (2000 à 5000 Hz) para fonema /a/	136
Tabela 51 - Análise estatística da energia espectral (5000 à 8000 Hz) para fonema /a/	136
Tabela 52 - Análise estatística da energia espectral (60 à 400 Hz) para fonema /e/	137
Tabela 53 - Análise estatística da energia espectral (400 à 2000 Hz) para fonema /e/	138
Tabela 54 - Análise estatística da energia espectral (2000 à 5000 Hz) para fonema /e/	138
Tabela 55 - Análise estatística da energia espectral (5000 à 8000 Hz) para fonema /e/	139
Tabela 56 - Análise estatística da energia espectral (60 à 400 Hz) para fonema /i/	140
Tabela 57 - Análise estatística da energia espectral (400 à 2000 Hz) para fonema /i/	140
Tabela 58 - Análise estatística da energia espectral (2000 à 5000 Hz) para fonema /i/	141
Tabela 59 - Análise estatística da energia espectral (5000 à 8000 Hz) para fonema /i/	141
Tabela 60 - Análise estatística da medida HFPR	143
Tabela 61 - Análise estatística da medida BR	144
Tabela 62 - Análise estatística da energia de ruído espectral (100 à 2600 Hz) para fonema /a/	144
Tabela 63 - Análise estatística da energia de ruído espectral (2600 à 5100 Hz) para fonema /a/	145
Tabela 64 - Análise estatística da energia de ruído espectral (5100 à 8000 Hz) para fonema /a/	146

Tabela 65 - Análise estatística da energia de ruído espectral (100 à 5100 Hz) para fonema /a/	146
Tabela 66 - Análise estatística da energia de ruído espectral (100 à 8000 Hz) para fonema /a/	147
Tabela 67 - Análise estatística da energia de ruído espectral (100 à 2600 Hz) para fonema /e/	147
Tabela 68 - Análise estatística da energia de ruído espectral (2600 à 5100 Hz) para fonema /e/	148
Tabela 69 - Análise estatística da energia de ruído espectral (5100 à 8000 Hz) para fonema /e/	148
Tabela 70 - Análise estatística da energia de ruído espectral (100 à 5100 Hz) para fonema /a/	149
Tabela 71 - Análise estatística da energia de ruído espectral (100 à 8000 Hz) para fonema /a/	149
Tabela 72 - Análise estatística da energia de ruído espectral (100 à 2600 Hz) para fonema /i/	150
Tabela 73 - Análise estatística da energia de ruído espectral (2600 à 5100 Hz) para fonema /i/	151
Tabela 74 - Análise estatística da energia de ruído espectral (5100 à 8000 Hz) para fonema /i/	151
Tabela 75 - Análise estatística da energia de ruído espectral (100 à 5100 Hz) para fonema /i/	152
Tabela 76 - Análise estatística da energia de ruído espectral (100 à 8000 Hz) para fonema /i/	152
Tabela 77 - Análise estatística do pico cepstral proeminente (CPP) para fonema /a/	154
Tabela 78 - Análise estatística da relação cepstral HNR (60 à 400 Hz) para fonema /a/	154
Tabela 79 - Análise estatística da relação cepstral HNR (400 à 2000 Hz) para fonema /a/	155
Tabela 80 - Análise estatística da relação cepstral HNR (2000 à 5000 Hz) para fonema /a/	155
Tabela 81 - Análise estatística da relação cepstral HNR (5000 à 8000 Hz) para fonema /a/	156
Tabela 82 - Análise estatística da relação cepstral HNR (60 à 2000 Hz) para fonema /a/	156
Tabela 83 - Análise estatística do pico cepstral proeminente (CPP) para fonema /e/	157
Tabela 84 - Análise estatística da relação cepstral HNR (60 à 400 Hz) para fonema /e/	158
Tabela 85 - Análise estatística da relação cepstral HNR (400 à 2000 Hz) para fonema /e/	158
Tabela 86 - Análise estatística da relação cepstral HNR (2000 à 5000 Hz) para fonema /e/	159
Tabela 87 - Análise estatística da relação cepstral HNR (5000 à 8000 Hz) para fonema /e/	159
Tabela 88 - Análise estatística da relação cepstral HNR (60 à 2000 Hz) para fonema /e/	160
Tabela 89 - Análise estatística do pico cepstral proeminente (CPP) para fonema /i/	160
Tabela 90 - Análise estatística da relação cepstral HNR (60 à 400 Hz) para fonema /i/	161
Tabela 91 - Análise estatística da relação cepstral HNR (400 à 2000 Hz) para fonema /i/	162
Tabela 92 - Análise estatística da relação cepstral HNR (2000 à 5000 Hz) para fonema /i/	162
Tabela 93 - Análise estatística da relação cepstral HNR (5000 à 8000 Hz) para fonema /i/	163
Tabela 94 - Análise estatística da relação cepstral HNR (60 à 2000 Hz) para fonema /i/	163
Tabela 95 - Análise estatística da medida HNR para fonema /a/	164
Tabela 96 - Análise estatística da medida HNR para fonema /e/	165
Tabela 97 - Análise estatística da medida HNR para fonema /i/	165
Tabela 98 - Análise estatística da medida HR para fonema /a/	166
Tabela 99 - Análise estatística da medida HR para fonema /e/	167

Tabela 100 - Análise estatística da medida HR para fonema /i/	168
Tabela 101 - Análise estatística da medida RA com 3 períodos de fundamental para fonema /a/	169
Tabela 102 - Análise estatística da medida RA com 6 períodos de fundamental para fonema /a/	169
Tabela 103 - Análise estatística da medida RA com 3 períodos de fundamental para fonema /e/	170
Tabela 104 - Análise estatística da medida RA com 6 períodos de fundamental para fonema /e/	170
Tabela 105 - Análise estatística da medida RA com 3 períodos de fundamental para fonema /i/.....	171
Tabela 106 - Análise estatística da medida RA com 6 períodos de fundamental para fonema /i/.....	171
Tabela 107 - Análise estatística da medida NS para fonema /a/.....	172
Tabela 108 - Análise estatística da medida NS para fonema /e/.....	173
Tabela 109 - Análise estatística da medida NS para fonema /i/	173
Tabela 110 - Análise estatística da medida NNE.....	174
Tabela 111 - Análise estatística da relação P/A para fonema /a/	175
Tabela 112 - Análise estatística da relação P/A para fonema /e/	176
Tabela 113 - Análise estatística da relação P/A para fonema /i/	176
Tabela 114 - Análise estatística da relação RPK para fonema /a/	177
Tabela 115 - Análise estatística da medida RPK para fonema /e/	177
Tabela 116 - Análise estatística da medida RPK para fonema /i/.....	178
Tabela 117 - Análise estatística do coeficiente de excesso (EX) para fonema /a/	179
Tabela 118 - Análise estatística do coeficiente de excesso (EX) para fonema /e/	180
Tabela 119 - Análise estatística do coeficiente de excesso (EX) para fonema /i/	180
Tabela 120 - Análise estatística do pico da autocorrelação (PA) para fonema /a/.....	181
Tabela 121 - Análise estatística do pico da autocorrelação (PA) para fonema /e/.....	182
Tabela 122 - Análise estatística do pico da autocorrelação (PA) para fonema /i/	182
Tabela 123 - Análise estatística do jitter para fonema /a/	183
Tabela 124 - Análise estatística do jitter para fonema /e/	184
Tabela 125 - Análise estatística do jitter para fonema /i/.....	184
Tabela 126 - Análise estatística da medida pico para o fonema /a/	185
Tabela 127 - Análise estatística da medida pico para o fonema /e/	186
Tabela 128 - Análise estatística da medida pico para o fonema /i/.....	186
Tabela 129 - Análise estatística da medida SFF para o fonema /a/	187
Tabela 130 - Análise estatística da medida SFF para o fonema /e/	187
Tabela 131 - Análise estatística da medida SFF para o fonema /i/	188
Tabela 132 - Análise estatística da medida SFR para o fonema /a/.....	189
Tabela 133 - Análise estatística da medida SFR para o fonema /e/.....	190
Tabela 134 - Análise estatística da medida SFR para o fonema /i/.....	190

Tabela 135 - Análise estatística da medida SNR para o fonema /a/	191
Tabela 136 - Análise estatística da medida SNR para o fonema /e/	191
Tabela 137 - Análise estatística da medida SNR para o fonema /i/	192
Tabela 138 - Relação de patologias e quantidade de pessoas para teste das redes neurais	195
Tabela 139 - Classificação patológica por redes neurais para medidas de perturbação do período e amplitude da fundamental.....	198
Tabela 140 - Classificação patológica por redes neurais para medidas de ruído.....	202
Tabela 141 - Classificação patológica por redes neurais para medidas residuais.....	204
Tabela 142 - Classificação patológica por redes neurais para medidas acústicas.....	206

Lista de Símbolos

a	coeficiente angular da regressão linear.
a, α	coeficientes de modelos digitais.
α	nível de significância.
$A(x)$	função de área.
$A(z)$	coeficientes da porção autorregressiva do modelo AR.
b	coeficiente linear da regressão linear.
β	constante de tempo para convergência.
c	velocidade do ar
$C(z)$	coeficientes da porção exógena do modelo ARX.
$\chi_s(\lambda)$	medida genérica obtida sobre sequência numérica.
E	erro de propagação de uma rede neural.
$E(z), E(\omega)$	resposta em frequência do modelo de excitação do trato vocal.
$F(x), F(x,y)$	função genérica.
$f(x)$	função de propagação em uma rede neural.
F_0	frequência fundamental.
F_i	frequência fundamental, ciclo-a-ciclo.
g	taxa de adaptação dos algoritmos adaptativos.
$g(n)$	sinal glotal.
$G(z)$	resposta em frequência do trato glotal.
η	taxa de aprendizagem.
H	entropia relativa.
$h(t)$	porção periódica no sinal.
$H(\omega)$	componentes harmônicas.
H, R_0	ganho.
H_0	hipótese nula em teste estatístico.
H_1	hipótese alternativa em teste estatístico.
i, j, k	índices.
κ, r	coeficiente de reflexão.
L	quantidade de harmônicos.
l	comprimento.

M	quantidade de amostras.
mse	função para minimização.
∇	gradiente de uma função.
N	quantidade de amostras.
n	tempo digital.
$N(\omega)$	componentes não-harmônicas.
NET	regra de propagação.
o	entrada de um neurônio.
P	quantidade de amostras em janelas espectrais.
p	função de pressão.
P	matriz de covariância.
P	número de pólos.
p	p-ésimo padrão para treinamento.
P	probabilidade obtida pela rede neural.
P_α	probabilidade a priori.
P_i	período da fundamental, ciclo-a-ciclo.
Q	probabilidade real.
r	autocorrelação.
ρ	erro médio quadrático.
ρ	taxa de convergência do filtro de Kalman.
ρ	densidade do ar.
$R(z)$	resposta em frequência da radiação dos lábios.
rc	cepstrum real.
r_p	coeficiente de Pearson.
σ	variância.
s	sinal de voz.
$\tilde{T}(\lambda)$	transformação genérica sobre sequência numérica.
T	período de amostragem.
t	tempo (em segundos).
t	unidade de tempo.
t_p	período de amostragem.
U	resultado numérico do teste Mann-Withney.
v	taxa de convergência dos algoritmos adaptativos.
v	função de velocidade do ar.
$V(z), V(\omega)$	resposta em frequência do modelo da supraglotal.
w	ruído branco.

$w(n)$	janela espectral.
ω	pesos sinápticos.
x	unidade de espaço.
X_i	variável genérica.
y	saída de um neurônio.
Z	impedância
z	variável normal.
Z, M, N, H	matrizes do modelo por espaço de estados.

Resumo

“Ver o corpo humano por dentro” sem a necessidade de intervenção cirúrgica é objetivo que motivou a criação de diversos instrumentos como eletrocardiogramas, eletroencefalogramas, equipamentos de ressonância magnética e raio-X. Através da avaliação de imagens ou resultados numéricos, pode-se detectar patologias nos primeiros estágios, permitindo uma ação decisiva de especialistas médicos na cura destas. Especialistas da fala normalmente empregam instrumentos como videolaringoscopia e videoestroboscopia para avaliar qualitativamente o comportamento da laringe e pregas vocais. Compreendendo que a voz transmite informações sobre alterações orgânicas ou funcionais nas estruturas de vocalização, este trabalho apresenta um conjunto de medidas acústicas baseadas neste sinal que evidenciam alterações na periodicidade do movimento das cordas vocais e quantidade de ruído turbulento que atravessa a glote. A partir de avaliação estatística da capacidade discriminatória destes índices acústicos e empregando-se redes neurais artificiais, define-se um método automático para identificação probabilística das patologias que afetam as estruturas da laringe.

Palavras-chaves: processamento digital de sinais; análise acústica da voz; diagnóstico automático de patologias; aplicação de redes neurais artificiais

Abstract

“To see the inside of the human body” without the necessity of surgical intervention is the objective that motivates the conception of several instruments like electrocardiogram, electroencephalogram, magnetic resonance and x-ray equipments. Through the image analysis or numerical results, it is possible to identify pathologies, allowing a decisive action of physician specialists in cure of these. Voice specialists, normally, use instruments as videolaryngoscopy and videostroboscopy to assess the vocal folds and larynx comportment, qualitatively. Understanding that the voice transmits information upon functional or organic alterations in vocalization structures, this work presents a set of acoustic measurements, based on this signal, that evidences alterations on vocal folds movement periodicity and quantity of turbulent noise through the glottis. From the statistic evaluation of discriminatory capacity of these acoustic indexes and using artificial neural networks, it defines an automatic method for the probabilistic identification of pathologies that affect the laryngeal structures.

keywords: digital signal processing; acoustical analysis of voice; automatic diagnosis of pathologies; application of artificial neural networks.

1. Introdução

O tratamento adequado de qualquer patologia exige um preciso diagnóstico do paciente. O método mais simples de diagnóstico é a utilização dos sentidos na avaliação de um órgão ou sistema do corpo humano. O tato, a visão e a audição permitem ao especialista identificar anormalidades, baseado em seu conhecimento anátomo-funcional do corpo humano e das patologias que podem estar modificando o comportamento orgânico do paciente.

Entretanto, estas técnicas perceptuais mostram-se ineficazes pois variam conforme a sensibilidade e conhecimento do especialista. Isto acarreta na dificuldade de padronizar-se este tipo de diagnóstico pela existência de expressões como “muito sopro”, “extremamente áspero”, “lenta pulsação”. A dificuldade de quantificação impossibilita uma clara comparação entre os julgamentos clínicos. O próprio conceito de normalidade, considerando esta metodologia, apresenta uma gama de possíveis termos qualitativos variando conforme o especialista, seu conhecimento técnico e sua experiência.

O desenvolvimento da tecnologia tem auxiliado grandemente a medicina pois provê uma grande quantidade de informações objetivas sobre as condições do paciente ao especialista, de maneira que o tratamento aplicado possa ser o mais adequado possível.

A avaliação das estruturas internas do corpo humano sem a necessidade de intervenção cirúrgica é possível graças à equipamentos como raio-X e tomógrafos por ressonância magnética. O funcionamento de alguns órgãos podem ser mapeados numericamente através de eletroencefalograma (EEG) para avaliação da atividade cerebral, eletrocardiograma (EKG) para monitorar a pulsação cardíaca, audiômetro para quantificação da capacidade auditiva. Métodos quantitativos permitem que os especialistas possam determinar precisamente as alterações orgânicas ou funcionais.

O sistema de produção de fala é uma importante estrutura sensorial pois permite ao ser humano comunicar-se, compartilhar informações de forma eficiente. A ineficiência destas estruturas ou sua completa ausência não representa para o paciente um risco à sua vida, mas dificulta sua convivência na sociedade. A perda temporária da voz para profissionais como radialistas, cantores, ou àqueles que necessitam falar em público representa um sério problema pois acarreta na incapacidade de exercer seu trabalho, podendo resultar posteriormente no afastamento definitivo do emprego. Detectar antecipadamente anomalias no trato vocal permite que se aplique um tratamento adequado, podendo evitar-se assim intervenções cirúrgicas mais complexas.

A técnica mais comum utilizada por especialistas da voz (otorrinolaringologistas e fonoaudiólogos) baseiam-se na avaliação perceptual das estruturas de vocalização e ressonância (laringe, pregas vocais e órgãos da supraglotal). Utilizando a própria audição, estes especialistas identificam as vozes em graduações de rouquidão, sopro e aspereza, sendo também considerado a intensidade da fonação (*loudness*). Vozes "muito" soprosas podem indicar a presença de nódulos ou fendas fusiformes. Entretanto, esta técnica apresenta três problemas:

1. É extremamente dependente do conhecimento e da acuidade auditiva do especialista, acarretando em diferenças de diagnóstico entre médicos, ou até mesmo para um mesmo especialista em instantes diferentes.
2. Não existe uma padronização adequada na graduação dos índices perceptuais apesar de metodologias como GRBAS (Imazumi, 1986) e PERC (Askenfelt e Hammarberg, 1986), motivado pelas considerações citadas no primeiro item.
3. Devido a complexidade e redundâncias do sistema de produção de fala, existem diversos modos pelos quais as estruturas do trato vocal produzem um mesmo sinal de voz. Pode-se identificar a condição perceptual, mas a gama de possibilidades para a geração desta fala impede a identificação precisa da natureza funcional ou orgânica de uma anomalia, que até mesmo pode inexistir.

Com isto, verifica-se a necessidade de visualizar diretamente o trato vocal para avaliar corretamente sua condição (normal ou não). Utilizando laringoscopia indireta (visão por espelho das pregas vocais) pode-se avaliar o movimento da mucosa que reveste as pregas vocais e a posição as estruturas que dão suporte à fonação, bem como sua coloração e configuração durante a fonação de alguma vogal sustentada.

O interesse de avaliar as estruturas de vocalização, mais precisamente laringe e cordas vocais, é a importância destas estruturas na produção da fala. O fluxo de ar proveniente dos pulmões "vibra" as pregas vocais, criando um sinal quase periódico que posteriormente sofrerá modificações espectrais (amplificação por ressonância de determinadas componentes harmônicas) de tal forma a produzir a voz propriamente dita. Alterações nesta estrutura afetam a geração de todos os fonemas, não existindo uma forma natural do ser humano compensar um dano nestas estruturas como ocorre com anormalidades na supraglote (esta estrutura possui configurações alternativas que permitem a ressonância adequada dos fonemas).

Para melhorar a visualização da laringe e pregas vocais, pode-se utilizar da videolaringoscopia e nasofibrosopia. Estes equipamentos permitem, pelo emprego de fibras ópticas, uma observação direta do funcionamento destas estruturas durante a vocalização. Com o emprego conjugado do estroboscópio, pode-se ver os movimentos de abertura e fechamento das pregas vocais (oscilações na mucosa) e diagnosticar diretamente alterações orgânicas ou funcionais na laringe.

Os filmes ultra-rápidos também são usados para verificar a existência de crescimentos orgânicos nas pregas vocais como cistos e pólipos, além de registrarem os movimentos da laringe

durante a produção da fala. Entretanto, as exigências de filmes com ASA elevada e iluminação adequada das pregas vocais (por fibra óptica) dificulta sua utilização.

Pode-se verificar que os instrumentos apresentados e que atualmente são utilizados pelos médicos especialistas e fonoaudiólogos ainda demandam uma boa acuidade visual, além do conhecimento exigido para identificar corretamente as modificações orgânicas ao longo da laringe e pregas vocais. A própria manipulação dos equipamentos exige destreza para que sejam produzidas boas imagens da cavidade glotal, além do custo de aquisição e manutenção destes equipamentos, resultando em exames caros.

Cabe citar alguns problemas ainda encontrados nos exames produzidos com o uso destes instrumentos. Primeiro, ainda são avaliações qualitativas (apesar da possibilidade de medição e localização de alterações orgânicas) e desta forma ainda podem resultar em diferentes diagnósticos, diferente das técnicas como EKG e EEG para cardiologias e neurologias, respectivamente.

Outro problema enfrentado são os atos reflexos que os pacientes podem apresentar quando o exame é realizado. O contato do tubo rígido do fibroscópio com algumas regiões de maior sensibilidade nas cavidades supraglóticas ou do tubo flexível do nasofibroscópio, pode impedir a boa visualização das pregas vocais, mesmo com o emprego de anestésicos como a xilocaína.

Como não se pode observar as pregas vocais a partir do pulmão, ou seja, de baixo para cima, crescimentos que ali ocorram não podem ser visualizados por estes métodos, caracterizando-se assim mais uma dificuldade no diagnóstico de patologias da laringe.

Com o intuito de suplantar estas dificuldades e prover auxílio aos especialistas no diagnóstico de doenças do trato vocal, particularmente as que afetam as pregas vocais, foi considerado o emprego do sinal de voz na avaliação das estruturas que o produziram. Através da análise da fala, indiretamente pode-se identificar as disfonias uma vez que esta informação está codificada na voz. Sendo um método não-invasivo, pode ser realizado em qualquer pessoa, inclusive eliminando o uso de anestésicos.

Dada a natureza numérica da análise acústica, os diagnósticos são quantitativos e portanto objetivos. Permitem que o especialista de voz possa indicar o tratamento mais adequado ao paciente, uma vez que terá subsídios que lhe possibilitam identificar corretamente a patologia. Além disto, elimina dúvidas eventuais que exames por laringoscopia indireta possam apresentar.

A própria evolução do tratamento pode ser avaliada, pois as medidas acústicas (extraídas do sinal de voz) indicam precisamente o nível de melhora do paciente, permitindo que o especialista mantenha ou altere a medicação ou a forma de conduzir o tratamento.

Nos estudos que empregam medidas acústicas, verifica-se que a grande maioria busca relacionar estes valores obtidos do sinal de voz com características perceptuais da voz. Como a classificação das vozes analisadas é realizada por em termos de variações de rouquidão, aspereza e sopro, existem naturais diferenças entre as avaliações dos especialistas de voz que classificam estas vozes para os estudos. Outra dificuldade com este método de utilizar as medidas acústicas para

avaliação automática da condição da laringe é que existe uma natural correlação entre estes índices perceptuais, acarretando na dificuldade de diferenciação entre vozes “praticamente normais” daquelas que apresentam “leves” níveis de alteração.

Considerando as dificuldades apresentadas no diagnóstico diferencial das patologias da laringe e do trato vocal, este trabalho busca relacionar as medidas acústicas extraídas do sinal de voz diretamente com as disfonias, ou seja, possibilitar uma identificação automática destas doenças apenas com o uso do sinal de voz digitalizado.

Apresenta-se primeiramente alguns aspectos anatômicos e fisiológicos do trato vocal e as anomalias que mais comumente afetam a população. Segue então a descrição dos modelos e técnicas matemáticas que definem o sistema de produção de fala. Muitas medidas acústicas baseiam-se em modificações espectrais produzidas pelas patologias, sendo desta forma importante compreender como alterações na região da glote se propagam ao longo das cavidades do trato vocal.

Realiza-se então a descrição dos parâmetros acústicos existentes na literatura e a metodologia de cálculo para a sua obtenção. Como muitas destas medidas são obtidas de “trechos mais estáveis” do sinal de voz, foi necessário modificar alguns dos passos indicados para permitir que não fosse necessário uma inspeção manual do intervalo de cálculo no sinal de voz.

Outra modificação realizada em alguns parâmetros acústicos foi avaliar seu comportamento ao longo do sinal de voz, sendo então empregados métodos estatísticos clássicos e robustos. Isto propicia uma melhor compreensão de como as medidas acústicas relacionam-se com as alterações no sinal de voz dada a extensão da disfonia analisada.

Algumas técnicas são sugeridas com a intenção de se obter medidas robustas baseado no método de filtragem inversa de Davis (1979). Empregando-se filtros adaptativos (Wiener e Kalman) sobre todo o sinal de fala sustentada, determina-se um conjunto de medidas que permitem a identificação de disfonias.

Com todas as medidas acústicas determinadas utiliza-se um teste não-paramétrico (Mann-Withney) para avaliar a capacidade discriminatória destas, para selecionar as que são mais eficientes individualmente. São confrontadas as patologias para verificar a existência de separabilidade estatística entre elas, considerando cada um dos parâmetros acústicos analisados.

Finalmente, emprega-se redes neurais artificiais para realizar a discriminação multivariada das patologias estudadas. Diversas configurações estruturais foram usadas (número de camadas, índices acústicos empregados), considerando algumas características comuns das patologias.

O objetivo fundamental deste trabalho é desenvolver métodos automáticos para identificação das patologias do trato vocal, particularmente da laringe, empregando-se unicamente o sinal de voz, considerando os benefícios que podem advir como:

1. Baixo custo pois é baseado apenas em algoritmos matemáticos e sinais digitais;
2. Facilidade de uso pois apenas necessita da fala sustentada e fornece como resultado a identificação probabilística das disfonias;

3. Pré-diagnóstico de patologias ainda em estágios que não exigem intervenção cirúrgica;
4. Monitoração de tratamentos clínicos realizados no paciente;
5. Possibilita a criação de uma base de dados clínica do paciente no decorrer dos anos.

Estes benefícios motivam o estudo da voz com objetivo de identificar as patologias do trato vocal permitindo que o especialista de voz (otorrinolaringologista ou fonoaudiólogo) tenha uma ferramenta extremamente útil no diagnóstico das disfonias da laringe e pregas vocais.

2. Fundamentação Teórica

Neste capítulo são apresentados as bases, tanto de anatomia e fisiologia como matemática, para compreensão das técnicas empregadas na avaliação e identificação das patologias do trato vocal. Primeiramente é apresentado o trato vocal, seu funcionamento e as patologias mais comuns. Em seguida, é apresentado o modelo matemático do sistema produtor de fala e as técnicas para sua avaliação.

2.1 Voz e Trato Vocal

O som é uma onda mecânica (e portanto, necessitando de um meio físico para propagação) que se movimenta em sentido transversal à direção de propagação. As sucessivas compressões e dilatações do ar, ou seja, alterações de pressão, se propagam até influenciar o ouvido (ou um transdutor de pressão), onde é adequadamente decodificado (traduzido para impulsos elétricos, por exemplo) e interpretado.

Uma forma inteligível do som é a voz, que é produzida a partir de um conjunto de órgãos que apresentam como função primária possibilitar a respiração humana e auxiliar nos mecanismos da deglutição. A produção dos fonemas (representação dos sons produzidos, que formam as linguagens propriamente ditas) estão sujeitos à limitações fisiológicas impostas por estes órgãos dada as exigências respiratórias do ser humano.

Para facilitar o estudo do sistema de produção de fala, o mesmo é normalmente reduzido à três grupos distintos, em relação à onda sonora que é produzida ou modificada pelos órgãos. Além disto, permite a construção de um modelo simplificado que se assemelha aos subgrupos. O primeiro grupo corresponde à produção de um fluxo de ar; por iniciar nos pulmões e encerrar no final da traquéia (*canus elasticus*), recebe o nome de subgrupo de respiração. Situando-se na faringe encontram-se os órgãos do segundo grupo, responsáveis pela produção de uma onda quase-periódica mediante vibração das cordas vocais em um movimento por variação de pressão (segundo a lei de Bernoulli); o período entre duas aberturas (ou fechamentos) sucessivos determina a frequência fundamental (ou *pitch*). Com a responsabilidade de produzir os fonemas que constituem o vocabulário acústico de uma língua encontra-se o sistema ressonador, constituindo assim o último grupo de órgãos do trato vocal.

Figura 1 - Trato Vocal

Apesar da aparente separação induzir a total independência entre os grupos, deve-se manter em mente que existe o acoplamento entre estes elementos pois a inexistência de um dos grupos automaticamente impede a produção adequada da fala. Além disto, a dinâmica da fala envolve a movimentação simultânea de diversas estruturas do trato vocal; isto leva a se considerar as redundâncias do trato vocal quanto a geração de voz, pois existe um conjunto de configurações dos órgãos que produzem os mesmos fonemas e coarticulações.

Para facilitar o estudo anatomo-fisiológico do trato vocal, os três grupos estruturais são apresentados nas próximas seções.

2.1.1 Respiração

O primeiro grupo de órgãos está ligado à etapa de respiração, ou seja, a geração de um fluxo de ar variável que excitará o sistema produtor de fala sito à laringe e órgãos posteriores. Como a

respiração tem associada a função primária de transporte de gases para renovação de oxigênio, existem limitações sobre a quantidade de som (ou mais precisamente, fonemas) que se pode sustentar por uma única expiração. Este é um fator importante quando da consideração dos experimentos de análise da voz, tanto normal quanto patológica.

A função respiratória é necessária à produção de voz pelo fato de que as ondas sonoras são carregadas pelo fluxo de ar proveniente dos pulmões. Como a capacidade respiratória é limitada pelo volume de ar passível de armazenamento no pulmão, a capacidade de fonação fica limitada tanto em altura e duração, apesar de ser possível elevar estas características através do controle do fluxo de ar proveniente do pulmão e posteriormente nas cordas vocais.

Anatomicamente, o sistema respiratório do ser humano - e que está relacionado à fonação - compreende um par de pulmões, que funciona como reservatório de ar. Como meio de transporte de ar a partir dos pulmões, encontram-se os sacos aéreos e os bronquíolos, onde as trocas gasosas são realizadas, entregando a partir destes a massa de ar para os brônquios (também em par, um para cada pulmão). Estes dois brônquios unem-se à traquéia, que serve de duto para o ar circular, em ambas as direções (de e para os pulmões).

Além destas estruturas, há o diafragma (composto de musculatura, tendão e membrana, com finalidade de separação da cavidade torácica da cavidade abdominal), importante na atividade respiratória e desta forma, intimamente ligada com o processo fonatório.

O funcionamento do processo respiratório é semelhante à um fole: com a separação das duas alças, cria-se uma condição de menor pressão interna que a externa, gerando um movimento de ar na direção fora-para-dentro, caracterizando a inspiração. Análogo à este procedimento, a expiração ocorre quando a caixa torácica expande-se (com o auxílio das vértebras, que possuem certa mobilidade, assim como da movimentação do diafragma), gerando uma condição de baixa pressão interna nos pulmões que caracteriza um fluxo de ar para dentro destes, ou seja, à inspiração.

A expiração entretanto é caracterizada não por uma contração ativa dos pulmões e sim por um relaxamento natural destes órgãos (colapso passivo do tórax), uma vez que tais estruturas foram forçadas para um posicionamento não-natural quando da inspiração. Isto é importante pois como a fonação é produzida pelo ar que é expelido dos pulmões, a fala é produzida pela componente passiva da respiração: qualquer tentativa de aplicação de força para produzir uma vazão de ar maior do que a natural (passiva) corresponde uma hiperatividade, isto é, quando se deseja manter por um maior intervalo de tempo a fonação, por exemplo.

Esta hiperatividade pode ser avaliada através das capacidades volumétricas pulmonares. Uma das principais medidas é o volume corrente (VC) que é a quantidade de ar inspirado e expirado durante um ciclo respiratório normal, sendo determinado apenas pelas necessidades de oxigênio (sem relação com fonação). A capacidade vital de um indivíduo está relacionada com o volume total de ar que pode ser expirado dos pulmões a partir de uma inspiração máxima; cabe lembrar que existe uma quantidade mínima de ar que não pode ser expirada, por mais forte que a expiração possa ser -

correspondente ao volume residual. Estas medidas possibilitam avaliação da capacidade pulmonar e respiratória de um ser humano e servem para se determinar diferenças entre o tempo de expiração e inspiração, verificando-se que o tempo de expiração é levemente superior ao da inspiração, principalmente devido as necessidades de fonação e controle do tempo de expulsão de ar para os órgãos glotais e supraglóticos.

2.1.2 Vocalização

A laringe e estruturas anexas não são prioritariamente voltadas para dar suporte a fonação. Sua função primária é impedir que corpos estranhos cheguem ao pulmão, através de movimento da epiglote e das estruturas próximas (como cordas vocais e aritenóides). Além disto, a segunda atividade básica destas estruturas é a fixação do tórax, bloqueando a passagem de ar através da glote para possibilitar que os membros superiores realizem tarefas que exijam força.

Mediante o controle coordenado da capacidade valvular da laringe e órgãos próximos (como pregas vocais) pode-se produzir um som de baixa intensidade, quase-periódico, que corretamente modulado (pelo trato supraglotal) torna-se “inteligível”. Pode-se verificar esta característica das pregas vocais na fig. 1, pelo posicionamento destas logo após a traquéia.

As estruturas da laringe são compostas por um conjunto de cartilagens, músculos e ligamentos que dão suporte as atividades exigidas. As cartilagens tem função estrutural; correspondem à porção “fixa”, sendo movimentada pelos músculos e ligamentos.

Existe um conjunto de cinco cartilagens principais (além de duas outras, as corniculadas, que têm pouca função na fonação), à saber, a cricóideia, a tireóideia, um par de aritenóideas e a epiglote. Corresponde ao “esqueleto” do sistema vibratório.

A cricóideia está na base da laringe como um anel de extensão, dando suporte às demais cartilagens e músculos, principalmente às aritenóideas. Envolvendo tanto a cricóideia quando as aritenóideas, encontra-se a tireóideia, como um revestimento protetor das estruturas. A epiglote apresenta a função de, durante a deglutição, bloquear a entrada de corpos estranhos nos pulmões, abrindo-se para a respiração e atividade fonética.

As cartilagens aritenóideas possuem uma função característica que é a de esticar ou relaxar a musculatura das cordas vocais criando um ambiente propício as movimentos da membrana que reveste as pregas durante o fenômeno vibratório da vocalização.

A musculatura existente é dividida em 2 grupos em relação à sua colocação no sistema de vocalização. Os que estão ligados somente às estruturas internas da laringe (intrínsecos) e os que se prendem também à elementos externos (extrínsecos). Os extrínsecos permitem a elevação ou abaixamento da laringe, além de mantê-la fixa; nas notas agudas, os músculos elevadores posicionam alto a laringe, enquanto que nas notas mais graves, estes relaxam-se enquanto a ação de rebaixamento da laringe fica a cargo dos depressores. O músculo cricofaríngeo (também externo) tem função fixadora da laringe e permite a produção de fala esofágica, quando adequadamente treinado

(situação em que o paciente é corpectomizado e necessita de uma fonte secundária de vibração para produzir sinal quase periódico).

Os músculos intrínsecos são em número de doze (seis pares) e agem diretamente sobre a fonação. Os cricoaritenóideos posteriores possibilitam a abertura das pregas vocais através de sua contração, afastando as cartilagens aritenóideas e relaxando assim as pregas vocais. Já o par cricoaritenóideo lateral tem função oposta, ou seja, atrai essas cartilagens, uma à outra, de modo a produzir uma adução nas cordas vocais. As fibras aritenóideas transversais auxiliam também no movimento da cartilagem aritenóidea, além de comprimir as pregas vocais. Os músculos aritenóideos oblíquos são então envolvidos no fechamento glotal.

Os tiroaritenóideos correspondem a porção muscular do que se denomina corda vocal, uma vez que esta ainda compreende a superfície interna da cartilagem aritenóidea e o ligamento vocal, sendo recoberto por uma membrana esbranquiçada, que efetivamente vibra. A inexistência desta membrana descaracteriza o processo de vocalização, pois as variações de pressão agem sobre esta, enquanto que a prega vocal apenas se abre e fecha, permitindo que pequena quantidade de fluxo de ar atravesse a glote e as características oscilatórias do sinal são geradas pelo movimento da membrana que recobre este par de músculos.

O último par de músculos da laringe (em relação à fala) são os cricótireóideos sendo responsáveis pelo distanciamento das cartilagens aritenóidea e tireóidea, provocando um estiramento das pregas vocais, fechando-as.

As pregas vocais, que efetuam a constrição da laringe, produzem o sinal quase-periódico através da vibração da mucosa (composta pelo epitélio e lâmina superficial - espaço de Reinke), que junto com a camada de transição - ligamento vocal - deslizam sobre o músculo tiroaritenóideo, por quase toda extensão do espaço glotal. A inexistência, ou defeito, desta superfície, seja por cirurgia ou patologia (como pólipos) incorre na produção de sinal de pouca ou nenhuma periodicidade.

Para produzir o sinal característico, os músculos intrínsecos adutores contraem-se de modo a unir as pregas vocais, de modo a obstruir a passagem de ar no espaço glotal, sendo que o cricoaritenóideo lateral e o tirearitenóideo contraem-se rapidamente (na ordem de até 15 ms), fazendo com que as pregas vocais se aproximem antes do início da vocalização. As pregas vocais não se fecham totalmente; uma fenda de até 3 mm possibilita a vibração das mucosas. Mediante alterações de nível de pressão e velocidade do fluxo do ar através desta fenda, as pregas vocais fecham-se e abrem-se, criando um ciclo de atividades em que por aumento da pressão subglote as membranas permitem a passagem de ar, e com o aumento da velocidade do ar, o efeito sugador de Bernoulli as reaproxima novamente. Este ciclo vibratório da mucosa repete-se aproximadamente 125 vezes por segundo nos homens e 215 vezes por segundo em mulheres, ambos adultos.

Apesar das pregas vocais “fecharem-se” em quase toda a sua extensão, é comum em algumas mulheres criar-se uma lacuna na porção posterior da glote, ou fenda fusiforme posterior, o que pode acarretar em um sinal com uma pequena quantidade de ruído turbulento, uma vez que a membrana

não consegue completar a adução total das cordas vocais. O não fechamento completo das cordas vocais pela mucosa, seja por características patológicas funcionais ou orgânicas permite que fluxo de ar escape, criando um sinal turbulento, caracterizando uma voz soprosa.

Características como o comprimento e a espessura das pregas vocais são importantes quando se considera a frequência fundamental do sinal produzido na abertura glotal. Pregas curtas, espessas tendem a ser frouxas, resultando em uma baixa velocidade na realização do ciclo de vibração; pregas mais finas e tensas levam a produção de sons mais agudos. A tensão exercida sobre as cordas vocais também é um elemento importante a ser considerado, pois nesta condição de extremo trabalho, a mucosa tende a não vibrar quase-periodicamente, acarretando em “erros” de vibração. O sinal resultante é estridente ou tenso, caracterizando uma voz áspera.

O aumento e a redução no comprimento das pregas vocais produzem sons mais altos ou mais baixos, uma vez que aumentam ou reduzem, respectivamente, a densidade do tecido das cordas vocais. Este movimentos são obtidos pela ação dos músculos cricotireóideo e cricofaríngeo, que influenciam no movimento das cartilagens (cricóidea e tireóidea). O aumento da tensão sobre as cordas vocais exige um maior fluxo de ar para produzir vibração na mucosa devido a alteração massa-volume desta.

A intensidade da voz (volume) está diretamente relacionada com alterações na queda de pressão de ar na glote e abaixo da mesma. A quantidade de ar, e sua velocidade de fluxo provida pelos pulmões, influenciam na velocidade do ciclo de adução/abdução das cordas vocais, sendo comum em pessoas não-treinadas elevarem tanto a frequência fundamental quando a intensidade do sinal glotal.

As características apresentadas nos últimos parágrafos influenciam no tipo de fonema que se produzirá com os efeitos de ressonância das estruturas supraglotalis. O nível de tensão ou relaxamento das pregas vocais auxilia na caracterização fonética. O fonema /i/, por exemplo, exige uma redução na densidade do material da mucosa, resultando num sinal mais estridente que o fonema /e/. Fonemas como /o/ e /u/ são mais soprosos que /a/, por serem caracterizados por uma condição de maior relaxamento das pregas vocais. A exigência de fluxo de ar é proporcional à estas exigências de tensão muscular.

2.1.3 Ressonância

O sinal proveniente da glote corresponde à um tom de baixa intensidade, que necessita ser amplificado e determinadas componentes harmônicas sofrerem ênfase, de maneira que os fonemas sejam caracterizados.

Visualizando a fig. 1, o trato supraglotal inicia a partir da hipofaringe, logo acima das pregas ariepiglóticas. A hipofaringe (cavidade diretamente acima do esôfago) compõem-se de constritores inferiores, desembocando na orofaringe, bem próximo à epiglote, seguindo pelos constritores médios da faringe (região denominada de orofaringe). Segue-se então pela valécua e inicia-se tanto o trato

bucal quanto o nasal. O acoplamento do trato nasal é realizado através do véu palatino (ou palato mole), na região da nasofaringe, finalizando com a cavidade nasal.

A cavidade bucal apresenta a língua como elemento importante de ampliação ou redução da área de ressonância, sustentado pela mandíbula, finalizando com os dentes e lábios.

O fenômeno que rege a produção de fonemas, mediante ênfase e atenuação de regiões do espectro, é conhecido por ressonância. Semelhante ao que ocorre em instrumentos musicais, onde ocorre o acoplamento de novas áreas para que o som seja amplificado, os diversos músculos e estruturas supraglóticas agem de forma a alterar a frequência natural que será amplificada. Assim, as componentes harmônicas e não-harmônicas são modificadas, seguindo uma envoltória que as modula adequadamente, de acordo com o fonema a ser produzido.

Conforme o volume (menor ou maior) produzido pelas cavidades do trato supraglotal, harmônicas mais baixas ou mais altas são ressoadas, respectivamente. Entretanto, estas estruturas estão alterando-se continuamente, conforme a expressão pronunciada. Mesmo com estas mudanças de configuração (que caracterizam principalmente os fonemas vocálicos), elas são lentas quando comparadas à frequência fundamental do sinal glotal. Assim, por breves instantes, pode-se admitir a estaticidade do mesmo. Consoantes, por exemplo, caracterizam-se por alteração nas cavidades do trato supraglotal, e a mudança entre os fonemas é denominada coarticulação.

O formato anatômico destas estruturas também influencia na produção da fala; acoplamento da cavidade nasal, grau de tensão ou relaxamento nos constritores faríngeos, nível de abertura mandibular e dos lábios influem no fonema a ser produzido. Os elementos ressonadores fornecem a percepção do dialeto regional e auxiliam na identificação da pessoa.

É importante citar que a frequência fundamental advinda da cavidade glotal sofre modificação ao longo do trato vocal, por modificações na intensidade das componentes harmônicas. A escolha de vogais para avaliação patológica é de suma importância para uma correta análise acústica.

2.2 Patologias do Trato Vocal

A produção de fala exige o bom funcionamento e condições das estruturas do trato vocal. Qualquer interferência na dinâmica implica na geração de um sinal acústico diferente do que a pessoa emite normalmente, geralmente exigindo algum esforço adicional da mesma na sua produção. As patologias podem ocorrer em qualquer um dos três grupos de órgãos do trato vocal (respiratório, vibratório ou ressonador), sendo convencionalmente classificadas quanto à origem das mesmas: origem funcional, onde por um mal uso ou funcionamento de parte do trato vocal, origina-se uma fala não-normal; e origem orgânica, em que a patologia é proveniente de anomalias físicas no sistema de produção de fala.

As patologias do trato vocal são apresentadas na tabela abaixo conforme a classificação etiológica das mesmas. Algumas informam a condição das estruturas laringeas do que pelas modificações acústicas que elas produzem.

Tabela 1 - Patologias e suas origens

Origem Funcional	Origem Orgânica
Edema de Reinke	Carcinoma
Diplofonia	Cistos
Disfonia Espasmódica	Disartria
Disfonia Ventricular	Granuloma
Espessamento	Hiperceratose
Falsete	Laringite infecciosa
Laringite traumática	Leucoplasia
Nódulos	Membrana laringiana
Pólipos	Mutações endócrinas
Quebras de altura	Paralisia de pregas vocais
Quebras de vocalização	
Úlceras de contato	

Acusticamente, as patologias são diferenciadas em relação à padrões perceptuais como rouquidão, aspereza e soprosidade, e à escala destas características. Entretanto, uma quantidade de variações destes termos são encontrados na literatura e diferentes especialistas podem fornecer diferentes diagnósticos baseado no uso da sua própria percepção. Em alguns casos, este problema de discriminação patológica afeta a identificação das patologias até mesmo em exames como a videolaringoscopia e nasofibroscopia devido à carência de uma padronização e compreensão completo sobre o que é um trato vocal normal e sua dinâmica.

Em termos perceptuais, pode-se classificar genericamente a voz como normal, áspera, soprosa e rouca, independente da origem patológica. Uma voz áspera caracteriza-se por um elevado estiramento das pregas vocais, causada pelos músculos que a aduzem, exigindo assim elevados níveis de fluxo de ar vindos do pulmão. A voz soa estridente, e é caracterizada acusticamente por variações na frequência fundamental e amplitude do sinal glotal. Em patologias leves, pode-se identificar que de modo geral as pregas vocais não apresentam lesão. Isshiki *et al* (1966) menciona que um baixo fluxo de ar durante a fonação ocorre durante o comportamento hiperfuncional da laringe, acarretando a presença de pouco ruído nas altas frequências.

A soprosidade por sua vez caracteriza-se por um funcionamento adequado da mucosa e musculatura da laringe; entretanto, seja por alguma alteração orgânica nas pregas vocais ou mal funcionamento dos músculo adutores, ocorre escape de ar na região glotal, que origina ruído espectral em todas as frequências, predominando por um aumento principalmente nas altas frequências (acima do segundo formante). Isto advém do fato de que as mucosas de ambas as pregas não estão totalmente aduzindo a corda vocal nos instantes necessários; ou seja, o efeito de Bernoulli não é capaz de, por sucção, manter o espaço glotal fechado durante os instantes de elevação da pressão subglotal.

Existem estudos relacionando a quantidade de ruído com fatores que provocam turbulência por fechamento incompleto da glote, pois a velocidade do ar pela cordas vocais excede o valor crítico de Reynolds; o número crítico encontrado para um modelo físico, tamanho real, de laringe foi 2000 (Isshiki, 1978) para a percepção de sopro.

O último elemento perceptual da voz é a rouquidão, que nada mais é que a combinação dos efeitos tanto de sopro quanto de aspereza. A corda não vibra mais tão periodicamente e ocorrem escapes de ar, produzindo assim ruído turbulento e variações acentuadas na frequência fundamental.

Avaliando-se acusticamente a voz em termos perceptuais, pode-se chegar ao conjunto patológico que originou tal modificação acústica. Cabe lembrar que pode ocorrer situações em que “ouve-se” um determinado padrão acústico e nenhuma patologia é verificada em exames de videolaringoscopia e videoestroboscopia. Isshiki (1969) alerta para o fato que devido a complexa estrutura de funcionamento do sistema produtor de fala, patologias com características soprosas podem ser confundidas com as perceptualmente definidas como roucas.

Algumas patologias de origem funcional apresentam este problema, uma vez que as estruturas não apresentam qualquer problema orgânico, o fluxo respiratório é normal e a movimentação das estruturas é adequada. Entretanto, o mal funcionamento de algum músculo ou ligamento pode incorrer na incompleta adução das pregas vocais, caracterizando assim uma sopro, pela existência de ruído turbulento.

O próprio mal uso das estruturas de fonação podem acarretar no aparecimento de alterações orgânicas pelo uso excessivo de determinadas estruturas, principalmente as cordas vocais e musculatura anexa. Nódulos e úlceras de contato podem desenvolver-se pela produção constante de fala em alturas elevadas, acarretando, acusticamente, em sopro e/ou rouquidão.

O adequado estado das cordas vocais, sua densidade muscular, o estado da mucosa que a reveste permitem que ambas sejam aduzidas otimamente, possibilitando a criação de uma condição ideal de pressão e velocidade do fluxo de ar, gerando assim o sinal quase-periódico que excitará a estrutura supraglotal.

Um mal funcionamento de qualquer músculo intrínseco (principalmente) acarretará uma vocalização soprosa. Quando ocorrer uma superadução prolongada das cordas vocais, estas ficam passíveis de originarem uma fenda na região posterior, também causando escape de ar. Além disto, a voz naturalmente apresenta uma aspereza, pois as exigências de fluxo de ar aumentam (para permitir a vibração da mucosa) com a maior distensão das pregas vocais. Qualquer interferência nas bordas de aproximação das pregas vocais resultará em disфонia e ruído turbulento, ou seja, aparecimento de sopro na avaliação perceptual.

A cavidade supraglotal quando não funcionando adequadamente pode produzir efeitos sobre a fala, dada sua capacidade de produzir ênfase ou atenuação sobre o sinal glotal. O mal acoplamento das cavidades em regiões como das falsas pregas vocais, pregas ariepiglóticas, constrições da faringe

e língua podem bloquear a passagem de ar, produzindo disfonia. Além disto, a ênfase inadequada pode produzir fonemas não desejados, efeitos coarticulatórios impróprios com as ansiedades da pessoa, modificando-se assim a envolvória espectral e as frequências formantes.

Para melhor compreensão, segue uma descrição das patologias apresentadas na tab. 1, definindo-as principalmente quando a percepção acústica, para posterior compreensão de quais as melhores medidas acústicas identificarão tal patologia, que outras.

2.2.1 Úlceras de contato

Consistem de “crateras” com uma estrutura escamosa muito espessa sobre o tecido conectivo, aliada à alguma inflamação. Utilizando laringoscopia pode-se notar ulcerações em ambas as cordas vocais com grande acúmulo de tecido granuloso ao longo das bordas de aproximação na porção posterior das pregas.

A origem deste distúrbio é relacionada primeiramente com o uso freqüente de limpeza de garganta e tosse, associado com a utilização de freqüência fundamental baixa (som grave) e com ataque vocal excessivamente brusco. A segunda origem é o refluxo gastroesofágico, que irrita as paredes das estruturas laríngeas; e a terceira etiologia é a entubação do paciente em cirurgias.

Acusticamente os sintomas desta patologia é o aparecimento de níveis de rouquidão e aspereza, sendo inicialmente verificado a aspereza, devido às exigências que a limpeza das cordas vocais e os ataques bruscos às cartilagens aritenóideas. Com isto, aliado a fala grave (e/ou refluxo), as pregas vocais ficam susceptíveis às alterações citadas.

2.2.2 Diplofonia

Qualquer estrutura do trato vocal que se aproxime, formando uma constrição, pode acarretar uma diplofonia. A principal causa desta patologia é a alteração de massa-tamanho-tensão nas pregas vocais de tal forma que estas vibrem de modo diferente uma da outra.

Isto acarreta no aparecimento, em termos acústicos, de um sinal com duas componentes harmônicas distintas, não necessariamente múltiplas entre si. É semelhante ao sinal produzido por duas fontes excitatórias distintas. Geralmente, uma destas frequências “fundamentais” é a principal, com mais energia espectral, em relação à outra.

2.2.3 Falsete

Caracteriza-se por uma voz com freqüência fundamental acima da gama de alturas atingida por pessoas normais e geralmente produzido por mal uso, e portanto funcional (podem existir casos de origem endócrina ou orgânica).

Em termos acústico a voz de falsete apresenta elevada altura, associada à escapes de ar e desta forma caracterizando soprosidade. Ocorrem também quebras de freqüência fundamental para baixo.

As cordas vocais na porção anterior vibram adequadamente, enquanto que na porção posterior elas não aproximam o suficiente para a mucosa bloquear o fluxo aéreo.

2.2.4 Afonia Funcional

Definida como um sintoma histérico ou de conversão, normalmente encontrada em pacientes que tornaram-se afônicos temporários antes de sua condição permanente. Também pode ocorrer em pacientes após alguma patologia laríngea ou outra doença.

Os pacientes emitem fonemas sussurrados, o que caracteriza a rouquidão, ou seja, intensa soproidade, aliada à aspereza, como se é notado pelo som fraco agudo e estridente.

2.2.5 Disfonia Funcional

Corresponde aos problemas vocais que não são de origem orgânica, em que as pregas vocais aproximam-se de modo relaxado, não permitindo a total constrição da laringe por parte da mucosa, ou por uma aspereza devido à alta tensão exercida sobre as cordas vocais.

Outros movimentos irregulares caracterizam a disfonia funcional, como fechamento produzido pelas pregas ventriculares (falsas pregas), ou das pregas ariepiglóticas.

A percepção da voz, neste caso pode ser rouca, áspera ou soproosa, de acordo com o tipo de movimento que a prega vocal realiza.

2.2.6 Quebras de Vocalização

Perda da voz por um curto período de tempo. O indivíduo fala normalmente, e repentinamente tem uma cessação completa da voz. Geralmente caracterizado por uma hiperfuncionalidade das pregas vocais, ou por perda de pressão subglótica e até mesmo pela presença excessiva de muco sobre as paredes das cordas vocais.

Em termos sonoros, caracteriza-se pelo aparecimento de um trecho não-vocálico, como uma rouquidão intensa, durante a emissão de um sinal vocálico.

2.2.7 Quebras de Altura

Duas são as origens deste distúrbio: a primeira é observada na puberdade quando o indivíduo sofre modificações hormonais e as estruturas laríngeas sofrem marcante desenvolvimento. A segunda causa advém do uso intensivo da voz por um período prolongado, geralmente quando a mucosa vibra em uma frequência inapropriada à que está acostumada.

Analisando o sinal de voz, verifica-se que esta apresenta uma mudança repentina na frequência fundamental em uma oitava, elevando-a ou reduzindo-a de acordo com o nível, inapropriado, de vibração glotal.

2.2.8 Disfonia Espasmódica

A voz do paciente é caracterizada por uma voz disfônica ou normal, com súbitos momentos de afonia, seguida de restituição da condição anterior. Os movimentos da laringe podem ser hiperadutores ou hiperabdutores, correspondendo às formas mais extremas de hiperfuncionalidade da voz.

Os pacientes apresentam, em instantes não periódicos, condições de aspereza ou soprosidade intensa por breves períodos quando da pronúncia de um fonema sustentado. Em termos espectrais, no tipo adutor pode-se perceber ruídos em alta frequência, estrutura formante bem definida e irregularidade energética nas componentes harmônicas intra-formante, enquanto que no abductor ocorre um excessivo nível de ruído em altas frequências e descaracterização da estrutura formante.

2.2.9 Espessamento de Pregas Vocais

Caracteriza-se pelo aumento volumétrico ao longo da borda as pregas vocais, resultante na maioria das vezes de um abuso contínuo da voz, podendo propiciar o aparecimento de nódulos. Perceptualmente, a voz apresenta-se como rouca, devido à modificação de massa nas pregas vocais, influenciando assim a vibração da mucosa.

2.2.10 Laringite traumática

As cordas vocais têm suas bordas inchadas e espessas, acompanhado por maior concentração sangüínea. Desta forma, com o aumento de massa e volume das pregas vocais, sua vibração torna-se inadequada, provocando assim uma voz rouca, de baixa intensidade.

2.2.11 Disfonia Ventricular

Definida por uma adução das pregas ventriculares (falsas pregas) que vibram com a passagem do ar oriundo do pulmão. As pregas verdadeiras estão ligeiramente mais abertas durante a fase de vocalização.

A voz distingue-se por ter pouca variação de altura e ser de baixa frequência, motivada pela maior massa destas pregas em relação às cordas vocais verdadeiras. Algumas vezes pode-se perceber um padrão diplofônico, uma vez que ambas as cordas vibram.

2.2.12 Nódulos

Produzidos por uso contínuo e abusivo das pregas vocais, são caracterizados por uma protuberância esbranquiçada na borda glótica das cordas vocais, geralmente bilateral, localizados na porção média da membrana vocal. Por estarem posicionados em uma camada muito superficial das pregas vocais, interferem com a vibração da mucosa, produzindo aperiodicidade no sinal glotal.

Na adução das pregas vocais, os nódulos impedem a sua completa aproximação, produzindo uma fenda de cada lado por onde o ar turbulento escapa, gerando assim sopro.

2.2.13 Cistos

Lesão benigna resultante do acúmulo de fluídos em uma estrutura do tipo saco. Como as pregas vocais estão alinhadas com pequenas glândulas de secreção de muco (que auxilia as cordas a vibrarem mais facilmente), alguma destas pode parar a sua emissão de fluido e este começa a se acumular, produzindo uma protuberância na prega vocal. O cisto, diferente do nódulo, ocorre na maioria das vezes somente em uma das pregas vocais e em qualquer posição destas.

Por ser uma protuberância na parede glotal, impede sua total adução, permitindo escape de ar, turbulento, dando o aspecto sopro à voz do paciente.

2.2.14 Pólipos

Relacionada com hiperfuncionamento da vocalização, são resultados de um único evento vocal (não contínuo), podendo ser identificado como uma haste penducular ou sésil (base ampla) na borda interna de uma das pregas vocais, pouco irritando a mucosa devido a sua maciez.

A voz de um paciente com esta patologia é sopro (devido a adução não ser total) e dependendo do tamanho e localização, pode abafar a vibração da mucosa, produzindo assim rouquidão.

2.2.15 Edema de Reinke

Também chamado de degeneração polipoidal, é descrito como um aumento de massa, bilateral, nas cordas vocais causado por edema no corpo gelatinoso de Reinke, logo abaixo da mucosa.

Com a elevação do tamanho das pregas vocais, a frequência fundamental é reduzida, tornando mais grave a voz, tornando-se instável, além de ser acompanhada de sopro. Esta conjunção de efeitos, pode produzir uma percepção de rouquidão na voz do paciente.

2.2.16 Câncer

Também denominado de carcinoma, pode afetar qualquer porção do trato vocal, produzindo efeitos diversos sobre a voz produzida por um paciente.

Considerando as lesões que afetam as pregas vocais (câncer glotal), dependendo o tipo (T1, T2, T3 ou T4), localização e extensão, pode-se ter desde um sopro leve, até o aparecimento de severa rouquidão.

2.2.17 Disartria

Problema motor da fala decorrente de dano em alguma parte do sistema nervoso central ou periférico, podendo resultar em paralisia unilateral ou bilateral das cordas vocais, ou musculatura anexa, que são acionados por nervos recorrentes.

O descontrole nervoso do trato vocal pode caracterizar problemas de prosódia, coarticulação e alterações na intensidade da voz, assim como altura e qualidade de vocalização.

2.2.18 Mudanças endócrinas

Alterações hormonais tem relevância sobre as características vocais das pessoas, causando modificações sobre a frequência fundamental, que pode ser extremamente grave ou aguda.

Alterações na produção hormonal da tireóide pode produzir voz perceptualmente rouca, com altura extremamente baixa. Níveis de estrogênio e progesterona nas mulheres pode produzir leve espessamento das pregas, causando redução de altura e rouquidão. Hormônios androgênicos na menopausa tornam as paredes das cordas vocais mais espessas, podendo torná-la áspera.

Os efeitos deste tipo de disfunção sobre as estruturas vocais devem ser considerados quando for realizada a seleção de pacientes (e até mesmo o grupo de controle normal) para os estudos de discriminação patológica.

2.2.19 Granuloma

Caracterizado por níveis de rouquidão e sopro elevados, são resultados de traumatismo nas estruturas da laringe (produzidas por entubação, mal uso das pregas vocais e até mesmo refluxo gastroesofágico). A mucosa apresenta hiperplasia e granulação, de aspecto rígido.

2.2.20 Hemangioma

Semelhante ao granuloma, diferenciando no aspecto das pregas vocais, em que se tem uma bolsa bem menos rígida, preenchida de sangue.

É geralmente encontrada sobre a região posterior das cordas vocais, com mesmas características etiológicas e vocálicas que o granuloma.

2.2.21 Hiperkeratose

Resultante da irritação contínua do tecido laríngeo, esta patologia caracteriza-se por placas rosadas, que surgem sobre as pregas vocais, instalando-se geralmente sobre a comissura anterior.

Seus efeitos sobre a voz podem ser de desprezíveis à severos, com o aparecimento de níveis de sopro e rouquidão.

2.2.22 Laringite infecciosa

Pacientes advindos de laringite funcional que apresentam abuso ou mal uso das estruturas de fonação acabam experimentando esta patologia tão logo tenham resfriado.

A voz destes é caracterizada por intensa soproidade uma vez que esta, apesar de inchadas, apresenta vibração de mucosa.

2.2.23 Leucoplasia

Lesões de acréscimo sobre a membrana superficial da mucosa, apresentando coloração esbranquiçada, sendo consideradas pré-cancerosas.

Com o acréscimo de massa, reduz-se a altura da voz, provocando redução da frequência fundamental e de intensidade fraca. Em várias ocasiões as pregas vocais tornam-se assimétricas, gerando diplofonia.

2.2.24 Papiloma

Limitando a passagem do fluxo de ar pelas cordas vocais, os papilomas são crescimentos verrugosos de origem viral. Ocasionalmente a percepção de rouquidão na voz do paciente, pelo aumento de massa.

2.2.25 Paralisia de Pregas Vocais

Qualquer dano sobre o nervo laríngeo superior ou no nervo laríngeo recorrente, que controlam as atividades musculares nas estruturas de vocalização, resultam em uma paralisia (unilateral ou bilateral) das pregas vocais, impedindo a produção de um sinal glotal adequado. No caso unilateral, uma das pregas vocais fica na posição abduzida, cabendo à outra a função de aproximar-se, compensando a inatividade da primeira.

A voz apresenta-se extremamente soproosa, uma vez que as pregas vocais não conseguem aduzir-se à uma mínima abertura para prover vibração da mucosa. No caso unilateral, havendo compensação da prega vocal normal, pode-se obter um padrão diplofônico, com percepção de soproidade.

2.2.26 Membrana Laringiana

Crescimento de uma membrana pela glote, nas duas pregas vocais, impedindo a vibração correta destas, podendo causar respiração curta, tendo origem congênita ou resultado de um trauma nas bordas internas das cordas vocais.

A voz é descrita perceptualmente como áspera, com alta frequência fundamental, uma vez que esta membrana tende a reduzir o espaço glotal, geralmente iniciando pela comissura inferior.

Analizando as patologias e os padrões acústicos descritos, pode-se perceber que a tarefa discriminatória através de informações unicamente perceptuais é extremamente complicada. A simples diferenciação entre pólipos e nódulos pode ser trabalhosa se conduzida desta forma. Considerar diretamente os índices acústicos que podem ser obtidos dos sinais de voz pode produzir resultados mais diretos e de maneira mais eficiente.

A maioria das patologias advém do mal uso da função vocalizadora das cordas vocais. À medida que prolongam-se os excessos, mudanças nos tecidos que revestem a laringe produzem diversos distúrbios.

Verifica-se que mesmo em patologias de etiologia diferente apresentam características sonoras semelhantes. A discriminação patológica então serviria como uma ferramenta útil no auxílio ao diagnóstico.

2.3 Modelo Matemático do Trato Vocal

De modo a fundamentar adequadamente as análises que serão realizadas sobre os sinais de voz (tanto normal quanto patológica), é importante definir um modelo matemático do trato vocal simples e que defina adequadamente a produção da fala, uma vez que as alterações patológicas podem produzir sinais de voz cujo sistema gerador não corresponde ao modelo proposto, e desta forma, as análises podem induzir à conclusões errôneas.

1. O trato vocal é constituído por todas as estruturas responsáveis pela produção da fala, citadas em detalhe anteriormente, podendo ser dividido funcionalmente em três estruturas distintas, apesar da interrelação que ocorre fisicamente no ser humano:
2. Um elemento produtor de excitação, caracterizado biologicamente pelo pulmão, suas estruturas de transporte de ar, traquéia e esôfago.
3. Um sistema produtor de oscilações periódicas, definido organicamente pelas cordas vocais e órgãos anexos (aritenóides, falsas cordas, entre outras), provocando constrições na laringe, que as suporta. Desta forma o fluxo de ar, advindo da traquéia, sofre modificações de pressão ao longo do tempo, caracterizando uma onda de pressão quase-periódica não senoidal.
4. Um sistema produtor de ressonância, compreendendo os órgãos da cavidade supraglotal (faringe, língua, véu palatino, entre outros), produzindo uma envoltória sobre as componentes harmônicas, caracterizando as estruturas formantes da fala. Ou seja, molda através de ressonância (de modo análogo ao que ocorre em órgãos) os fonemas e a coarticulação entre eles (a prosódia é um elemento que exige a atuação sincronizada dos três elementos aqui citados).

Para facilitar a construção do modelo matemático que represente o sistema fonador humano, simplifica-se as estruturas acima descritas em somente dois blocos distintos, responsáveis pela produção da fala, através da junção do excitador “pulmonar” ao gerador de oscilações “glotal”, de modo que se tenha somente um único “subsistema” capaz de produzir o sinal oriundo das cordas vocais. Outra simplificação útil é a de considerar o “subsistema” ressonador como uma estrutura matemática linear, mesmo que variante no tempo (pelas características fonéticas associadas com as alterações nos órgãos supraglotalis), pois as equações tornam-se mais fáceis de serem manipuladas de modo algébrico. Com tais pensamentos, tem-se abaixo a representação gráfica de tais considerações.

Figura 2 - Representação Simplificada do Sistema Produtor de Fala

2.3.1 Excitação glotal e considerações físicas associadas

Analisando agora o sinal proveniente das pregas vocais (já considerando o modelo simplificado), três diferentes excitações são possíveis de geração pelo primeiro “subsistema”:

1. um sinal quase-periódico resultante da ação vibratória das cordas vocais quando da passagem do fluxo de ar proveniente dos pulmões;
2. um sinal ruidoso, proveniente da turbulência sofrida pelo fluxo de ar pulmonar em grande velocidade quando da passagem pelas cordas vocais, que estão relaxadas;
3. um sinal de excitação transiente, resultado da elevação de pressão na porção anterior às cordas vocais, causado pela constrição total da mesma, seguido de uma rápida abertura glotal.

Estes sinais são importantes na produção fonética, pois cada um está associado a grupos de fonemas distintos, em que a configuração da cavidade ressonadora pode ser a mesma, onde temos sons vocálicos, fricativos ou não vocálicos e plosivos, respectivamente associados aos três tipos de sinais glotais.

A caracterização dos sinais de excitação, bem como a fisiologia do trato vocal para a sua produção, é importante para que seja possível idealizar um modelo representativo para cada padrão patológico, e as influências da alteração sobre as considerações simplificadoras quando estas não mais ocorrerem, como por exemplo, a produção de voz soprosa advinda da adução incompleta das pregas vocais.

O processo de vibração das cordas vocais (na produção de uma onda quase-periódica para fonemas vocálicos) inicia-se pela elevação da pressão nos pulmões, criando um fluxo de ar que atravessa a abertura glotal. Baseado nas leis de Bernoulli para fluidos, com o aumento da velocidade do ar, a pressão nas paredes das cordas vocais é menor que na sua vizinhança. Considerando as características de tensão aplicadas pela musculatura acessória sobre as pregas vocais, ambas começam a constrição da laringe, devido a transferência de energia cinética do movimento fluxo aéreo em energia potencial elástica nas cordas vocais. Com este acréscimo de energia e perda consecutiva de pressão de ar (energia potencial), as paredes da porção subglotal fecham-se e, conseqüentemente, com a energia potencial elástica acumulada no restante da corda, esta aduz-se completamente.

Neste ponto, inicia-se o incremento de energia potencial abaixo da glote, uma vez que a mesma encontra-se fechada (velocidade do fluxo de ar nulo por entre as pregas vocais). A medida

que a energia potencial elástica torna-se cada vez mais inferior à pressão do ar, as cordas vocais iniciam o processo de abdução, forçando a abertura das cordas vocais. No momento de abertura plena da corda vocal, a energia potencial começa a reduzir-se pela elevação da energia cinética do ar por entre as paredes da glote.

Este regime oscilatório é mantido mediante o controle da pressão exercida pelos pulmões, tensão e espessura das cordas vocais e área de abertura glotal. Estes parâmetros são importantes na formulação do modelo de produção vocal, e são de igual validade na avaliação das modificações acústicas (e por conseguinte, no modelo) quando a glote não apresenta mais estas condições de operação (por alguma patologia).

Já a produção de sons não-vocálicos, duas situações permitem a produção de ruído turbulento. A primeira envolve a elevação da velocidade do fluxo de ar além de um patamar característico (dependente de características anatômicas das cordas vocais). A segunda situação caracteriza-se pela adução inadequada das pregas vocais, promovendo parcial fechamento de tais estruturas nos instantes necessários.

A terceira forma sonora suportada pelas pregas vocais envolve a retenção de ar na subglote e conseqüente elevação de pressão de ar, seguido da liberação rápida resultando em um sinal turbulento devido ao posicionamento glotal e a velocidade do ar. A transformação de energia potencial em energia cinética, aliada ao relaxamento da corda vocal impede que os movimentos vistos durante a produção de fonemas vocálicos sejam possíveis nesta configuração do sistema produtor de fala.

Principalmente a segunda e a terceira forma de sinal glotal terão relevância neste trabalho quando forem consideradas as patologias da laringe e cordas vocais, visto que apesar do paciente estar emitindo fonemas vocálicos, os sons serão semelhantes aos produzidos como se a configuração pulmão-glote desse suporte à sinais fricativos ou plosivos.

2.3.2 Sistema Linear Variante no Tempo - Supraglotal

A estrutura de ressonância (organicamente representada pelos órgãos supraglotalis) é responsável pela caracterização fonética, transformando o sinal de excitação em fonemas, mediante movimento articulatorio de músculos e cartilagens que promovem acoplamento e desacoplamento de cavidades; estas cavidades influenciam diretamente no processo de ênfase energética de determinadas frequências e a atenuação de outras.

As componentes espectrais, correspondentes às frequências naturais de ressonância, levam o nome de formantes, podendo-se determinar infinitas frequências formantes (iniciando-se pelas mais inferiores); entretanto, seu número na prática fica restrito a faixa de 3 à 5 formantes devido aos requisitos do teorema de Nyquist para sinais amostrados. Cabe lembrar que devido à características particulares (anatômicas e fisiológicas) entre pessoas, as frequências de ressonância apresentam certa variabilidade, tanto na frequência central quanto na largura de banda.

Além disto, como já citado anteriormente, a alteração das cavidades do trato supraglotal, mesmo que lentas, fazem com que seja considerado o aspecto de variabilidade temporal dos modelos descritivos deste “subsistema”.

2.3.3 Modelo Acústico da Fala

Para iniciar a descrição matemática do sistema de produção de fala, descreve-se primeiramente a propagação da onda sonora pelo trato vocal. Os princípios físicos de conservação de energia, momento e massa através dos órgãos da fonação, além das considerações de baixa viscosidade e compressibilidade do ar são levadas em conta na formulação.

Além do formato do trato vocal, sua variabilidade temporal, as perdas que a onda sonora tem devido à condução térmica e fricção do ar sobre suas paredes, sua rugosidade, o efeito de radiação nos lábios, o acoplamento nasal realizado no véu palatino. Considerando estas peculiaridades do trato vocal, modelá-lo de modo conexionista seria extremamente difícil.

Uma abordagem interessante é considerar que o trato vocal é formado por tubo cilíndrico, com área variante no tempo, de seção transversal não-uniforme. A onda sonora então propaga-se por este tubo, admitindo-os sem perdas de energia ao longo do traçado. Assim, as equações diferenciais de movimento da onda (Rabiner e Schaffer, 1978) são:

$$\begin{aligned} -\frac{\partial p}{\partial x} &= \rho \frac{\partial(v/A)}{\partial t} \\ -\frac{\partial v}{\partial x} &= \frac{1}{\rho c^2} \cdot \frac{\partial(vA)}{\partial t} + \frac{\partial A}{\partial t} \end{aligned} \quad (1)$$

onde $p = p(x, t)$ é a variação da pressão sonora na posição x , no instante t .

$v = v(x, t)$ é a variação da velocidade do ar na posição x , no instante t

ρ é a densidade do ar no tubo

c é a velocidade do som

$A = A(x, t)$ é a área da secção transversal na posição x , no instante t .

Supondo a uniformidade do tubo e admitindo que nas aberturas deste não exista variação de pressão, tendo-se apenas variação na velocidade volumétrica do ar, tem-se que a solução para o par de equações diferenciais é:

$$\begin{aligned} v(x, t) &= [v^+(t - x/c) - v^-(t + x/c)] \\ p(x, t) &= \frac{\rho c}{A} [v^+(t - x/c) + v^-(t + x/c)] \end{aligned} \quad (2)$$

onde os superscritos na função v representam a direção do movimento das ondas.

Considerando as representações no domínio da frequência (através do uso de transformadas) e aplicando as condições de contorno para a glote e os lábios, tem-se a relação entre as velocidades de entrada (onda glotal) e velocidades de saída (lábios) nos tubos sem perda dada por:

$$V(s) = \frac{2e^{-sl/c}}{1 + e^{-s2l/c}} \quad (3)$$

onde l corresponde ao comprimento total dos tubos concatenados que representam os órgãos supraglóticos (aproximadamente 17 cm).

Sabendo que os pólos da função $V(s)$ (que é um sistema linear invariante no tempo, pela suposição da lenta alteração temporal do mesmo, e desta forma, podendo ser considerado estacionário por um janela de tempo) correspondem às frequências naturais do sistema. No caso da fala, são chamadas formantes. São em número infinito e dadas por:

$$f_s^k = j\left(\frac{(2k+1)\pi c}{2l}\right), \quad k=0, \pm 1, \pm 2, \dots \quad (4)$$

Verifica-se que a modelagem através da propagação da onda sonora no tubo, apesar dos resultados mostrarem-se de acordo com os obtidos através do sistema de produção de fala humano, não se percebe uma correlação específica entre os órgãos supraglóticos (à exceção dos lábios que tem uma grande influência na perda energética das altas frequências no sinal de voz e o efeito de radiação produzido será caracterizado nas equações). Esta característica é definida pelo que se conhece por modelo terminal-análogo em que apesar de algumas estruturas internas do modelo não são necessariamente equiparáveis às do sistema produtor de fala, mas os sinais produzidos por ambos são análogos.

Agora, aplicando os efeitos de perda de energia gerados pela condução térmica e fricção do ar nas paredes da supraglotal, constatou-se que desconsiderando-os da modelagem, seu efeito não se compara à perda oferecida pela difração da onda sonora nos lábios. Na modelagem do efeito de radiação, constata-se que existe uma atenuação progressiva na energia com o aumento da frequência, e desta forma, atenuando também a estrutura formante.

Estas considerações matemáticas realizadas no domínio da transformada de Laplace, aliadas à caracterização do trato supraglotal em termos de um tubo cilíndrico, têm por objetivo permitir uma transição destes modelos contínuos para modelos discretos, no tempo.

Considerando agora um conjunto de tubos concatenados, de áreas A_k , representando uma aproximação do tubo anteriormente considerado (discretização do tubo) e considerando que as perdas devido a rugosidade, fricção e condução térmica das paredes podem ser contabilizadas pelas perdas nos lábios, pode-se obter equações para cada tubo, ou seja:

$$\begin{aligned} v_{k+1}^+(t) &= (1+r_k) \cdot v_k^+(t-\tau_k) + r_k \cdot v_{k+1}^-(t) \\ v_k^-(t+\tau_k) &= -r_k \cdot v_k^+(t-\tau_k) + (1-r_k) \cdot v_{k+1}^-(t) \end{aligned} \quad (5)$$

onde

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \quad (6)$$

A_i a área do i -ésimo tubo representativo do trato vocal

r_i o coeficiente de reflexão

T_i o tempo para a onda propagar-se pelo i -ésimo tubo

$v_i^+(t)$ a velocidade do fluxo de ar na direção positiva no i -ésimo tubo no tempo t .

$v_i^-(t)$ a velocidade do fluxo de ar na direção negativa no i -ésimo tubo no tempo t .

Este modo de equacionamento permite a construção de representações de fluxo de sinal, muito utilizada em processamento digital de sinais, além de permitir a simulação mais direta dos modelos em ambientes computacionais. A utilização destas equações deve considerar as condições iniciais, ou condições de contorno, impostas pela glote e pelos lábios, dadas por:

$$\begin{aligned} v_1^+(t) &= \frac{(1+r_G)}{2} \cdot v_G(t) + r_G \cdot v_1^-(t) \\ r_G &= \left(\frac{Z_G - \frac{\rho c}{A_1}}{Z_G + \frac{\rho c}{A_1}} \right) \end{aligned} \quad (7)$$

e

$$\begin{aligned} v_L(t) &= (1+r_L) \cdot v_N^+(t-\tau_N) \\ r_L &= \left(\frac{\frac{\rho c}{A_N} - Z_L}{\frac{\rho c}{A_N} + Z_L} \right) \end{aligned} \quad (8)$$

respectivamente, onde Z_G e Z_L correspondem à impedância glotal e dos lábios respectivamente para o modelo digital tratado.

Verifica-se neste momento que uma parte do sinal propagado em cada tubo é transmitido ao tubo sucessor e outra porção ao tubo antecessor. Avaliando este fato, certamente pode-se concluir que efeitos de realimentação devem ser considerados quando da transformação para domínio da frequência, dando suporte à pólos e zeros em funções de transferência.

Das estruturas de fluxo de sinal (que podem ser obtidas das expressões acima), pode-se perceber a sua relação com filtros digitais uma vez que considerando que o comprimento de todos os

tubos concatenados, que representam um trecho do trato supraglotal, têm o mesmo tamanho e desta forma apresentam um atraso equivalente. Esta nova “conversão” possibilita o emprego de um conjunto ferramental de teoria de sistemas de tempo discreto, permitindo a avaliação da complexidade das operações à serem realizadas para a produção da fala.

Refletindo novamente sobre o modelo terminal-análogo, levando-se em conta a relação com sistemas de tempo discreto, pode-se determinar uma relação entre o sinal glotal e o sinal de fala dada por:

$$V(z) = \frac{H}{1 - \sum_{k=1}^N \alpha_k z^{-k}} \quad (9)$$

onde H e $\{\alpha_k\}$ são valores dependentes da função de área do trato supraglotal (H corresponde ao ganho geral do sistema e o denominador representa os atrasos temporais equivalente de propagação da onda sonora pelos N tubos concatenados). Cabe salientar que o sistema de produção de fala é variante no tempo; assim, os parâmetros H e $\{\alpha_k\}$ devem ser modificados de modo a adequarem-se aos novos sinais de voz (e efeitos de coarticulação) durante o tempo. Como já citado, estas variações temporais são lentas no sistema orgânico, sendo que mantém-se parcialmente fixas por um período entre 10 a 20 ms. Esta consideração será importante quando da construção de algoritmos de análise do sinal de voz, seja normal ou patológico.

Diante das reflexões feitas, o subsistema ressonador pode ser representado adequadamente por um modelo somente-pólo para a maioria dos fonemas (sons fricativos e nasalados necessitam de zeros adicionais para serem produzidos pelo modelo matemático, ou pela inclusão sistemática de mais pólos). Utilizando a eq. (9), tem-se que os pólos são dados por (em pares conjugados, com a presença de uma raiz real se N for ímpar):

$$f_z^k, f_z^{k*} = e^{-\sigma_k T} e^{\pm j 2\pi F_k T} \quad (10)$$

onde a formante é encontrada em $2\pi F_k$, com largura de banda aproximadamente de $2\sigma_k$.

Figura 3 - Espectro densidade de potência para o modelo de uma configuração específica da estrutura supraglotal

A fig. 3 apresenta a resposta em frequência de uma determinada configuração do trato supraglotal para exemplificar, onde não são computadas as perdas energéticas devido a temperatura e rugosidade das paredes daquelas estruturas e principalmente da perda por radiação nos lábios. As ressonâncias são visíveis (frequências formantes) nas regiões dos pólos.

O sinal de fala entretanto ainda não é totalmente descrito somente por $V(z)$. Existe ainda a consideração sobre o efeito de radiação dos lábios que, como já mencionado, reduz a energia das altas frequências. Desta forma é necessário prover tal característica no modelo de produção de fala. Uma aproximação deste efeito é dada por:

$$R(z) = R_0(1 - z^{-1}) \quad (11)$$

por ser a relação entre a velocidade e a pressão do ar nos lábios uma operação de filtragem em passa-alta. A perda energética das componentes de altas frequências devido a radiação pode ser vista na figura 4.

Figura 4 - Espectro densidade de potência considerando as perdas devido a radiação nos lábios.

Conforme já citado anteriormente, o efeito da radiação dos lábios deve ser considerado na formulação de medidas acústicas para identificação das patologias.

2.3.4 Modelo Completo

Uma vez que o trato vocal foi estudado em porções distintas, através da avaliação do gerador de excitação, do trato supraglotal e dos efeitos de perda energética (concentrados no modelo de radiação dos lábios), determina-se agora, através da junção de todas as “peças”, o modelo completo do trato vocal. Tal modelo pode ser visto na figura abaixo.

Figura 5 - Modelo Representativo do Trato Vocal

O valores de ganho, A_v e A_n , levam em consideração os efeitos de maior ou menor pressão exercido pelo pulmão na fonação.

O modelo do pulso glotal é necessário para que se produza um sinal similar ao produzido pela vibração das cordas vocais de modo a excitar adequadamente a supraglotal artificial e produzir o sinal de voz desejado. Considerando que para sinais vocálicos tem-se uma onda quase-periódica, utilizando um gerador de trem de pulsos, separados por um período de tempo denominado período de *pitch* (ou frequência fundamental), o sistema que produz o sinal advindo das cordas vocais é dado por:

$$g(n) = \begin{cases} \frac{1}{2} \left[1 - \cos\left(\frac{\pi n}{N_1}\right) \right], & 0 \leq n \leq N_1 \\ \cos\left[\frac{\pi(n - N_1)}{2(N_1 - N_2)}\right], & N_1 \leq n \leq N_2 \\ 0, & c.c. \end{cases} \quad (12)$$

Este sistema não é único e definitivo. Foi construído com base em propriedades espectrais e similaridade do sinal produzido com uma onda real glotal. Os parâmetros N_1 e N_2 são definidos com base na curva desejada para o sinal de saída deste modelo e no efeito de filtragem passa-baixa característico deste sistema. A figura abaixo apresenta um exemplo de sinal glotal obtido a partir da eq. 12, definida empiricamente.

Figura 6 - Trecho periódico do sinal glotal

O modelo para o sinal quando na necessidade de emissão de sinal não-vocálico é obtido na produção de sinal aleatório sem necessidade de consideração restritiva de distribuição de probabilidade das amostras produzidas. Normalmente emprega-se geradores de sinais aleatórios com espectro de ruído branco.

Com todos os elementos agora definidos, todos os sistemas podem ser cascadeados, dando origem à uma única função de transferência que rege este modelo simplificado de trato vocal

$$H(z) = G(z) \cdot V(z) \cdot R(z) \quad (13)$$

onde G representa o modelo do pulso glotal, responsável pela produção da onda quase-periódica para sinais vocálicos e ruído branco para sinais não-vocálicos, V representando a influência moduladora do trato supraglotal, caracterizando assim as formantes e deste modo, os fonemas, e R produzindo a influência das perdas acústicas ao longo do trato vocal e principalmente devido à difração da onda sonora nos lábios.

Indicar as limitações deste modelo é importante para que se compreenda corretamente os resultados que serão produzidos pelas análises sobre as vozes (principalmente pela condição patológica que se deseja compreender e até certo ponto, reproduzir).

O modelo é adequado para sinais de voz que pouco variam durante o tempo, ou seja, vocálicos e não-vocálicos (considerando a estacionariedade por períodos de tempo entre 10 a 20 ms). Sinais plosivos não são adequados para a representação apresentada, dada sua natureza transiente. Outra característica direta da variabilidade temporal de $V(z)$ é que os parâmetros $\{\alpha_k\}$ devem ser computados (num modelo somente-pólo) à cada período Δt caso o estudo dirija-se para análise de frases (agrupamento fonético).

O não emprego de zeros no modelo do trato supraglotal também implica em limitações se a produção ou análise estiver voltada para sinais nasalados ou fricativos.

Para suprir estas “falhas” do modelo aqui apresentado, pode-se associar tais condições, obtendo-se modelos mais sofisticados. Em alguns algoritmos que serão apresentados, foram levadas estas considerações, visto que a fonação de pacientes com alguma patologia (orgânica ou funcional) não é capaz, em alguns casos, manter o sinal glotal o mais periódico possível, ocasionando assim, perturbações neste, que acarretarão sinais de voz incompatíveis com o modelo simplificado.

Apesar de sucinto, os elementos básicos apresentados neste capítulo servem de base para compreender as manipulações matemáticas que serão realizadas sobre o sinal de voz, a fim de obter-se parâmetros acústicos que representem o sinal de voz, além da descrição da nomenclatura utilizada em análises acústicas e sistema de produção de fala.

2.4 Elementos Matemáticos da Análise Acústica

Esta seção tem por objetivo apresentar a fundamentação matemática sobre a qual se apoiam as análises acústicas dos sinais de voz e apresentação das características temporais e espectrais das patologias conhecidas.

A principal ferramenta de análise da voz é apresentada (espectrograma), exigindo-se sua definição matemática por técnicas de tempo-curto (*short-term*). Por possibilitar a visualização da característica espectral do sinal de voz, seja normal ou patológico, a maioria das medidas acústicas que serão realizadas são baseadas em padrões visuais que podem ser identificados no espectrograma.

Posteriormente é apresentado um resumo das medidas acústicas conhecidas na literatura, começando pelas medidas de perturbação e seguindo por índices de nível de ruído espectral. Outras técnicas disponíveis são apresentadas como emprego de sinal residual na avaliação patológica.

Alguns comentários são realizados para a compreensão da metodologia de análise da voz para que seja dada sustentação aos algoritmos desenvolvidos nesta pesquisa.

2.4.1 Janelas

Uma característica importante da voz é a sua alteração temporal, mediante mudanças na configuração do trato vocal. Efeitos coarticulatórios são importantes quando se considera a produção de frases, pois um conjunto de fonemas estão conectados ali e o trato vocal deve variar durante o tempo para produzi-los, caracterizando transientes no sinal. Até mesmo durante a fala sustentada (emissão prolongada de um fonema, normalmente vocálico) existe uma certa movimentação muscular no trato vocal (na maior parte das vezes, inconsciente), como variações na vibração da mucosa, leves alterações na posição da língua e mandíbula, que afetam desta forma o sinal de voz produzido.

Verifica-se em alguns esquemas de processamento de fala que os sinais de voz são considerados invariantes no tempo para permitir, por exemplo, avaliação da periodicidade, estimação paramétrica de modelo do trato vocal. Outro ponto importante são os efeitos que o processo de amostragem e discretização podem produzir sobre os resultados, relacionado com os estudos sobre ruído de alta frequência que pode ser afetado pelo “aliasing”.

Considerando a não estacionariedade do sinal de voz, é necessário buscar uma forma de efetuar as análises sobre as amostras levando-se em conta que por um instante de tempo (entre 100 a 200 ms) as medidas desejadas permanecem estacionárias. Assim, utiliza-se a metodologia de cálculo empregada para grandes trechos de sinal (“long-term”), aplicando-as para estes trechos curtos (“short-term”), levando em conta os desvios teóricos que esta técnica produz. A base desta suposição

é a lenta alteração das propriedades do sinal de fala e que o trecho curto isolado para análise foi retirado de um sinal sustentado com propriedades invariantes.

Os trechos do sinal que são analisados são denominados “frames” e definidos por:

$$f_s(n; m) = s(n) \cdot w(m - n) \quad (14)$$

onde $s(n)$ corresponde à n -ésima amostra do sinal de fala, $w(n)$ define uma janela de comprimento finito P , onde $w(n)=0$ para $n < m-P+1$ e $n > m$. A multiplicação destes dois sinais provê um sinal finito, de energia finita, formalmente definido.

Considerando uma medida sobre um sinal amostrado $\chi_s(\lambda)$, pode-se obter uma versão de período curto através do princípio de construção. Definindo-se primeiramente a operação sobre o sinal $s(n)$ completo por:

$$\chi_s(\lambda) = \tilde{T}(\lambda)\{s(n)\} = \xi\{T(\lambda)\{s(n)\}\} \quad (15)$$

onde $\tilde{T}(\lambda)$ é a transformação que se deseja efetuar sobre o sinal $s(n)$, que pode ser, geralmente, decomposta na expectância de uma operação em uma seqüência numérica, representado por $\xi \circ T(\lambda)$.

Uma vez definido o método de transformação, aplica-se uma janela de modo a isolar um trecho, caracterizando a versão “short-term”.

$$\chi_s(\lambda; m) = \frac{1}{P} \sum_{n=-\infty}^{\infty} T(\lambda)\{s(n) \cdot w(n - m)\} = \frac{1}{P} \sum_{n=-\infty}^{\infty} T(\lambda)f_s(n; m) \quad (16)$$

Outra forma do princípio de construção é utilizá-lo sobre o resultado da operação, aplicando-se à este a janela $w(n)$, ou seja:

$$\chi_s(\lambda; m) = \frac{1}{P} \sum_{n=-\infty}^{\infty} s_T(n; \lambda) \cdot w(n - m) = \frac{1}{P} s_T(n; \lambda) * w(n - m) \quad (17)$$

A implementação das transformações sobre os sinais baseada na aplicação da janela sobre os dados permite o emprego mais direto dos algoritmos de processamento digital de sinais em computadores, devido a restrições de memória que o mesmo possui; mesmo assim, existem medidas acústicas que são úteis considerando a segunda formulação do princípio.

Deve-se levar em conta os efeitos dos lóbulos laterais e a largura de banda do lóbulo principal produzidas pela janela (correspondendo a um filtro passa-baixa). De modo a manter a linearidade de fase, normalmente são utilizadas janelas causais, com início em $n = 0$ e simetria em $(P-1)/2$.

Um cuidado é necessário ser feito na escolha de P , o tamanho da janela. Isto porque a não estacionariedade do sinal implica em mudanças ao longo do tempo, que são importantes na análise de sinais de voz, principalmente nos fonemas vocálicos em fonação sustentada. O emprego de P

elevado permite que o lóbulo principal da janela, no domínio da frequência, tenha sua largura de banda reduzida, ao mesmo tempo que a atenuação dos lóbulos laterais seja cada vez menor; desta forma, eleva-se a resolução espectral da janela, melhorando-se a estimação da medida acústica.

Entretanto, isto acarreta problemas de resolução temporal, uma vez que a consideração de um maior número de valores para o trecho “visível” através da janela faz com que modificações rápidas no sinal não sejam “percebidas” pelo estimador. A escolha de P deve ponderar estas duas características inerentes ao processamento de sinais não estacionários.

Novamente considerando a variação temporal do trato vocal, tem-se que a variação de n não necessita ser realizada em passos unitários. Uma vez que se pode admitir a estacionariedade por um determinado período de tempo, os efeitos de sobreposição de janelas poderão ser desconsiderados se o número de passos para cada avaliação em trechos curtos de uma operação sobre o sinal for suficiente para manter a resolução temporal.

2.4.2 Espectrogramas

Esta operação é muito comum na construção de espectrogramas. Estes são extremamente úteis como forma de registro da voz, tanto de pessoas normais quanto de pacientes (Rontal, Rontal e Rolnick, 1975), sendo uma das mais importantes técnicas de análise de voz (Baken, 1996) por prover uma grande quantidade de informação mesmo durante uma fala curta. Através da transformação (discreta de Fourier ou banco de filtros), obtém-se a representação espectral para trechos curtos do sinal, que é avaliada em intervalos regulares de tempo (passos).

A resolução temporal e espectral é determinada pela largura de banda dos filtros passa-faixa usados na decomposição do sinal de voz, empregando-se 45 Hz ou 300 Hz. Estes correspondem à filtros banda-larga e banda-estreita, respectivamente.

Filtros com largura de banda de 45 Hz permitem uma melhor resolução espectral pois consegue separar duas linhas espectrais até esta frequência. Entretanto, alterações rápidas no trato vocal não podem ser visualizadas. Usando a segunda opção de filtros, pode-se avaliar melhor as alterações coarticulatórias e alterações rápidas na frequência fundamental (que se propagam pelo espectro). Entretanto, apresenta pior resolução espectral pois não é possível avaliar alterações entre distâncias harmônicas abaixo de 300 Hz. O compromisso frequência-tempo deve sempre ser considerado, seja empregando equipamentos analógicos ou técnicas digitais.

Esta relação entre larguras de banda é traduzida para as técnicas matemáticas como aumento ou redução no tamanho da janela sobre o qual determina-se o espectro de potência. Filtros de banda estreita correspondem à uma pequena quantidade de amostras por janela e desta forma, não possuem resolução espectral suficiente, apesar de possibilitar a percepção de modificações rápidas no tempo. Em contrapartida, o aumento do tamanho da janela permite uma elevação desta resolução, tornando o espectrograma digital mais preciso, com uma redução da precisão temporal. Estas diferenças

podem ser melhor compreendidas pela observação da fig. 7, onde são apresentados os espectrogramas (45 e 300 Hz) para um mesmo sinal de voz (fonema /a/ masculino).

Figura 7 - Espectrograma para o fonema /a/ masculino digitalizado à 22050 Hz, obtido por filtros passa-faixa de 45 Hz e 300 Hz, respectivamente

Quando se emprega um espectrograma de 45 Hz, percebe-se mais claramente as componentes harmônicas (estrias horizontais ao longo do gráfico). Já o segundo espectrograma possibilita a análise de alterações rápidas que possam ocorrer ao longo da fonação.

Por estas características, o emprego do espectrograma no auxílio à especialistas de saúde se dá em sinais de voz provenientes da fala sustentada, pois a configuração do trato vocal altera-se levemente durante a fonação. Assim, alterações bruscas corresponderam à disfunções (orgânicas ou funcionais) na estrutura do sistema de produção de fala.

As primeiras utilizações do espectrograma datam de 1947 através de estudos nos laboratórios Bell por Potter *et al* (1947). Dada a potencialidade visual, tanto como registro de evolução de tratamentos como avaliação de cirurgias, a sua utilização no diagnóstico de patologias torna-se útil pois pode-se identificar no espectrograma padrões acústicos de aspereza, sopro e rouquidão.

Avaliando alterações nas componentes harmônicas e quantidade de ruído no espectrograma, Yanagihara (1967) caracterizou a voz rouca como resultado de três fatores:

1. Componentes ruidosas na região da primeira formante.
2. Ruído de alta frequência a partir de 3000 Hz.
3. Perda de energia espectral nas componentes harmônicas a partir de 3000 Hz.

Verificou-se também que as patologias que produzem características de rouquidão apresentam uma grande tendência para rápidas e frequentes alterações nos períodos de oscilação das pregas vocais (frequência fundamental), além dos fatores acima citados.

Com isto, classificou-se os padrões patológicos conforme os padrões visuais do espectrograma:

- a) Tipo 1 - Componentes harmônicas regularmente espaçadas com baixo nível de ruído, principalmente na região das formantes
- b) Tipo 2 - Predominância de ruído na região da segunda formante (nas vogais /e/ e /i/, inglesas), com aparecimento de componentes ruidosas a partir de 3000 Hz.

- c) Tipo 3 - Substituição das componentes harmônicas na região do segundo formante (fonemas /ε/ e /i/) por ruído, com elevação do ruído de alta frequência.
- d) Tipo 4 - O segundo formante dos fonemas /a/, /e/ e /i/ é substituído por ruído, com perda de periodicidade por incremento de componentes ruidosas (alterações na primeira formante). Aumento de ruído de alta frequência.

Também relacionando os padrões visuais do espectrograma com as patologias, Rontal *et al* (1975) verificaram que a soprosidade está associada à presença de ruído de alta frequência (sem especificar a partir de qual frequência tal consideração é válida), além de reafirmarem as evidências de Yanagihara (1967).

Também constatou-se, naturalmente, a existência de diferenças nas componentes harmônicas entre homens e mulheres. Por possuir uma frequência fundamental média superior à dos homens, a voz das mulheres apresenta maior proximidade entre as estrias espectrais (que caracterizam o *pitch*). Posteriormente, verificou-se que a discriminação de vozes patológicas de percepção áspera será dificultada em mulheres pois a condição de alto estiramento das pregas vocais (elevado *pitch*) é semelhante ao padrão normal quando a extensão da disfonia é pequena.

Verificou-se que até mesmo a presença de muco nas pregas vocais (causado possivelmente por fadiga vocal) pode influenciar o espectrograma, causando uma a repentina aperiodicidade das pregas vocais e retorno natural à condição normal.

O espectrograma é sem dúvida uma ferramenta extremamente útil para diagnóstico de disfonias e avaliação de tratamento do paciente. Entretanto, devido a grande quantidade de informação visual, o mesmo exige uma percepção do especialista na identificação dos padrões gráficos e sua relação com as patologias (e extensão desta). Por isto, não é possível o emprego somente do espectrograma de forma automática para avaliação das estruturas da laringe.

Além disto, algumas patologias (como nódulos, cistos e pólipos, por exemplo) não são discrimináveis por esta ferramenta, novamente pela grande quantidade de informação existente. Medidas acústicas baseadas nas evidências espectrais (periodicidade e quantidade de ruído) são importantes para discriminação das patologias.

2.5 Medidas de Perturbação

As perturbações, ciclo-a-ciclo, da frequência fundamental e de sua amplitude são importantes na avaliação de patologias do trato vocal. Isto inclusive é comprovado pela análise de quadros de filmes de alta velocidade.

Alterações extremas e abruptas na frequência fundamental durante a fonação fazem com que estas vozes sejam percebidas como ásperas. O acréscimo de ruído de alta frequência, proveniente da incorreta adução das pregas vocais caracteriza a voz rouca. Desta forma, pode-se perceber a importância da avaliação da frequência fundamental (ou *pitch*) na análise das disfonias.

A avaliação das perturbações ciclo-a-ciclo de frequência fundamental (*jitter*) e de amplitude (*shimmer*) necessita que seja estimado o período de abertura e fechamento das pregas vocais, e o nível com que a glote abduz-se e aduz-se. Nas primeiras pesquisas, o emprego de papéis milimetrados e medidores de tensão do sinal de voz gravado em meio magnética permitiam o registro da fonação e a avaliação das perturbações. Com o incremento da tecnologia digital, computadores possibilitaram o desenvolvimento de técnicas automáticas para determinação dos períodos sucessivos do sinal glotal.

2.5.1 Métodos de aquisição do sinal de voz

O sinal ideal para análise desta perturbação é o pulso glotal, que possui um conjunto de harmônicas, múltiplas da frequência fundamental, com um decaimento de 6 dB por oitava, que corresponde diretamente ao fluxo de ar produzido pela movimentação cíclica das pregas vocais. Entretanto, o nível de energia deste sinal é muito baixo, além de ser extremamente difícil de se obtê-lo diretamente. Cabe então desenvolver métodos para estimá-lo, ou determinar suas propriedades temporais (*jitter* e *shimmer*).

As técnicas de estimação do *pitch* dependem do modo com que a voz é obtida do paciente. O emprego de microfones tipo cardióide é muito difundido pela facilidade com que é operado e adquirido, sendo especificado com ganho dinâmico, máxima resposta em frequência e unidirecional, de modo que o sinal obtido seja a representação mais fiel da voz produzida pela pessoa. Este sinal contém informações sobre a laringe e cordas vocais, mas também traz as características das estruturas de ressonância.

Microfones tipo acelerômetro, ou de contato, por sua vez, produzem um sinal de voz livre da influência da cavidade supraglotal. A colocação deste na pele do pescoço do paciente, na região da tireóide, gera um sinal de voz com características próximas ao sinal glotal. Entretanto, a atenuação produzidos pelas paredes do pescoço pode mascarar disfonias ainda em um estágio inicial.

Um sinal mais próximo ao glotal é obtido pela eletroglotografia (Rothenberg e Mahshie, 1988; Childers *et al.*, 1990), onde um transdutor é colocado ao redor do pescoço do paciente e alterações na sua resistência, compatíveis com movimentação das pregas vocais. Entretanto, ainda questiona-se a relação entre o sinal glotal e o obtido por esta técnica: variação inter-pessoas da anatomia do pescoço e problema com patologias onde não ocorre o contato entre as cordas vocais podem levar a erros de estimativa (Askenfelt, 1980).

2.5.2 Estimativa do período e amplitude da frequência fundamental

Uma vez obtido o sinal a ser utilizado na avaliação das perturbações, determina-se o valor da amplitude e período de *pitch* ao longo do sinal estimado. Para os sinais advindos de acelerômetros e eletroglotógrafo, o emprego de técnicas de busca de pico (*peak-picking*) ou pesquisa de cruzamentos por zero (*zero-crossing*) são comumente empregadas devido a clara definição dos períodos e amplitudes nestes sinais.

Figura 8 - Estimativas de período e amplitude ciclo-a-ciclo no fonema /a/

O sinal advindo de microfones convencionais (tipo cardióide, por exemplo) trazem consigo as características espectrais do trato supraglotal e desta forma, a determinação dos pontos de busca (valor máximo ou mínimo do sinal de voz, ponto de cruzamento por zero) é dificultada.

Na fig. 8 são mostrados simultaneamente os instantes onde são comumente determinadas as medidas de período da fundamental (instantes referenciados graficamente por círculos). Para este fonema, a determinação visual é simples, entretanto, vozes patológicas podem não apresentar esta clareza, exigindo algoritmos computacionais específicos para determinar a periodicidade do sinal de voz.

Classicamente, existem dois grupos de métodos para determinação do período de *pitch*: baseados em eventos e os baseados em médias de tempo curto (*short-term*). Os métodos baseados em eventos buscam pontos chaves no sinal de voz e a diferença entre estes caracteriza o *i*-ésimo período de frequência fundamental. Neste grupo encontram-se as medidas por busca de pico, cruzamento por zero e mais recentemente, formas de onda similares (*waveform matching*), onde esta última é baseada em uma minimização do erro médio quadrático entre sucessivos períodos de *pitch* (Milenkovic, 1987). Estas medidas tornam-se versáteis quando é aplicado um filtro passa-baixa no sinal de voz. Titze e Liang (1993) recomendam que o filtro seja do tipo FIR (fase linear) com frequência de corte entre o primeiro e o segundo harmônico (supõem-se que seja com base em uma estimativa prévia de F_0).

Dentro do grupo de estimadores baseados em médias de tempo curto, encontram-se métodos baseados em autocorrelação (Dubowski, 1976), por diferença média de magnitude, ou AMDF (Ross, 1974; Moorer, 1974), por filtragem inversa (Sondhi, 1968; Maksym, 1973; Davis, 1979), por análise cepstral (Noll, 1964; 1966), por compressão harmônica (Schroeder, 1968), além de outros. Apesar da precisão destas técnicas, diretamente não devem ser usadas na avaliação das perturbações de frequência fundamental por realizarem uma ponderação média do valor de F_0 , e assim desconsiderar algumas oscilações do valor médio que correspondem à alterações na laringe e pregas vocais. O emprego de métodos híbridos pode ser considerado útil na estimação tanto da amplitude quanto do período da frequência fundamental.

Com relação às medidas de amplitude ciclo-a-ciclo, não há uma padronização sobre quais pontos do sinal de voz devem ser medidos e sua diferença considerada como *shimmer*. Os valores comumente medidos como amplitude de um ciclo são o ponto máximo e mínimo do sinal, apesar dos estudos ainda não indicarem a relação direta entre estes picos e o sinal glotal propriamente dito. Verifica-se que o pico mínimo em um período do sinal de voz é posterior ao fechamento real da prega vocal (energia sonora mínima), mesmo sem a aplicação de filtros passa-baixa (Titze, Horii e Scherer, 1987). A influência da cavidade de ressonância no sinal de voz é responsável por este distanciamento entre os sinais (glotal e de voz). Na fig. 8 pode-se perceber que as amplitudes mínimas são determinadas para avaliação posterior do *shimmer*.

Uma vez que o processamento do sinal de voz é realizado em computadores, as peculiaridades do processo de discretização devem ser consideradas na análise das perturbações de F_0 .

As medidas de *jitter* são limitadas pela taxa de amostragem enquanto que o *shimmer* tem precisão relacionada à discretização das amostras. Estudos consideraram este problema. Verificou-se

que a quantização das amostras deve ser, sem o emprego de interpolação, no mínimo de 9 bits para garantir uma margem de erro de 0,1% em vozes normais na avaliação do *shimmer* (Titze, Horii e Scherer, 1987).

Horii (1979) constatou que a partir de 20 kHz de taxa de amostragem, obtém-se valores reduzidos de *jitter*, definindo assim esta frequência como mínima para estudos de perturbação de *pitch*, uma vez que a resolução desta medida é de 0,05 ms. Posteriormente, Titze *et al* (1987), consideraram que 500 amostras por período de frequência fundamental são necessárias para ter-se um erro de estimativa do *jitter* em 0,1%. Este requisito é elevado demais pois para uma voz masculina, por exemplo, de 100 Hz exigiria uma taxa de amostragem de 50 kHz.

O uso de interpolação nas medidas de amplitude e frequência fundamental permite a redução destes requisitos de taxa de amostragem e discretização. Realizando interpolação parabólica pode-se reduzir o requisito de taxa de amostragem para 50 amostras por ciclo, mantendo o nível do erro no cálculo do *jitter*.

2.5.3 *Jitter e shimmer*

Com a determinação precisa das amplitudes e períodos de frequência fundamental, é possível efetuar as medições acústicas para identificação de patologias ou característica perceptual. Os índices acústicos encontrados na literatura apresentam diferenças entre o tipo de microfone empregado, armazenamento da voz gravada e, principalmente, tipo de fala produzida pelo paciente (contínua ou sustentada).

Estes parâmetros baseiam-se em uma estabilidade ideal das cordas vocais, obtendo-se então uma variância nula. Alterações involuntárias ou proveniente de patologias produzirão padrões errôneos de vibração nas pregas vocais que serão detectados por medidas baseadas na alteração ciclo-a-ciclo da amplitude ou frequência fundamental. Pode-se definir *jitter* e *shimmer* como uma medida de dispersão do período e da amplitude ciclo-a-ciclo, respectivamente.

Considerando a magnitude da dispersão das perturbações ciclo-a-ciclo, Horii (1979) obteve o *jitter* médio, sendo calculado pela seguinte expressão:

$$JM = \frac{\sum_{i=1}^N |P_i - P_{i-1}|}{N - 1} \quad (18)$$

onde P_i corresponde ao período de *pitch*.

O emprego do módulo desta medida pode ser substituído pelo desvio padrão das diferenças entre períodos sucessivos, mantendo-se o sinal, uma vez que estas medidas (*jitter* sinalizado) possuem distribuição normal (teste χ^2).

Lieberman (1961, 1963) verificou através da distribuição das diferenças entre períodos sucessivos que os casos patológicos apresentavam-se na região desta curva, acima de 0,5 ms. Com

isto, definiu a medida PFF (*Perturbation Frequency Factor*), ou seja, a integral da distribuição da frequência a partir de períodos sucessivos superiores à 0,5 ms. Inclusive comentou que esta medida é sensível ao tamanho e posição de crescimentos (pólipos, cistos, nódulos) na laringe do paciente. Mas Hall (1995), empregando medidas de *jitter* e *shimmer* em sinais de voz e de eletroglotografia verificou que não foi possível discriminar significativamente nódulos vocais em mulheres.

Considerando apenas as alterações de sinal na frequência fundamental, Hecker e Kruel (1971) desenvolveram a medida DPF (*Directional Perturbation Factor*), sendo definido pelo percentual de alterações de sinal da frequência fundamental. Entretanto, estes pesquisadores empregaram fala contínua (*Rainbow Passage*) e assim, seus resultados devem ser ponderados com cuidado pois a prosódia e características psicológicas influenciam a F_0 . Posteriormente, Sorensen e Horii (1984) empregaram fala sustentada para avaliação desta medida. Juntamente com Murry e Doherty (1980), verificaram que esta medida (para perturbações tanto em amplitude como em frequência fundamental, mesmo que os primeiros autores não a tenham usado para amplitude), é estatisticamente significativa na discriminação de carcinomas na laringe. Também verificou-se que não há diferença significativa entre homens e mulheres quando o DPF é empregado. Sorensen e Horii tentaram inclusive explicar os valores elevados para esta medida - característica patológica - em termos de alterações nos mecano-receptores do sistema reflexogênico que mantém o equilíbrio de forças na musculatura das pregas vocais, apesar desta explicação ser apenas uma suposição teórica.

Reconhecendo a dependência da perturbação de períodos de *pitch* em relação à frequência fundamental, Horii (1979) empregou a medida JR (*Jitter Ratio*) para evitar que com o aumento de F_0 , ocorresse uma redução do *jitter* médio. Isto advém da estabilidade “artificial” obtida nesta circunstância pois o aumento de tensão na prega vocal aumenta a adução da mesma, restringindo o espaço glotal e permitindo que os níveis volumétricos de ar sejam menores. Esta medida é definida pela média da diferença absoluta das perturbações dos período de F_0 pelo período médio no trecho analisado. Multiplica-se este valor por 1000 apenas para evitar valores muito pequenos. Horii (1982) verificou estatisticamente que não existem diferenças entre medidas usando este índice obtidas para oito diferentes fonemas vocálicos ingleses. Sorensen e Horii (1983) encontraram diferenças significativas entre homens e mulheres para esta medida.

Seguindo uma abordagem semelhante, Murry e Doherty (1980) usaram a media JF (*Jitter Factor*) que é definida por:

$$JF = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |F_i - F_{i+1}|}{\frac{1}{N} \sum_{i=1}^N F_i} \quad (19)$$

onde F_i denota o recíproco do período fundamental.

Verificou-se que esta medida pode ser empregada na diferenciação entre pessoas normais e pacientes disfônicos, discriminando inclusive pacientes com tumor de pacientes com paralisia unilateral. Entretanto, verificou-se que DPF para perturbações em F_0 é mais sensível às alterações de massa na laringe.

Ainda considerando a relação entre frequência fundamental e medida de perturbação, Honjo e Isshiki (1980) consideraram o uso de medidas de frequência em semi-tons. Com isto, determina-se a média da variação em semi-tons, da frequência fundamental à cada período. Constatou-se que valores acima de um semi-ton caracterizam-se por elevada disфонia.

Wendhal (1963, 1966) estudou os desvios ciclo-a-ciclo de período e amplitude através de síntese de voz. Produzindo alterações de *jitter* e *shimmer* no sinal sintetizado proporcionais, este pesquisador determinou que estas medidas relacionam-se diretamente com patologias de característica áspera, pois o aumento de *jitter* e *shimmer* é proveniente do aumento da percepção da aspereza na voz do paciente.

Considerando a avaliação da aspereza, Deal e Emanuel (1978) desenvolveram um índice acústico que se relaciona-se com o aumento de ruído espectral característico deste padrão de voz. Para evitar a polarização por diferenças naturais na frequência modal de diferentes pessoas, empregou-se o coeficiente de variância sobre os períodos e amplitudes de frequência fundamental. Obteve-se então as medidas PVI (*Period variability index*) e AVI (*Amplitude variability index*), que correspondem ao coeficiente de variância multiplicado por uma constante (1000). Aplica-se o logaritmo decimal sobre a medida AVI por se relacionar linearmente com a percepção de aspereza. Apesar de bons resultados, estes pesquisadores, que empregaram equipamentos analógicos, não as consideraram eficientes na identificação de aspereza na voz de pacientes (a medida de ruído SNL por eles descrita oferece melhor poder de discriminação, sendo posteriormente apresentada).

Entretanto, Klingholz e Martin (1985) acharam que medidas de *jitter* (mais especificamente o *Jitter Factor*) são capazes da diferenciação entre distúrbio de vozes hipofuncional e hiperfuncional (a hiperfuncionalidade é uma característica das vozes ásperas) e questionaram os impulsos dente-de-serra usados por Wendhal como sinal sintetizado na avaliação da aspereza. Estes pesquisadores, através de voz sintetizada, mostraram que o nível de ruído espectral proporcionado por perturbações de frequência fundamental é muito superior que os produzidos por alterações de amplitude.

Cabe ressaltar que Gelfer (1995) cita que com o aumento da frequência fundamental, ocorre um mascaramento da condição áspera da voz, pois a hiperfuncionalidade da prega vocal cria uma condição de estabilidade “artificial” desta. Mulheres geralmente apresentam maior aspereza do que homens, apesar de diagnosticadas (através de equipamentos) como normais.

Apesar da possibilidade de uso na avaliação de tremor vocal (Winholtz e Ramig, 1992), as alterações lentas nos períodos de *pitch* em fala sustentada não estão relacionadas com patologias e podem elevar as medidas de *jitter*, o mesmo sendo válido para o *shimmer*. Koike (1973) desenvolveu a medida RAP (*Relative average perturbation*), também denominada PPQ (*Period perturbation*).

quotient) por Takahashi e Koike (1975) para evitar que tais efeitos acústicos indiquem falsamente uma disfonia.

Esta medida também foi aplicada sobre medidas de amplitude, ciclo-a-ciclo, sob o nome de APQ (*Amplitude perturbation quotient*). Davis (1979) e Koike *et al* (1977) buscaram estabelecer a fundamentação matemática destas medidas - RAP, PPQ e APQ - para a discriminação entre patologias e grupo de controle normal.

No cálculo destes parâmetros, determina-se a diferença entre o período/amplitude atual da frequência fundamental e uma estimativa empregando média móvel deste valor. Assim, tem-se a seguinte expressão:

$$RAP = \frac{\frac{1}{N-P+1} \sum_{i=1}^{N-P+1} \left| \frac{1}{P} \sum_{j=1}^P (X_{i+j-1}) - X_{i+0.5(P-1)} \right|}{\frac{1}{N} \sum_{i=1}^N X_i} \quad (20)$$

onde N é a quantidade de períodos (ou amplitudes) usados no cálculo

P é o tamanho da janela para média móvel

X_i é a i -ésima medida de período (ou amplitude), ciclo-a-ciclo

O tamanho da janela variou conforme o pesquisador: Koike (1973) empregou uma janela de 3 amostras para medidas de oscilação de F_0 e 11 amostras para medidas de amplitude, enquanto Davis (1979), através de teste sistemático do tamanho das janelas sobre um sinal proveniente de filtragem inversa, sugere o uso de $k=5$ para medidas de perturbação em amplitude.

Patologias como laringite crônica, paralisia unilateral e tumor foram discriminadas por estes parâmetros, enquanto seu uso em vozes de pacientes com nódulos não foram significativamente diferentes as vozes normais.

Seguindo um caminho distinto, Schoentgen e Guchteneere (1995) analisaram o *jitter* através de séries temporais. Sua fundamentação está na existência de correlação entre os sucessivos períodos de *pitch*. Considerando que as medidas de dispersão supõem independência estatística entre amostras, é necessário aplicar um método de descorrelação das medidas de variação de ciclo. Isto é obtido pelo uso de séries temporais por dispensarem a hipótese de independência estatística. A estimação de um modelo AR possibilita a determinação de resíduos (erros de estimação) e determina-se o PPQ modificado, ou seja:

$$PPQ_{\text{mod}} = \frac{\frac{1}{N-q} \sum_{i=q+1}^N |Z_i|}{\frac{1}{N} \sum_{i=1}^N P_i} \quad (21)$$

onde Z_i é o i -ésimo resíduo (ou erro) da estimação do modelo AR

q é a ordem do modelo AR

N é a quantidade de períodos de frequência fundamental estimados

P_i é o i -ésimo período de *pitch*.

A ordem do modelo é determinada a posteriori baseado em teste estatístico χ^2 para verificação da brancura do sinal residual, uma vez que o algoritmo de Levinson permite o incremento da ordem do modelo a cada iteração. A aceitação da hipótese no teste motiva a parada da recursão nesse algoritmo.

Esta abordagem permite a redução da diferença entre pessoas, pois elimina a correlação do *jitter*. Verificou-se que a curva de distribuição dos resíduos da estimação é mais estreita (leptocúrtica) que a distribuição do *jitter*, uma vez que o intervalo total é reduzido. Esta distribuição sofre uma redução do seu centro (deslocamento da média). Verificou-se que a distância entre as distribuições dos resíduos entre pessoas normais e pacientes (utilizada na verificação de separação entre populações) foi reduzida, provocada pela redução no intervalo interquartil.

Através do coeficiente de correlação serial (r_k) sobre as medidas de amplitude da frequência fundamental, Koike (1969) e Iwata (1972) foram capazes de determinar padrões no correlograma para o padrão normal de voz e patologias como paralisia unilateral e carcinoma. Empregando acelerômetro e “lag” variando de 1 até 15, demonstra que as medidas de amplitude são úteis na discriminação de patologias.

De modo a padronizar a maioria das medidas de perturbação (*jitter* e *shimmer*) em torno dos conceitos que as fundamentam, Pinto e Titze (1990) tentaram unificar os parâmetros acústicos em torno de derivadas de ordem superior das medidas de perturbação. Foi sugerido o uso de medidas estatísticas como média, mediana e desvio padrão para avaliar a extensão da perturbação. Isto é confirmado por Askenfelt e Hammarbert (1986) que consideraram o desvio padrão das derivadas discretas de primeira ordem como melhor índice para aplicação no diagnóstico de patologias (eles empregaram fala contínua em seu trabalho).

Foi avaliado a robustez desta medidas em rejeitarem modulações (senoidal, ruído gaussiano e tendência linear). Esta unificação teve o intuito de criar uma terminologia comum entre os pesquisadores e facilitar a comparação entre resultados.

A maioria das medidas acústicas para determinação de *jitter* e *shimmer* foram descritas nesta seção. Variações deste conjunto são possíveis e comumente encontrados, como por exemplo, emprego de função logarítmica e eliminação da tendência linear. Entretanto, alguns esclarecimentos ainda se fazem necessários.

Na determinação dos parâmetros de perturbação, é necessário considerar a quantidade de ciclos que serão analisados para obtenção de uma medida confiável. Titze *et al* (1987) recomendam o emprego de no mínimo 20 ciclos para determinação dos índices acústicos. Verificou-se que as medidas de perturbação decaem assintoticamente para um valor estável a partir desta quantidade de

ciclos. Deve-se notar que a maioria das medidas acústicas empregam a média estatística; o emprego de estimativas robustas como a mediana possivelmente permite a redução desta quantidade.

A faixa de idade, tanto das amostras normais quanto dos pacientes tem influência sobre as medidas acústicas. Murry e Doherty (1980) verificaram que DPF e JF são mais elevados para pessoas de maior idade; isto provavelmente está associado ao envelhecimento das pregas vocais e órgãos anexos envolvidos na produção da fala. É importante ser cuidadoso na comparação com resultados da literatura e valores dos índices produzidos para pessoas idosas.

Um ponto de discordância entre pesquisadores é o tipo de voz a ser empregado na análise das perturbações (sendo válido para demais tipos de medidas). Alguns trabalhos empregam a fala contínua considerando-a como a mais natural que a pessoa pode produzir, reduzindo os efeitos psicológicos da metodologia para gravação.

Entretanto, Horii (1979) afirma que as medidas obtidas para este tipo de fala devem ser cuidadosamente consideradas pois alterações articulatórias do trato supraglotal influenciam os parâmetros acústicos, mascarando patologias e evidenciando disfonias inexistentes. O emprego de voz sustentada permite a fixação do trato supraglotal em uma configuração quase fixa, e “somente” alterações na laringe e pregas vocais influenciaram o sinal de voz. Através do emprego de eletroglotografia ou acelerômetros, estas considerações não se fazem necessárias.

As vozes sustentadas são caracterizadas pela produção de fonemas vocálicos. A escolha destes também varia conforme o pesquisador. Os fonemas mais empregados na literatura são o /a/, /e/ e /i/, sendo esta escolha motivada pelas características da estrutura formante, menor amplificação das componentes não harmônicas (ruído) pela cavidade supraglotal, facilidade na determinação dos períodos e amplitudes de frequência fundamental.

Cabe citar que o fonema /i/ tem características mais ásperas que os outros dois fonemas. Isto advém da característica de maior tensão das pregas vocais, causada pela elevação das aritenóides. A estabilidade do sinal produzido é superior, mas devido a característica ressonadora da configuração do trato supraglotal, uma quantidade de ruído proveniente da glote é amplificada. Com características opostas, o fonema /a/ é menos estável, mas o sinal de voz possui menor quantidade de ruído.

2.6 Medidas de ruído espectral

O ruído espectral corresponde, geralmente, ao aumento da produção de ar turbulento devido à incompleta adução das pregas vocais. A percepção acústica destas alterações no sinal de voz é a soproidade.

Patologias que produzem percepção de rouquidão também permitem que grande quantidade de ar atravesse a glote durante a fonação. Assim é importante determinar a quantidade de ruído existente no sinal de voz. Além disto, alterações na estrutura espectral, como deterioração energética das componentes harmônicas e aumento da largura de banda nas frequência formantes devem ser contabilizados para uma melhor discriminação das patologias.

A presente seção apresenta as medidas acústicas que consideram estas modificações no espectro, buscando representar o comportamento incorreto das pregas vocais, seja por disfonias funcionais ou orgânicas.

Figura 9 - Espectro densidade de potência para fonema /a/ em uma voz normal

Na fig. 9, onde é apresentado um fonema /a/ pronunciado por uma pessoa normal, pode-se claramente identificar as componentes harmônicas até 3800 Hz, aproximadamente. A partir desta frequência, não existe uma separação clara entre as componentes harmônicas e as ruidosas. Percebe-se a diferença energética entre as amplitudes de baixa e alta frequência e pode-se inclusive identificar os formantes apenas por inspeção visual desta curva espectral.

Figura 10 - Espectro densidade de potência para fonema /a/ em uma voz áspera

Já para uma voz identificada perceptualmente por áspera, pode-se verificar a redução na amplitude das componentes harmônicas, além da elevação (apesar de discreta) das componentes de alta frequência. Nota-se que a estrutura formante não é totalmente identificada (principalmente a partir da terceira).

Figura 11 - Espectro densidade de potência para fonema /a/ em uma voz soprosa

A análise visual do espectro densidade de potências para uma voz soprosa (um pólip, no caso), percebe-se que as componentes harmônicas são praticamente mascaradas em sua totalidade,

restringindo-se à região abaixo de 1000 Hz. A estrutura formante é praticamente eliminada e a partir de 3000 Hz, pode-se verificar a característica de nivelamento espectral (*spectral flatness*) devido à elevação da quantidade de ruído promovido pelo escape de ar nas pregas vocais.

Figura 12 - Espectro densidade de potência para fonema /a/ em uma voz rouca

Claramente na fig. 12 pode-se perceber a elevação da quantidade de energia de ruído de alta frequência, reforçando o nivelamento espectral. As componentes harmônicas situadas abaixo de 1000 Hz ainda permitem a identificação visual da frequência fundamental, mas a estrutura formante é praticamente inexistente.

Baseado nestas considerações gráficas são realizadas as análises de relação sinal ruído que possibilitam a discriminação de um rol elevado de tipos de patologias. Perceber-se-á que muitas medidas acústicas são definidas em termos do espectro densidade de potência, motivando a apresentação das figs. 9, 10, 11 e 12.

Empregando um analisador de frequência de alta resolução (equivalente à densidade espectral de potência), Remacle e Trigaux (1991) verificaram que no intervalo entre 0 e 1 kHz é possível perceber visualmente diferenças entre uma voz normal e voz patológica provenientes de cistos e nódulos na corda vocal, com o emprego de uma escala linear ao invés de logarítmica. Descreveram o nódulo como possuindo o primeiro harmônico inalterado, mas os demais sofrem decréscimo de intensidade. Identificaram que casos de nódulos de pequena extensão eram muito similares à voz normal. Apesar do carácter qualitativo, pode-se perceber que o espectro permite identificar características particulares das disfonias.

Utilizando-se da densidade espectral de potência médio em composições fonéticas consoante-vogal, Naranjo *et al* (1994) à uma taxa de amostragem de 48 kHz, verificaram que a energia acima de 6 kHz a energia é significativamente maior para pessoas normais do que para pacientes

disfônicos. Apesar de empregar sílabas, somente a porção vocálica é analisada, pois a coarticulação representa uma não estacionariedade no trato supraglotal.

Shoji *et al* (1992-1), considerando pacientes com voz soprosa e usando frequência de amostragem em 44,1 kHz, determinou também a espectro de potência média para 20 segmentos de 243,8 ms, com sobreposição de 50%. Determinando-se a energia a partir de uma frequência de corte, percebeu-se que a partir de 6 kHz os pacientes com soprosidade são discriminados adequadamente.

Outra forma de visualizar os ruídos é através do espectro médio de longo tempo (LTAS ou *Long-term average spectra*). Frokjaer-Jensen e Prytz (1976) que corresponde a determinação da densidade espectral de potência média por um grande número de amostras.

Estas medidas da densidade espectral correspondem à retirada da ordenada tempo do espectrograma. Uma simplificação que pode ser empregada em voz sustentada pois admite-se novamente a estacionariedade do sinal de voz. Assim, as discussões sobre resolução temporal e espectral permanecem válidas para estas medidas, qualitativas ou quantitativas.

Uma das medidas mais encontradas nos estudos de disfonias é o HNR (*Harmonic-to-noise ratio*), desenvolvido por Yumoto, Gould e Baer (1982). Originalmente, são consideradas as vozes roucas. O método baseia-se na suposição de que o sinal de voz é composto de uma componente periódica e outra, ruidosa (distribuição de amplitudes com média zero).

Determina-se uma estimativa de um período de frequência fundamental sem ruído, $s_a(t)$, através da média de um conjunto N de trechos de comprimento igual ao i -ésimo período de *pitch*. Assim, esta medida pode ser obtida pela seguinte relação

$$HNR = \frac{N \cdot \sum_{t=1}^{T_a} s_a^2(t)}{\sum_{i=1}^N \sum_{t=1}^{T_i} [s_i(t) - s_a(t)]^2} \quad (22)$$

onde T_a é o comprimento da estimativa livre de ruído

T_i é o comprimento do i -ésimo trecho do sinal de voz

s_i é o i -ésimo trecho

O denominador corresponde à energia do ruído estimado. Foi considerado $N=50$ nesta estimação. Pode-se verificar a necessidade de uma estimativa precisa do período de frequência fundamental. Estes pesquisadores citam que esta técnica não pode ser empregada em casos extremos de rouquidão por não ser possível identificar os trechos periódicos, e assim, determinar HNR; porém verificaram que esta medida é estatisticamente significativa na avaliação de rouquidão.

Para contornar o problema da determinação dos períodos de F_0 na estimação de HNR, Awan e Frenkel (1994) desenvolveram um algoritmo empregando medidas de perturbação (*jitter* e *shimmer*), além de busca de picos e cruzamentos de zero, possibilitando uma redução na quantidade de erros na determinação dos trechos utilizados nessa medida acústica. Isto permite uma redução do desvio

padrão de HNR para os grupos de patologia, aglutinando-os espacialmente, o que possibilita uma discriminação patológica mais robusta.

Outra medida para identificação de níveis de rouquidão é o Hr (*Relative harmonic intensity*), baseando-se na existência de proeminentes componentes de frequência fundamental no espectro de potência, usando escala linear, enquanto que as regiões laterais contém ruído proveniente da turbulência do ar na glote. Assim, determinando esta curva sobre 4096 pontos no trecho mais estável do sinal de voz. Hiraoka *et al* (1984) definiram então esta medida como a percentagem de energia das componentes harmônicas a partir da segunda frequência harmônica, pela energia total do sinal de voz. Não foi utilizado a primeira harmônica pois verificaram que algumas vozes roucas apresentam elevada amplitude nesta componente, fazendo com que a medida acústica possua um grande valor e a discriminação patológica seja prejudicada; sem seu emprego, Hr apresenta valor limite para pessoas normais em 67,5%, enquanto que se esta componente fosse utilizada, este limiar sobe para 76,5%, para o fonema /a/ (que apresenta alta energia na formante mais baixa).

Empregando expansão de séries de Fourier, Kojima *et al* (1980) construiu um parâmetro acústico relacionando a energia entre componentes acústicas e ruidosas de um sinal construído a partir de três trechos com período de frequência fundamental.

Determina-se então a energia das componentes harmônicas (S_i) e não-harmônicas (N_i) da densidade espectral obtida do sinal gerado para cada segmento de três períodos de *pitch* até a frequência de 5500 Hz, percorrendo 325 ms do sinal de voz. Finalmente determina-se a medida R_a (*Relative harmonic*) através da expressão

$$R_a = 10 \cdot \log \frac{\sum_{i=1}^N H_i}{\sum_{i=1}^N N_i} \quad (23)$$

Esta técnica tenta ampliar a resolução da transformada de Fourier por utilizar um sinal altamente periódico, sendo capaz de diferenciar vozes soprosas de ásperas.

A medida da soprosidade desenvolvida por Fukazawa *et al* (1988) permite uma avaliação simples do ruído de turbulência produzido pela glote nesta condição disfônica. A medida Br (*Breathiness ratio*) corresponde ao percentual de energia de alta frequência por energia total do sinal. Esta energia de alta frequência é obtida pela derivada discreta de segunda ordem (filtro passa alta com +12 dB/oitava) do sinal de voz, ou seja:

$$Br = \frac{\sum_{i=1}^{N-2} [s(t_i) - 2s(t_{i-1}) + s(t_{i-2})]^2}{\sum_{i=1}^N [s(t_i)]^2} \quad (24)$$

A vantagem desta medida é não depender da determinação de período de frequência fundamental, uma vez que algumas patologias dificultam esta estimação. Algumas medidas que foram apresentadas (HNR, por exemplo) dependem destas avaliações para permitirem a identificação de patologias. Além da extrema simplicidade de obtenção, possibilita discriminar entre disfônicos soprosos e pessoas normais. Os autores sugerem o uso adicional de outras medidas para melhorar a separabilidade entre câncer, pólipos e paralisia unilateral.

Investigando vozes ásperas, alguns pesquisadores (Sansone e Emanuel, 1970; Lively e Emanuel, 1970; Emanuel *et al*, 1973; Emanuel e Whitehead, 1979; Trullinger *et al*, 1988) desenvolveram uma medida denominada SNL (*Spectral noise level*) baseada na determinação das menores intensidades espectrais em intervalos de 100 Hz na curva média de 37 espectros de potência, com sobreposição de 256 pontos. Os intervalos estudados são 100 a 2600 Hz, 2600 a 5100 Hz, 100 a 5100 Hz, 100 a 8000 Hz e 100 a 2000 Hz.

Determinaram que existem vogais que apresentam mais características de aspereza que outras (/u/ e /i/), possivelmente pela característica das formantes estarem deslocadas na direção das altas frequências, pelo posicionamento da língua, e maior tensão das pregas vocais. Além disto, verificaram que o ruído espectral está relacionado com perturbações na frequência fundamental. Isto também é explicado pelos estudos de simulação de aspereza de Klingholz e Martin (1985). Verificaram que existe uma alta correlação entre esta medida e a aspereza quando empregado os intervalos 100 a 2600 Hz e 100 a 2000 Hz. Isto porque as componentes harmônicas são confundidas pelas amplitudes de ruído nas altas frequências.

Empregando um método adaptativo de filtragem, mais robusto que HNR e R_a , Kasuya *et al* (1986) construíram uma medida no domínio da frequência para estimar o ruído entre as componentes harmônicas, que suporta alterações suaves na amplitude e período de F_0 . Considerando que o sinal de voz é composto por componentes periódicas e ruído aditivo, a medida NNE (*Normalized noise energy*) é quantificada pela expressão:

$$NNE = 10 \cdot \log \frac{\frac{1}{L} \sum_{\omega=N_L}^{N_H} \sum_{m=1}^L |\hat{N}_m(\omega)|^2}{\frac{1}{L} \sum_{\omega=N_L}^{N_H} \sum_{m=1}^L |S_m(\omega)|^2} \quad (25)$$

onde $\hat{N}_m(\omega)$ é a estimativa espectral do ruído;

$S_m(\omega)$ é a densidade espectral de potência do sinal de voz;

L é a quantidade de janelas de dados que são utilizadas para análise;

N_L e N_H são os limites inferiores e superiores, em amostras, sobre o qual a medida é desenvolvida, sendo $N_L = N f_L T$ e $N_H = N f_H T$. As frequências f_L e f_H formam o intervalo de frequências na curva de densidade espectral de potência onde a estimativa de ruído é feita. T é o

recíproco da frequência de amostragem. N é o número de pontos da transformada de Fourier utilizada no cálculo da curva de energia espectral.

A estimativa do ruído é feita com a seguinte suposição: na região das componentes não harmônicas existe apenas ruído; assim, nestes locais, a energia do ruído é igual a energia das próprias componentes. Já as componentes periódicas são compostas por sinal e ruído, utiliza-se o valor médio da energia das regiões não harmônicas laterais como estimativa do ruído. Assim, fica determinado o numerador da expressão, ou seja, ruído espectral.

As larguras de banda dos componentes harmônicos, utilizadas na estimação do ruído, são obtidas por processo adaptativo, baseado na frequência fundamental. São utilizados 7 períodos de *pitch* para a janela (Hamming) de dados de 40 ms de extensão, onde se determina o valor de NNE. Este processo é repetido a cada 20 ms e o valor médio de NNE representa a medida de quantidade de ruído para a voz analisada.

Patologias como câncer (do tipo T2, T3 e T4), paralisia unilateral e nódulos foram passível de identificação utilizando este parâmetro acústico para um intervalo de frequências entre 1 e 5 kHz, parecendo informar adequadamente a quantidade de ruído turbulento que atravessa a glote. Entretanto, as disfonias pólipos, Edema de Reinke e laringite não foram identificadas adequadamente.

O emprego de análise cepstral foi considerada por de Krom (1993) na determinação da relação harmônico-ruído. O princípio do método é considerar que os picos na *quefrequency* (ver fig. 13) equivalente a frequência fundamental correspondem às componentes *harmônicas* do sinal. Retirando-os por um processo de *liftering*, obtém-se então o ruído cepstral (componentes não harmônicas) modulado pela estrutura formante, pois estas características ainda permanecem - baixo *quefrequency* - no cepstrum modificado. Retornando ao domínio da frequência, a diferença de energia, em uma determinada banda de frequência, entre o sinal original e a componente ruidosa (ruído cepstral), fornece o valor de HNR.

Figura 13 - Exemplo de cepstrum real para fonema /a/

Um processo adaptativo é empregado para determinar os picos *harmônicos* baseado na alteração de sinal da primeira derivada do cepstrum à direita e à esquerda destes picos. Também é realizado uma compensação no valor da componente ruidosa (domínio da frequência) pois as

componentes harmônicas devem estar sobre o eixo das abscissas. Esta correção de linha de base é realizada, subtraindo-se do ruído espectral estimado, o módulo do valor mínimo da componente espectral de menor energia no intervalo entre picos harmônicos do espectro livre de ruído (transformada inversa do cepstrum modificado).

Verificou-se posteriormente que esta medida baseada em análise cepstral (Krom, 1995) consegue discriminar adequadamente vozes soprosas de ásperas, nos intervalos de frequência de 400 a 2000 Hz e 2000 a 5000 Hz.

Sabendo que a amplitude do primeiro harmônico é muito maior para vozes soprosas do que para vozes normais, Krom (1990) determinou a diferença entre a energia dos dois primeiros harmônicos. Considerando que vozes soprosas apresentam queda suave de energia no intervalo entre a segunda formante e a banda de ruído turbulento de alta frequência (2000 a 5000 Hz), foi considerado a diferença de energia entre os intervalos acima apresentados. Também foi considerado as bandas entre 60 a 400 Hz e 5000 a 8000 Hz, além de uma medida de frequência fundamental baseada em técnica cepstral de Noll (1966).

Estas medidas de relação sinal-ruído (HNR), energia e declividade espectral, energia de harmônicos e algumas relações entre elas, tem objetivo de descrever o espectro (grande quantidade de informação) através de poucos valores e relacioná-los com as patologias.

Verificou-se que a percepção de aspereza varia conforme a vogal, confirmando estudos anteriores (Sansone e Emanuel, 1970; Lively e Emanuel, 1970; Emanuel *et al*, 1973; Emanuel e Whitehead, 1979; Trullinger *et al*, 1988; Klingholz e Martin, 1985), tendo apresentado resultados de discriminação inferiores aos obtidos com patologias soprosas. Isto indica que é necessário empregar medidas de perturbação de amplitude e período de *pitch* para identificar tais patologias, apesar de se conhecer a influência do *jitter* e *shimmer* no domínio espectral (Klingholz e Martin, 1985).

Empregando um método de *pitch* síncrono para avaliação da rouquidão, Muta *et al* (1988) desenvolveram a medida NS (*Noise to signal*) para determinar a quantidade de ruído entre os picos harmônicos no espectro de potência. Primeiro, é extraído um esboço do período de frequência fundamental para um trecho de voz, através de um método de estimação de *pitch* qualquer; em seguida, através da autocorrelação de um novo trecho de voz de comprimento quatro vezes o período de F_0 determinado se obtém um novo período de frequência fundamental que é comparado ao seu esboço. Caso sejam diferentes, um novo tamanho de amostras é utilizado, até ambos serem idênticos.

Com a aplicação de uma janela de Hanning sobre as amostras de tamanho quatro vezes superior ao período de *pitch* encontrado, determina-se o espectro de potência destas amostras janeladas, do qual mantém-se somente 4 amplitudes, $S(k)$, por período de *pitch* (igualmente espaçadas).

Uma vez que se têm exatamente quatro períodos da fundamental de amostras da voz analisada, os picos harmônicos são separados exatamente à cada quatro unidades. Assim, determina-se a relação entre picos e vales (sinal/ruído) por:

$$NS = 10 \cdot \log \frac{\sum_{k=3}^{4L+2} \min_{i=-1,0,1,2} S(4k+i)}{\sum_{k=3}^{4L+2} S(k)} \quad (26)$$

onde L é o número de harmônicos empregados no cálculo (no caso, $L=16$).

Primeiramente, verificou-se a capacidade da medida na avaliação terapêutica de pacientes (pré- e pós-operatórios), mostrando ser útil nesta função. Resultados mostraram que o método é sensível à *jitter*, *shimmer* e ruído aditivo, e insensível à pequenas modulações em amplitude e frequência.

Considerando patologias que produzem percepção soprosa, Hillenbrand *et al* (1994) empregaram um conjunto de medidas de modo a relacionar com este tipo de voz. Suas medidas acústicas baseiam-se na suavidade do sinal glotal, o que caracteriza a grande amplitude do primeiro harmônico; a perda de periodicidade nos harmônicos, principalmente nas médias e altas frequências; e na maior energia de alta frequência devido ao ruído turbulento.

Definiu-se a medida CPP (*Cepstral peak prominence*) como a diferença em amplitude entre o pico cepstral e o seu valor correspondente em uma regressão linear em um intervalo no cepstrum de *quefreny* de 1 msec até o máximo *quefreny*. Opera como uma medida de periodicidade pois existe relação direta entre a amplitude do pico e quão periódico é o sinal.

Outra medida acústica é o RPK (*Pearson r at autocorrelation peak*), que é obtida pelo cálculo de coeficiente de Pearson entre o sinal de voz e a sua versão deslocada k amostras, onde k corresponde ao ponto de ocorrência do segundo máximo da função de autocorrelação do trecho analisado. Esta medida é realizada a cada 10 ms, usando uma janela de 30 ms, sendo que o valor médio destas estimativas é apresentado como resultado final. Novamente, esta medida serve para avaliar a periodicidade do sinal de voz., pois espera-se obter valores baixos quando o parâmetro é calculado sobre vozes soprosas.

Também foi empregado a medida P/A (*Peak to average ratio*) dada pela razão entre pico máximo do trecho analisado pela amplitude média deste mesmo trecho, retificada em onda completa. O resultado final é dado pela média dos valores determinados em sucessivos segmentos de 10 ms.

Além destas medidas sugeridas, Hillenbrand *et al* (1994) sugerem a utilização dos parâmetros Br, da razão entre a energia acima de 4 kHz (alta frequência) pela energia abaixo desta frequência (baixa frequência) e da amplitude do primeiro harmônico. Verificou-se que as medidas RPK e CPP correlacionam-se com as patologias com características soprosas, além de simples de serem determinadas. Confirmando os estudos de Fukazawa *et al* (1988), Br também serviu para discriminação de soproidade.

Em um estudo para determinação de rouquidão, Shoji *et al* (1992-2) sugere o emprego de sentença fonética /hai/ para medição da eficiência glotal. O parâmetro H-Index é calculado pela diferença entre os dois primeiros máximos do envelope de uma curva de densidade espectral de

potência. Esta curva é determinada sobre os primeiros 275 ms da sentença fonética, sendo que cada ponto desta curva é determinado pela média de 20 amplitudes no espectro de potências do trecho analisado e ao final é aplicado um filtro suavizante por janela triangular. O princípio fisiológico desta medida é que durante o fonema /h/, a capacidade harmônica da glote na produção de um sinal é baixa (ruído turbulento). Após a coarticulação, o fonema /a/ deve apresentar elevada energia harmônica. A capacidade humana de aduzir a corda é então determinada, permitindo avaliar a existência ou não de patologias na glote.

Pode-se perceber pelas medidas apresentadas que a capacidade de produção de ruído turbulento na glote é caracterizado pela proeminência do primeiro harmônico em relação aos demais. A partir do segundo harmônico, a amplitude destas componentes é reduzida, fazendo com que a curva de densidade espectral de potência (para médias frequências) torne-se cada vez mais horizontal. Além disto, o ruído de alta frequência eleva-se. Estas três características descrevem as patologias perceptualmente definidas por soprosas. A percepção da rouquidão se deve pela produção simultânea de aspereza e soprosidade, sendo que a aspereza é um fenômeno caracterizado principalmente por *jitter* e *shimmer*.

Notou-se que a maioria dos trabalhos não tentam discriminar patologias diretamente; apenas relacionar seus índices com identificação perceptual.

2.7 Avaliação por Resíduos

Os métodos anteriormente apresentados empregam medidas tanto no domínio do tempo como no domínio da frequência, buscando representar a forma de onda (normal ou patológica) através de uma pequena quantidade de medidas. A avaliação das estruturas da laringe e cordas vocais exigiria que o sinal considerado não sofresse influência do efeito de ressonância proporcionado pela cavidade supraglotal.

O emprego de fala sustentada (usando fonemas vocálicos) visa minimizar esta influência; entretanto, conforme o fonema empregado, características mais ou menos ásperas podem desviar a discriminação para uma patologia, mesmo que seja de pouca extensão.

Sugere-se então o emprego de uma estimativa do sinal de velocidade volumétrica de ar através da estimativa inversa do modelo de produção de voz. Este sinal glotal permitirá, assim, a avaliação do funcionamento e condição das pregas vocais. Rothenberg (1973) empregou uma máscara pneumotacógrafa para avaliar a velocidade volumétrica de ar na boca, juntamente com um filtro inverso para avaliação de apenas 2 formantes (a resposta em frequência da máscara opera como um passa-baixa com $f_c=1$ kHz) e obteve sinais glotais bastante próximos do real.

Entretanto, surgem alguns problemas nesta avaliação:

1. é difícil determinar os parâmetros do modelo inverso do sinal de voz
2. o sinal glotal contém somente informação de baixa frequência
3. são considerados modelos lineares para o trato vocal, que são invalidados para grandes aberturas glotais.

Como alternativa para esta técnica, pesquisadores (Koike e Markel, 1975; Davis, 1979; Prosek *et al*, 1987; Eskenazi *et al*, 1990) sugerem o uso do resíduo proveniente de uma filtragem inversa como fonte de informação sobre a estrutura glotal, uma vez que elimina-se a correlação proporcionada pela supraglotal. Além disto, o sinal residual mantém informação de alta frequência e de periodicidade, permitindo, sozinho, a discriminação de grande número de patologias.

O sinal residual corresponde ao erro de estimação de um modelo AR para o trato vocal, baseado nas amostras de fala sustentada. O emprego da fala sustentada tem por objetivo manter a supraglotal em um estado estacionário (sem variação das formantes).

Figura 14 - Sinal residual para um fonema /a/ de uma voz normal

A fig. 14 ilustra um sinal residual estimado para o fonema /a/, correspondente ao sinal de voz representado na fig. 8. Pode-se perceber claramente os picos indicando os períodos da fundamental. Além disto, entre estes picos existe informação sobre a quantidade de ruído correspondente ao escape de ar na glote.

Para determinar-se este sinal, é necessário primeiramente determinar os coeficientes do modelo AR, através das equações de Yule-Walker. Estas equações do modelo paramétrico são solucionadas pela técnica de autocorrelação (usando o algoritmo de Levinson), onde o número de coeficientes estimados é 14. Davis (1979) investigou a influência da quantidade de coeficientes e verificou que com o aumento deste, ocorre um incremento na quantidade de picos no espectro de potência (alta resolução) do filtro AR sem uma relação com o número de formantes. Também não emprega-se pré-ênfase de alta frequência de forma a compensar a influência dos lábios no sinal de voz pois não há melhoras significantes na discriminação das disfonias.

O segmento de voz analisado compreende, em média, 138 ms de duração, sendo que o ponto inicial de estimação dos coeficientes é feito a partir de 400 ms do começo da fonação. Este procedimento visa obter a porção mais estável da voz através de uma padronização que permita a construção de métodos automáticos de avaliação do trato vocal, eliminando a entonação inicial.

Definiu-se um conjunto de parâmetros acústicos que caracterizem o sinal residual obtido através de poucas medições. A periodicidade é determinada diretamente através dos índices acústicos PPQ (*Pitch perturbation quotient*) e APQ (*Amplitude perturbation quotient*), com procedimento de cálculo semelhante ao de Koike (1973). São empregados 5 pontos para o cálculo das médias móveis destas duas medidas. O intuito é obter-se avaliação de *jitter* e *shimmer*, minimizando o efeito de variações lentas que ocorrem nos sinais de voz, que tendem à induzir discriminações patológicas errôneas.

Para também avaliar a periodicidade do sinal residual (e assim, do sinal de voz), emprega-se a medida P/A (*peak of autocorrelation*) correspondente ao segundo pico de maior amplitude na autocorrelação normalizada do sinal residual. Assim, quanto mais próximo proeminente o pico for, mais periódico é o sinal de voz.

Considerando a distribuição das amostra de resíduo, verificou-se que a avaliação da curtose pode ser útil na identificação de patologias. O índice acústico EX (*Coefficient of excess*) avalia o grau de achatamento da distribuição de amplitudes do sinal residual. Assim, curvas leptocúrticas caracterizam vozes normais e platicúrticas as vozes disfônicas. O cálculo da curtose corresponde à razão entre o momento de quarta ordem da distribuição pelo quadrado do momento de segunda ordem.

Emprega-se a avaliação no grau de nivelamento (*flatness*) da magnitude espectral, tanto do filtro inverso, quanto do sinal residual pois a proeminência dos picos espectrais determina vozes saudáveis. Assim, os parâmetros SFR (*Spectral flatness of residue*) e SFF (*Spectral flatness filter*) correspondem à medidas de mascaramento de formantes e mascaramento de harmônicos da frequência fundamental, respectivamente, por ruído.

Verificou-se que as alterações orgânicas ou funcionais nas pregas vocais e laringe refletem-se diretamente no sinal residual e por consequência, estes parâmetros podem ser empregados na discriminação patológica. A medida acústica P/A é de grande importância, pois correlaciona-se fortemente com a percepção acústica das patologias.

2.8 Outras técnicas acústicas

Individualmente, os parâmetros acústicos determinados não conseguem identificar perceptualmente as patologias, e assim, efetuar um diagnóstico diferencial correto. Diversos pesquisadores realizaram tentativas de empregar mais de uma medida acústica buscando melhores resultados.

Sanderson e Maran (1992), considerando medidas de frequência fundamental, intervalo de *pitch*, percentual vocálico e *jitter* médio, para fonação contínua, concluíram que é possível avaliar tratamento terapêutico de pacientes disfônicos através da análise de suas vozes, sendo confirmado por Fex *et al* (1994), que empregaram as medidas F_0 , APQ, PPQ, NNE e variação entre níveis espectrais entre períodos da fundamental. Wolfe e Steinfatt (1987) empregaram medidas de desvio padrão e coeficiente de variação das perturbações de primeira ordem do período da fundamental, *jitter* médio, *jitter factor* (JF), *jitter ratio* (JR) e DPF para discriminação entre características perceptuais da voz.

Foi verificado que avaliação acústica da fala permite diagnosticar fadiga laríngea (*myasthenia laryngis*), que corresponde a um primeiro estágio das patologias orgânicas (Eustace *et al*, 1996). Hartman (1984) cita que é possível detectar acusticamente disfonias centrais, geralmente associadas a problemas neurológicos que acabam afetando o funcionamento das pregas vocais.

Entretanto, outros trabalhos indicam a dificuldade de uso de medidas acústicas no pré-diagnóstico de patologias. Martin, Fitch e Wolfe (1995), empregando medidas de *jitter*, *shimmer* e HNR, citam sobre a dificuldade de diagnosticar sopro-sidade. Nittrouer *et al* (1990), usando outras medidas (SNR, relação entre primeiro e segundo harmônicos, perturbação de frequência e amplitude da F_0) citam que os resultados não são suficientemente significativos para um uso exclusivo de avaliação acústica.

Utilizando modelagem biomecânica da fala, Smith *et al*, 1992 mostra ser possível alterar estes modelos de forma a representar paralisias unilaterais e bilaterais. Herzel *et al*, 1994, suscita o uso de teoria de dinâmica não-linear (teoria do caos) para diagnóstico de patologias, visto que o sistema de produção de fala é complexo e não linearidades são características do mesmo.

Gubrynowicz, Mikiel e Zarnecki (1980) utilizando um classificador *fuzzy* com medidas de dispersão e assimetria das diferenças entre períodos sucessivos de *pitch* foram capazes de distinguir entre pessoas normais e pacientes disfônicos. Através de classificação por vetores de quantização para medidas acústicas, Childers e Bae (1992) foram capazes de avaliar a condição das pregas vocais, também separando pacientes disfônicos de pessoas normais. Cita que a classificação empregando sinal proveniente de eletroglotografia prevê resultados inferiores que os obtidos usando sinais de voz.

3. Discriminação Patológica

Neste capítulo são apresentadas as técnicas matemáticas utilizadas no trabalho. A discriminação patológica é realizada através de redes neurais artificiais que são alimentadas por medidas acústicas extraídas dos sinais de voz. A escolha das medidas para o classificador não-linear é realizada através de avaliação estatística.

3.1 Aquisição de vozes

A gravação da voz para análise foi realizada nos ambulatórios de Otorrinolaringologia e de Cirurgia de Cabeça e Pescoço, no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, da Universidade de São Paulo. No período compreendido entre agosto de 96 até agosto de 97, um conjunto de 73 pessoas em que 25 normais e 48 pacientes disfônicos.

A tab. 2 apresenta os tipos de vozes que foram coletadas e a quantidade destas. Um total de 21 disfonias foram encontradas entre os pacientes, sendo que em alguns casos, um mesmo paciente apresentava duas patologias. Isto explica a diferença entre o total de casos desta tabela e a quantidade de pessoas utilizadas na análise (as patologias são descritas no capítulo 2).

O grupo denominado "Normal" corresponde às pessoas sem qualquer disfonia, orgânica ou funcional, não-fumantes, que servem de controle nos estudos. Para os diagnósticos (tanto pacientes quanto das pessoas normais) foi empregado a videolaringoscopia. Em alguns casos foi empregado videolaringoestroboscopia para dirimir dúvidas que ocorriam (com baixa frequência). Normalmente, mais de um especialista efetuava os exames e avaliavam a condição da laringe e das cordas vocais.

Tabela 2 - Relação de patologias e quantidade de pessoas

Patologia	N. de casos
Carcinoma	1
Cisto	4
Corpectomia Parcial	4
Corpectomia Total	1
Disartria	1
Disfonia Funcional	3
Edema	2
Edema de Reinke	6
Espessamento de Pregas Vocais	2
Fenda Fusiforme	8
Granuloma	1
Hipercinesia	1
Hiperemia	1
Hipocinesia	1
Laringite Crônica	2
Leucoplasia	2
Nódulo	5
Normal	25
Outras Disfonias	2
Paralisia Unilateral	1
Pólipo	6
Úlceras de Contato	4
Total	83

3.1.1 Gravação da voz

Na literatura, verifica-se o uso predominante de equipamentos analógicos, com gravação em fita magnética para posterior estudo, seja em computadores ou outro dispositivo como osciloscópio. Holmes (1975), analisando a gravação analógica em meio magnético, verificou que mesmo gravadores analógicos de boa qualidade provocam distorções de fase nas baixas frequências do sinal. Assim, as informações patológicas como frequência fundamental, sinal residual ou glotal podem ser descaracterizadas. Sugere-se o uso de gravação em FM ou utilização direta de equipamento digital.

Este trabalho se apoia no uso direto do computador, recebendo o sinal a ser digitalizado de um microfone. Com apenas uma conversão A/D, elimina-se grande quantidade de ruídos espúrios e assim, evita-se que o sinal de voz digitalizado possa apresentar características de vozes patológicas.

O microfone empregado na aquisição de voz foi um modelo LeSon SM58, unidirecional e de ganho dinâmico, permitindo que informação espúria do ambiente sobre o sinal de voz seja desprezível, juntamente com a boa resposta em frequência deste tipo de microfone.

A conversão analógico-digital foi realizada através de uma placa de aquisição de dados convencional marca SoundBlaster SB16, sendo que a taxa de amostragem utilizada foi de 22050 Hz e discretização das amostras em 16 bits. O uso destes valores é confirmado pelos estudos de Horii

(1979) e Titze, Horii e Scherer (1987) de modo a manter a margem de erro para *jitter* e *shimmer* abaixo de 0,1%. O controle automático de ganho foi definido no programa de gravação para reduzir perturbações lentas de grande amplitude na voz do paciente. Outra necessidade foi a de especificar um valor máximo para o ganho da placa de aquisição para evitar-se o ceifamento de amplitudes e produzir sinal ruidoso aditivo na conversão.

Trabalhos anteriores mostram divergência entre o tipo de fala empregada na gravação (sustentada ou contínua). Apesar da fala contínua ser mais próxima da realidade do paciente, as análises sobre este tipo de fonação devem ser cuidadosas, pois alterações nas articulações da supraglote modificam a densidade de potência espectral. Considerando também que a escolha dos fonemas sustentados varia conforme a medida acústica realizada, efetuou-se a gravação dos fonemas /a/, /e/ e /i/.

Com o microfone posicionado à 5 centímetros em direção da boca, solicitou-se que os pacientes pronunciassem os fonemas (separadamente) por aproximadamente 5 segundos em média, tempo este controlado pelo operador, em níveis de intensidade e altura confortáveis para evitar tensão excessiva sobre as pregas vocais. Caso o paciente não suportasse este período de fonação, solicitava-se que o mesmo cessasse imediatamente de emitir o fonema pois verifica-se um aumento de tensão das pregas vocais para manutenção da vibração destas mediante uma menor quantidade de ar proveniente dos pulmões; isto produz uma estabilidade artificial na produção dos fonemas sustentados (Gelfer, 1995). Espera-se que o fonema /i/ produza menor nível de *jitter* e *shimmer* que os demais fonemas pois as cordas vocais estão em uma condição de maior tensão que nos demais sons vocálicos.

A quantidade de tempo considerada (5 segundos) deve-se aos requisitos de convergência assintótica das perturbações - de amplitude e frequência fundamental - necessários para uma melhor avaliação patológica do paciente (Titze, Horii e Scherer, 1987).

Toda gravação é precedida de uma explicação deste processo de registro da voz ao paciente e realizado efetuado-se um teste para que este sofresse pouca influência psicológica, alterando significativamente sua voz. Algumas gravações foram descartadas e refeitas pois constatou-se diferença entre a emissão da vogal sustentada e o modo com que o paciente falava durante conversas com o mesmo antes da gravação.

Estes cuidados fazem-se necessários para evitar a produção de ruídos espúrios e assim, novas variáveis de que não se tem controle ou conhecimento sobre sua influência no sinal de voz, e assim, dificultar a análise acústica das patologias.

A cada gravação, automaticamente tomava-se nota de algumas informações do paciente. Cadastrou-se o nome completo dos pacientes, sua idade, sexo, atividade que exerce, fonema gravado e a patologia que possui, sendo esta última fornecida pelo especialista.

3.1.2 Cálculos preliminares sobre o sinal de voz

Uma vez que todo o sinal de voz apresenta diferenças em amplitude entre pacientes, algumas medidas acústicas podem ser influenciadas. Assim, é realizado um pré-processamento para normalização das amplitude entre -1 e +1. Isto é realizado por uma ponderação de todas as amostras digitalizadas em relação ao maior valor absoluto destas, que padroniza todas as medidas acústicas, desprezando-se efeitos de intensidade na voz.

Outra necessidade é definir os trechos vocálicos do sinal digitalizado. Apesar da possibilidade de desenvolver-se algoritmo específico para isto, foram eliminados os trechos não-vocálicos manualmente. Sabe-se que os sinais de voz apresentam diferenças de energia sonora entre as porções inicial e final dos trechos vocálicos, devido ao “ataque” do fluxo de ar pela glote e perda de capacidade volumétrica de ar, respectivamente. Krom (1995) mostrou que estes intervalos no sinal de voz produzem maior perturbação em amplitude e frequência, ciclo-a-ciclo, o que pode caracterizar apenas uma condição de energia acústica do fluxo de ar e não uma evidência patológica completa. Apesar de Krom (1995) mostrar que o intervalo inicial do trecho vocálico apresenta maior correlação com a percepção de aspereza na voz de pacientes, estes dois momentos no sinal de voz são retirados manualmente. São apagados aproximadamente 10 períodos de frequência fundamental em cada um destes intervalos. Se houvesse mais de 5 segundos de vocalização ainda restantes, o excedente ainda foi retirado para padronizar o tamanho das amostras de voz.

Verificou-se nos sinais digitalizados que a placa de aquisição utilizada (SoundBlaster SB16) produz um nível DC nos sinais digitalizados indevido, sendo necessário sua retirada. Apesar da influência sobre as medidas acústicas realizadas ser muito pequena, utilizou-se um algoritmo para retirada de tendência linear, que para grande quantidade de amostras, apenas centra o valor médio das amplitudes em zero.

Assim, o pré-processamento é realizado na seguinte ordem:

1. Retirada de trecho não-vocálico, e de ataque inicial e final, do sinal de voz digitalizado
2. Retirada de tendência linear
3. Normalização em amplitude do sinal

O intuito deste pré-processamento é padronizar todos os sinais de voz para que a análise seja uniforme, evitando que os parâmetros acústicos sejam influenciados por características de gravação (mesmo com o controle realizado).

3.2 Medidas de Perturbação

As medidas de perturbação, ciclo-a-ciclo, da frequência fundamental e amplitude são importantes para avaliação das pregas vocais. Pesquisadores, começando por Lieberman (1963), descreveram um conjunto de medidas cuja magnitude de variação permitem diagnóstico diferencial das disfonias. Os objetivos deste tipo de parâmetro é primariamente auxiliar na discriminação de patologias que afetam a periodicidade de adução e abdução das pregas vocais. Disfonias como hipercinesia produzem um movimento irregular da mucosa que reveste as cordas vocais, apesar de organicamente as estruturas de vocalização não apresentarem moléstia. A aspereza (percepção detectada neste tipo de patologia) advém da tensão exagerada sobre a prega vocal.

Outro grupo de patologias que influenciam no movimento glotal são as que produzem voz rouca. Carcinoma, paralisia unilateral, corpectomia são casos em que as pregas vocais também apresentam periodicidade irregular, além de propiciarem o fluxo de ar turbulento pela glote.

Considerando estas características, determinou-se um conjunto de medidas de perturbação, baseado na literatura, empregando em alguns casos elementos estatísticos robustos para obter-se resultados mais precisos, que possam ser empregados na discriminação patológica.

Uma primeira necessidade é estimar o sinal glotal, ou os períodos e amplitudes ciclo-a-ciclo do sinal de voz, para que se possa determinar as perturbações propriamente ditas. Este é um fator crítico em pesquisas da área pois no caso de utilizar-se sinais de voz, a influência da estrutura supraglotal dificulta sua determinação. Objetivamente, o sinal glotal é sempre estimado. Mesmo com o emprego de eletroglotografia ou acelerômetros, os sinais produzidos não são exatamente idênticos ao produzido na laringe:

Considerando apenas o uso da voz, técnicas síncronas e assíncronas, uso de filtragem passa-baixa e medição manual dos períodos de *pitch* em papel milimetrado são os métodos mais empregadas na literatura.

Esta seção apresenta um algoritmo para estimação das amplitudes e períodos da fundamental utilizando o método cepstral e minimização de erro médio quadrático. Posteriormente, as medidas acústicas de perturbação, e o seu procedimento de cálculo, são apresentadas.

3.2.1 Estimador síncrono para período e amplitude de F_0

O algoritmo implementado para determinar os sucessivos períodos e amplitude da frequência fundamental realiza duas etapas distintas, sendo a primeira assíncrona e a seguinte, síncrona. A etapa assíncrona realiza uma estimação do período médio de frequência fundamental para um determinado segmento do sinal de voz, possibilitando que na segunda etapa seja realizado um

refinamento desta estimativa. Isto se faz necessário pois algoritmos síncronos, ou orientados à eventos, em avaliações de laboratório, foram imprecisos para vozes com grande quantidade de oscilação. Resolveu-se então adotar uma técnica híbrida que pudesse ser robusta, independente dos requisitos computacionais.

A fase assíncrona realiza uma estimativa do *pitch* para um segmento de 4096 amostras (185.8 ms) utilizando o método de cepstrum (Noll, 1967). Considerando o uso do cepstrum real, este é matematicamente definido por:

$$rc_s(n) = \text{real}\left\{\mathfrak{T}^{-1}\left\{\log|\mathfrak{T}\{s(n)\}|\right\}\right\} \quad (27)$$

ou seja, o cepstrum real $rc_s(n)$ corresponde a porção real da transformada discreta inversa de Fourier do logaritmo do módulo da transformada discreta de Fourier do sinal de voz. O princípio é que a aplicação do logaritmo sobre um produto equivale à soma dos dois logaritmos. Com este intuito, procura-se realizar a separação do sinal de excitação (variação rápida) e a função de transferência do trato supraglotal (variação lenta), que são convoluídos para produzir o sinal de voz.

Assim, primeiramente teríamos o espectro do sinal de voz representado por:

$$S(\omega) = G(\omega) \cdot V_{VR}(\omega) \quad (28)$$

onde $S(\omega)$ é a transformada discreta de Fourier do sinal de voz

$G(\omega)$ é a transformada discreta de Fourier do trem de pulsos periódico

$V_{VR}(\omega)$ é a transformada discreta de Fourier do trato supraglotal e radiação dos lábios

Aplicando-se o logaritmo nesta expressão, transforma-se este produto em soma. Como a transformada de Fourier é linear, obtém-se então que o cepstrum real de tempo curto de um segmento do sinal de voz, ou seja:

$$rc_s(n, m) = \text{real}\left\{\mathfrak{T}^{-1}\left\{\log|P(\omega)|\right\}\right\} + \text{real}\left\{\mathfrak{T}^{-1}\left\{\log|V_{VR}(\omega)|\right\}\right\} \quad (29)$$

Apenas considera-se a porção real pois o logaritmo do espectro de potência é par. A análise do sinal resultante mostra que as componentes de variação lenta localizam-se nas baixas *quefreny*, enquanto que as variações rápidas, provenientes dos harmônicos no espectro de potência do sinal de voz, ficam dispostos nas altas *quefreny*. A linearidade (adição) obtida no cepstrum real permite que através de *liftering* (filtragem) seja possível separar as componentes harmônicas da estrutura formante.

Empregando um algoritmo de busca de pico, procura-se então o máximo pico cepstral na região de existência do período de frequência fundamental da voz, onde a abscissa deste pico corresponde ao *pitch* médio para o segmento de voz considerado. A região de pesquisa desta frequência corresponde ao intervalo entre 40 e 500 Hz, pois considera-se que a variação do período da fundamental pode ser extrema para vozes ásperas e roucas.

Noll (1967) recomenda um processamento adicional para refinamento do período da fundamental encontrado. Isto porque perturbações temporais entre períodos pode propiciar que o segundo pico cepstral (*rharmônico* do primeiro) tenha uma amplitude superior ao primeiro (e correto) pico do cepstrum, produzindo erros de estimativa. Como isto é algo comum para pessoas com disfonias, foi considerado uma regra heurística que leva em consideração medidas as medidas de período da fundamental realizadas anteriormente:

1. Período de F0 corresponde *quefreny* ATUAL (abscissa do pico máximo ATUAL);
2. Se a *quefreny* do pico máximo ATUAL $\geq 1,6$ vezes o *quefreny* do pico máximo ANTERIOR;
 - 2.1. Busca pico máximo entre $\frac{1}{2}$ da *quefreny* do pico máximo atual - 0,5 ms e $\frac{1}{2}$ da *quefreny* do pico máximo ATUAL + 0,5 ms;
 - 2.2. Se este novo pico máximo localiza-se entre o *quefreny* do pico máximo ANTERIOR - 1 ms e o *quefreny* do pico máximo ANTERIOR + 1 ms;
 - 2.2.1. Então Se a magnitude deste novo pico $> \frac{1}{4}$ da magnitude do pico máximo ATUAL;
 - 2.2.1.1. Período de F0 corresponde à nova *quefreny*;
 - 2.2.2. Senão Se a magnitude deste novo pico $> \frac{1}{2}$ da magnitude do pico máximo ATUAL;
 - 2.2.2.1. Período de F0 corresponde à nova *quefreny*.

Pode-se perceber a necessidade de se manter as estimativas (corrigidas) no ciclo anterior (amplitude e *quefreny* do cepstrum anterior), além de buscas na região de possível existência do primeiro *rharmônico* (passo 2.1). Os passos 2.2.1 e 2.2.2 são necessários para verificar se a amplitude do novo pico pesquisado é condizente com uma estimativa real para a frequência fundamental.

A partir desta estimativa assíncrona para o segmento considerado, é realizada a etapa síncrona, para determinar precisamente todos os períodos da fundamental neste trecho de voz analisado. Para auxiliar o processo de busca, é mantido paralelamente ao sinal de voz, uma versão filtrada do mesmo, na qual empregou-se um filtro passa-baixa FIR de ordem 100 com frequência de corte em 800 Hz.

Realiza-se uma pesquisa de todas as amplitudes mínimas (em segmentos periódicos baseado na estimativa da F_0) pois este evento é o que mais se aproxima de um efeito físico do sistema produtor de fala (adução completa das pregas vocais). Estes períodos de amplitude mínima ainda necessitam ser melhor definidos pois a discretização acarreta em perda de resolução. Para contornar este problema, empregou-se a técnica de minimização de erro médio quadrático entre amplitudes mínimas (Titze e Liang, 1993), baseado na metodologia de Milenkovic (1987).

Considerando que um sinal de voz sustentada pode ser definido pela soma entre uma componente periódica e outra aperiódica, ou seja:

$$s(t) = h(t) + e(t) \quad (30)$$

Considerando esta aproximação, a porção periódica pode ser definida por

$$p(t) = h(t - t_p) \quad (31)$$

onde t_p é o período de frequência fundamental para o segmento considerado.

A estimativa deste período se dá através do critério de minimização do erro médio quadrático. Para tanto, supõem-se que $e(t)$ é estacionário, com média zero e apresenta decorrelação entre períodos sucessivos. Sabendo que:

$$s(t) - s(t - t_p) = e(t) - e(t - t_p) \quad (32)$$

Assim, a função à minimizar-se é dada por:

$$mse(t_p) = \frac{1}{2} E \left\{ [e(t) - e(t - t_p)]^2 \right\} \quad (33)$$

ou

$$mse(t_p) = \frac{1}{2} E \left\{ [s(t) - s(t - t_p)]^2 \right\} \quad (34)$$

Para determinar o período de *pitch* que minimiza esta função, determina-se uma região de busca para que possa realizar-se a variação de t_p , reduzindo a quantidade de processamento. Titze e Liang (1993) sugerem que este intervalo tenha magnitude entre 5 a 15% da estimativa anterior do período de frequência fundamental, com centro definido nesta estimativa. No algoritmo implementado foi empregado 5%. Aplica-se uma interpolação parabólica usando o erro mínimo, $mse(t_p)$, calculado para determinar a melhor estimativa do período da fundamental.

Aplicado este critério sobre todos os períodos existentes no segmento atual de 4096 amostras, realiza-se um último refinamento. Usando a interpolação parabólica diretamente sobre o sinal de voz no instante definido por t_p , determina-se o valor da amplitude para o presente ciclo (que será utilizado no cálculo do *shimmer* do sinal de voz).

Este critério empregado garante a robustez da estimativa de frequência fundamental para um ruído aditivo. Isto é útil pois muitas patologias (que fornecem percepção soprosa) se caracterizam pelo aumento de oscilações aleatórias e métodos para estimação de F_0 puramente orientados à eventos como cruzamentos por zero e coleta de picos (*peak-picking*) tem sua capacidade deteriorada nesta situação, muito comum em disfonias.

Este processo (etapa assíncrona/síncrona) é realizada sobre todo o sinal de voz padronizado, com movimentação da janela de análise. Pode-se dizer que a fase assíncrona opera como um indicador na busca dos períodos de frequência fundamental, pois o algoritmo é fortemente baseado na segunda etapa.

3.2.2 PFF (Perturbation Frequency Factor)

Esta medida segue a definição de Lieberman (1963, 1967). Determinando as diferenças entre os períodos de frequência fundamental sucessivos, este pesquisador construiu sua curva de distribuição. Verificou que as patologias concentravam-se a partir de diferenças de período, ciclo-a-ciclo, acima de 0,5 ms. Assim, esta medida acústica é definida como a integral desta distribuição para perturbações superiores à 0,5 ms.

Uma definição equivalente é que PFF corresponde à percentagem de todas as diferenças de período da fundamental superiores à ½ ms.

3.2.3 DPF (Direction Perturbation Factor)

Ignorando as variações de amplitude das diferenças entre os períodos glotais adjacentes, define-se DPF, segundo Hecker e Kruel (1971), como sendo a percentagem de alterações do sinal algébrico destas diferenças em relação ao número de intervalos de pulsos glotais.

Para determiná-lo, determina-se primeiramente as diferenças entre períodos (ou amplitudes) de frequência fundamental, das quais se mantém apenas o sinal algébrico. Calculando a relação entre o número de mudanças de sinal pelo número de períodos (ou amplitudes) da fundamental, obtém-se o percentual de perturbação diferencial, ou DPF.

3.2.4 JR (Jitter Ratio)

Por verificar-se que o aumento da frequência fundamental propicia uma redução na perturbação média (redução de *jitter* médio), este parâmetro foi construído para minimizar tal efeito.

A expressão que define esta medida é dada abaixo

$$JR = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |P_i - P_{i+1}|}{\frac{1}{N} \sum_{i=1}^N P_i} \cdot 1000 \quad (35)$$

ou seja, a razão entre a perturbação média dos períodos de frequência fundamental pelo período de *pitch* médio. O fator aplicado (1000) na equação acima apenas evita a produção de valores de baixa magnitude.

Apesar desta medida apenas ser empregada na literatura para períodos sucessivos, esta expressão também foi utilizada considerando as amplitudes ciclo-a-ciclo, para verificar seu potencial na discriminação de patologias.

3.2.5 JF (Jitter Factor)

De modo análogo à medida acústica JR, considerando o recíproco do período (ou da amplitude) da fundamental determina-se o índice JF, também com o intuito de produzir uma medida que não sofra influência das alterações da frequência fundamental média. Assim, tem-se a seguinte expressão:

$$JR = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |F_i - F_{i+1}|}{\frac{1}{N} \sum_{i=1}^N F_i} \cdot 100 \quad (36)$$

Novamente um fator multiplicativo é empregado para que não se produza valores muito pequenos para o índice.

3.2.6 Séries Temporais

Baseado na sugestão de Shoentgen e Guchteneere (1995), deseja-se produzir uma medida que elimine a correlação existente entre perturbações adjacentes (tanto de período quanto amplitude).

Para determinar esta medida, determina-se os resíduos de uma estimação AR sobre as diferenças entre o i -ésimo período da fundamental pelo período de F_0 médio, ou seja, estima-se a seguinte série descorrelacionada:

$$Z_i = (P_i - \bar{P}) - a_1(P_{i-1} - \bar{P}) - \dots - a_q(P_{i-q} - \bar{P}) \quad (37)$$

onde q é a ordem do modelo,

P_i é o i -ésimo período (ou amplitude) da frequência fundamental,

$\bar{P}$ é a média do período (ou amplitude) da frequência fundamental.

Para determinar os coeficientes do modelo AR acima, e por conseguinte, os resíduos da estimação, utiliza-se o algoritmo de Levinson para solucionar as equações de Yule-Walker, pois este método permite que a ordem do modelo aumente a cada iteração.

Para determinar esta ordem, q , emprega-se um teste estatístico (teste χ^2) para avaliar a brancura do resíduo Z_i obtido. Neste teste, a hipótese nula é que os resíduos obtidos correspondem a um ruído branco. Em caso de rejeição, executa-se uma nova iteração do algoritmo de Levinson. Limitou-se em 20 o número máximo de iterações, pois nas análises de Schoentgen e Guchteneere (1995) constatou-se que poucos modelos AR para estes resíduos verificam a hipótese nula para um número de coeficientes superior a 8.

Determinados os resíduos da estimativa paramétrica, aplica-se então uma média móvel (também denominado PPQ modificado) de comprimento 5 para eliminação das perturbações lentas que possam ocorrer no sinal residual, ou seja:

$$PPQ_{\text{mod}} = \frac{\frac{1}{N-5} \sum_{i=6}^N |Z_i|}{\frac{1}{N} \sum_{i=1}^N P_i} \quad (38)$$

Apesar do trabalho desses pesquisadores apenas citar o uso para perturbações em frequência, considerou-se coerente aplicá-lo sobre as amplitudes da fundamental pois certamente estas são correlacionadas com os valores vizinhos.

As medidas para este algoritmo são realizadas à cada 256 amostras sobre um conjunto de 512 estimativas de período (ou amplitude) de F_0 , sendo que determina-se ao final a mediana destes valores como índice representativo do sinal de voz.

3.2.7 AVI/PVI (Amplitude/Perturbation Variation Index)

Baseado em estatística descritiva, aplica-se o coeficiente de variação dos períodos, ou amplitudes, da frequência fundamental, sobre todas as amostras padronizadas do sinal de voz. O coeficiente de variação é definido por:

$$CV = \frac{\frac{1}{N} \sum_{i=1}^N (P_i - \bar{P})^2}{\bar{P}^2} \quad (39)$$

onde P_i é o i -ésimo período (ou amplitude) da frequência fundamental,

$\bar{P}$ é a média do período (ou amplitude) da frequência fundamental.

O parâmetro acústico PVI corresponde ao coeficiente de variação aplicado sobre os períodos de frequência fundamental multiplicado por 1000 para que não se tenha valores muito pequenos, enquanto que a medida AVI, além de resultar do produto entre o coeficiente de variação pela mesma constante sobre as amplitudes da fundamental, aplica-se o logaritmo decimal, pois existe a indicação de alto grau de linearidade positiva entre o nível de percepção acústica de aspereza e a aplicação da escala logarítmica na medida AVI.

3.2.8 Médias Móveis

Na literatura estas medidas são encontradas sobre nomes como RAP (Relative Average Perturbation), PPQ (Period perturbation Quotient) ou APQ (Amplitude Perturbation Quotient). O princípio que rege a aplicação destas medidas sobre as estimativas tanto de período quanto de amplitude da frequência fundamental é a existência de variações lentas, sem relação com as disfonias, que ocorrem no sinal de voz. Estas oscilações afetam medidas de *jitter*, elevando seu valor, induzindo a identificação errônea de patologias da laringe.

A expressão genérica para médias móveis é:

$$MM = \frac{\frac{1}{N-P+1} \sum_{i=1}^{N-P+1} \left| \frac{1}{P} \sum_{j=1}^P (X_{i+j-1}) - X_{i+0.5(P-1)} \right|}{\frac{1}{N} \sum_{i=1}^N X_i} \quad (40)$$

onde N é a quantidade de períodos, ou amplitudes usados no cálculo

P é o tamanho da janela para média móvel

X_i é a i -ésima medida de período, ou amplitude, ciclo-a-ciclo

onde P é a ordem, ou comprimento da janela de dados.

No cálculo deste parâmetro foi considerado para mensuração das perturbações de F_0 , janelas de comprimento 5 e 11, condizente com o trabalho de Davis (1979). Para perturbações em amplitude, utilizou-se um comprimento de 11 amostras.

3.2.9 Perturbações de ordem superior

Classificou-se um conjunto de medidas que se aplicam sobre perturbações de ordem superior (tanto em período quanto amplitude da frequência fundamental), ou seja, sobre derivadas discretas com ordem variando entre 1 e 4, sendo que a perturbação de ordem zero é o desvio entre as amostras e a média destas amostras. Sobre este conjunto de 10 subsinais (5 para períodos e 5 para amplitudes de F_0) aplica-se um conjunto de medidas estatísticas, para avaliação da capacidade destas medidas em discriminar patologias.

Matematicamente tem-se que a perturbação de ordem zero é dada pela seguinte expressão:

$$p_i^0 = P_i - \bar{P} \quad (41)$$

onde P_i é o i -ésimo período (ou amplitude) da frequência fundamental,

$\bar{P}$ é a média do período (ou amplitude) da frequência fundamental.

A k -ésima perturbação é definida por:

$$p_i^k = \begin{cases} p_i^{k-1} - p_{i-1}^{k-1}, & k \text{ par} \\ p_{i+1}^{k-1} - p_i^{k-1}, & k \text{ ímpar} \end{cases} \quad (42)$$

Um conjunto de medidas estatísticas (clássicas e robustas) foi utilizado para caracterizar este “sinal” de perturbações de período e amplitude da fundamental. Como medidas de tendência central, utilizou-se a média e mediana das perturbações de ordem k .

Para avaliar a dispersão destas perturbações, utilizou-se desvio absoluto, desvio padrão, assimetria e curtose, para se avaliar o comportamento da distribuição. Finalmente, utilizou-se regressão linear para determinar o comportamento das amostras de perturbação através de estimação clássica (por mínimos quadrados) e robusta (por minimização do desvio absoluto do valor mediano).

Nesta última abordagem são considerados os coeficientes da regressão linear e o desvio quadrático das regressões.

A medida de assimetria, baseada no momento de terceira ordem, permite avaliar a tendência lateral da curva de distribuição das perturbações, sendo descrita pela seguinte equação:

$$assimetria = \frac{\sum_{i=1}^N (p_i^k)^3}{N \cdot \sigma_{p^k}^3} \quad (43)$$

Determinando o grau de achatamento da distribuição das perturbações, utiliza-se a medida de curtose, definida pela expressão:

$$curtose = \frac{\sum_{i=1}^N (p_i^k)^4}{N \cdot \sigma_{p^k}^4} - 3 \quad (44)$$

Esta medida é semelhante ao parâmetro EX definido por Davis (1979), onde constatou-se que as disfonias são caracterizadas por curvas de distribuição platicúrticas (curtose maior que 0), enquanto que as curvas leptocúrticas são encontradas em pessoas normais.

A regressão linear clássica foi definida através do critério de mínimos quadrados, ou seja, minimiza-se o somatório do quadrado das diferenças,

$$F(a, b) = \sum_{i=1}^N (p_i^k - a - b \cdot i)^2 \quad (45)$$

Os coeficientes a e b , são determinados por

$$b = \frac{N \cdot \left(\sum_{i=1}^N i \cdot p_i^k \right) - \left(\sum_{i=1}^N i \right) \cdot \left(\sum_{i=1}^N p_i^k \right)}{N \cdot \left(\sum_{i=1}^N i^2 \right) - \left(\sum_{i=1}^N i \right)^2} \quad (46)$$

e

$$a = \frac{\left(\sum_{i=1}^N p_i^k \right) \cdot \left(\sum_{i=1}^N i^2 \right) - \left(\sum_{i=1}^N i \cdot p_i^k \right) \cdot \left(\sum_{i=1}^N i \right)}{N \cdot \left(\sum_{i=1}^N i^2 \right) - \left(\sum_{i=1}^N i \right)^2} \quad (47)$$

O desvio da regressão linear é finalmente calculado como o desvio padrão dos resíduos desta estimação linear. O objetivo destas medidas de regressão é verificar se as perturbações apresentam elevação ou redução linear significativa durante a fonação. Além disto, através do desvio de

regressão, deseja-se determinar a dispersão das perturbações, eliminando-se a influência da inclinação linear que estes “subsinais” possam apresentar.

Para suportar oscilações rápidas aleatórias que possam ocorrer nas perturbações de ordem superior, utilizou-se uma regressão linear robusta, menos susceptível à ruídos espúrios. O princípio da técnica é minimizar a expressão:

$$F(a, b) = \sum_{i=1}^N |p_i^k - a - b \cdot i| \quad (48)$$

Utiliza-se uma simplificação na qual se minimiza uma soma absoluta qualquer empregando a mediana (x_M) destas amostras, ou seja:

$$\sum_i |x_i - x_M| \quad (49)$$

Fixando inicialmente o coeficiente b (estimativa), o coeficiente a pode ser obtido através da expressão:

$$a = \text{mediana}\{p_i^k - b \cdot i\} \quad (50)$$

Uma vez determinado o coeficiente a , pode-se finalmente determinar o novo coeficiente b através da equação:

$$\sum_{i=1}^N i \cdot \text{sinal}\{p_i^k - a - b \cdot i\} = 0 \quad (51)$$

onde a função *sinal* é -1 para valores negativos, +1 para valores positivos e *sinal*(0)=0.

O objetivo de se minimizar a discrepância entre cada perturbação e a sua mediana é obter uma equação de reta linear mais representativa. Para esta determinação, utiliza-se algoritmos iterativos e numéricos para obter-se os coeficientes a e b . Empregou-se o método da bisseção para obter estes valores, sendo utilizada uma estimativa inicial empregando o método dos mínimos quadrados. De forma análoga à regressão linear por mínimos quadrados, o desvio da regressão robusta é obtida pelo desvio padrão dos resíduos desta estimação.

A partir deste contingente de medidas estatísticas, pretende-se determinar a capacidades destas na discriminação patológica. A grande quantidade de medidas sobre os sinais de perturbação de ordem superior visa caracterizar adequadamente suas amostras e verificar o comportamento destas para cada disfonia.

3.3 Medidas de Ruído

A presença de ruído no sinal de voz é um grande indício de anormalidades nas pregas vocais. Patologias como pólipos, nódulos e cistos impedem a correta adução das pregas vocais, resultando em escape de ar que torna-se turbulento e produz a sensação acústica de sopro. Quando ocorrem perturbações na frequência fundamental (seja em magnitude ou nos períodos sucessivos), como em carcinomas, a voz torna-se rouca.

Baseado nestas considerações, é importante examinar a curva de densidade espectral de potência à procura de oscilações aleatórias entre as componentes harmônicas do sinal de voz, verificar o comportamento da envoltória espectral, relacionando a perda energética das formantes com o aumento da magnitude nas regiões espectrais não-formantes. Além disto, torna-se imprescindível quantificar estas informações, além da própria quantidade de ruído de alta frequência. Para isto, algumas medidas acústicas foram utilizadas para fornecer uma quantificação da turbulência glotal, sendo extraídas do trecho normalizado do sinal de voz.

3.3.1 HPFR

Para avaliar o ruído de alta frequência, empregou-se a estimativa HPFR sugerida por Shoji *et al* (1992-1). Determina-se a partir do sinal de voz normalizado a média das curvas de densidade de potência espectral em escala logarítmica (contendo 1024 pontos cada uma), com deslocamento de 25,6 ms entre cada estimativa. Por fim, aplica-se a seguinte expressão, que corresponde a média das magnitudes espectrais em determinadas faixas de frequência.

$$HPFR_i = \sum_{\omega=f_l}^{f_h} \overline{|S(\omega)|^2} \quad (52)$$

onde f_l e f_h indicam o intervalo de estimação da energia média espectral.

Utiliza-se uma janela de Blackman para reduzir o espalhamento espectral sobre cada trecho avaliado. Os intervalos considerados na análise são 60 a 400 Hz, 400 a 2000 Hz, 2000 a 5000 Hz, 5000 a 8000 Hz e a energia acima de 6000 Hz até o limite de Nyquist (como empregou-se uma frequência de amostragem de 22050 Hz, o patamar máximo considerado é de 11025 Hz).

Uma vez determinadas estas medidas (HPFR_i) empregou-se medidas estatísticas (clássicas e robustas) para avaliar o comportamento da distribuição dos valores ao longo do sinal de voz. Utilizou-se a média, mediana, desvio absoluto, desvio padrão, índices de assimetria e curtose, coeficientes da regressão linear obtida por mínimos quadrados, coeficientes de regressão linear robusta, e os desvios médio quadrático de ambas as regressões. A descrição destes índices estatísticos pode ser encontrada na seção 3.2.9.

3.3.2 Medidas de Energia Espectral

Com o objetivo de avaliar o comportamento da energia espectral das disfonias, determinou-se para cada trecho do sinal de voz normalizado contendo 1024 amostras, com deslocamento temporal de 20 ms, a curva de densidade de potência espectral em escala logarítmica (utilizou-se uma janela de Hanning para redução do lóbulos laterais). Sobre este segmento espectral, determinou-se a magnitude média em determinadas faixas de frequência, a saber, 60 a 400 Hz, 400 a 2000 Hz, 2000 a 5000 Hz, 5000 a 8000 Hz.

Uma vez calculado todas estas quatro medidas espectrais para a totalidade das amostras do sinal de voz, aplicou-se técnicas estatísticas para avaliar o comportamento da distribuição destas medidas, ou seja, média, mediana, desvio absoluto, desvio padrão, índices de assimetria e curtose, coeficientes da regressão linear obtida por mínimos quadrados, coeficientes de regressão linear robusta, e os desvios médio quadrático de ambas as regressões. Este grupo de índices estatísticos será empregado em outros parâmetros acústicos, sendo totalmente descritos na seção 3.2.9.

3.3.3 Br

Para quantificar o ruído turbulento em patologias com características perceptuais de sopro, foi sugerido por Fukazawa *et al* (1988) a medida Br (*Breathiness Ratio*), fornecendo avaliação deste ruído em relação às demais componentes.

No procedimento de cálculo, primeiramente aplica-se um filtro de pré-ênfase, utilizando ganho de +6dB/oitava a partir da frequência nula até 10 kHz.

Considerando que o ruído turbulento caracteriza-se por componentes aleatórias de alta frequência, acompanhado de redução harmônica, determina-se o índice acústico sobre o sinal filtrado pela seguinte equação

$$Br = \frac{\sum_{i=1}^{N-2} [s(t_i) - 2s(t_{i-1}) + s(t_{i-2})]^2}{\sum_{i=1}^N [s(t_i)]^2} \quad (53)$$

onde N é a quantidade de amostras do sinal de voz normalizado.

Apesar dos autores citados desenvolverem o cálculo apenas sobre um trecho de 25,6 ms, este índice acústico foi determinado sobre toda a extensão do sinal normalizado. Comparando-se os resultados com patologias similares às utilizadas por estes pesquisadores, não houve diferença estatística significativa.

3.3.4 SNL

Para diferenciar vozes normais das que apresentam características ásperas, Sansone e Emanuel (1970), Lively e Emanuel (1970) e Emanuel *et al* (1973) desenvolveram a medida SNL (*Signal to noise level*) para quantificar esta característica perceptual. O procedimento de cálculo era realizado sobre curvas de densidade de potência espectral impressas em papel milimetrado.

Baseado nas descrições deste método, determinou-se a potência espectral para cada trecho de 4096 pontos, com deslocamento de 25,6 ms entre segmentos, sendo previamente aplicado uma janela de Hamming. Em cada uma destas curvas de energia, determinou-se o valor mínimo de magnitude à cada 100 Hz, dentro de um intervalo de frequências (100 à 2600 Hz, 2600 à 5100 Hz, 5100 à 8000 Hz, 100 à 5100 Hz e 100 à 8000 Hz).

Ao final, obtém-se a média destes valores mínimos, ou seja, valor parcial de SNL. Com todos os valores parciais extraídos das janelas do sinal de voz, aplica-se as medidas estatísticas descritas na seção 3.2.9.

3.3.5 HNR

Vozes roucas são caracterizadas por componentes aleatórias na região da frequência formante, aparecimento de ruído de alta frequência e perda de magnitude nas frequências harmônicas. Assim, para quantificar estes elementos do sinal de voz patológico, foi desenvolvido o índice acústico HNR (*Harmonic to noise ratio*) por Yumoto *et al* (1982).

O procedimento para obtê-lo exige a construção de um algoritmo para determinar os trechos periódicos de frequência fundamental sincronamente, pois é necessário extrair um conjunto de segmentos para se estimar esta medida acústica. O algoritmo utilizado emprega duas fases distintas na estimação dos períodos de *pitch*: uma assíncrona e outra síncrona.

A primeira fase corresponde à determinação da F_0 média para um conjunto de 4096 amostras. Isto é realizado através da técnica de estimativa cepstral (Noll, 1967).

A fase posterior utiliza o valor médio da frequência fundamental para buscar os cruzamentos por zero que definem o início e fim de um período da fundamental. Primeiramente determina-se a amplitude mínima dentro de um intervalo de existência e a partir deste, encontra-se o próximo cruzamento por zero, à direita. Com todas as estimativas dos cruzamentos de zero separados por um período da fundamental, pode-se determinar as formas de onda da fundamental, ciclo-a-ciclo.

No procedimento de cálculo da medida, é necessário determinar a média para um conjunto de 50 formas de onda periódicas, ou seja:

$$s_m(\tau) = \sum_{i=1}^{50} \frac{s_i(\tau)}{50} \quad (54)$$

onde τ corresponde ao máximo período da frequência fundamental, sendo τ_i o período do i -ésimo segmento periódico analisado (para os trechos com período $\tau_i < \tau$, considera-se $s_i(t) = 0$ para $\tau_i < t < \tau$).

Com a forma de onda média (s_m), que no domínio da frequência corresponde à um sinal imune de ruído, obtém-se a energia das componentes harmônicas através da expressão:

$$H = n \cdot \sum_{i=1}^{\tau} s_m^2(i) \quad (55)$$

A determinação da energia do ruído do sinal analisado baseia-se na diferença entre cada segmento $f_i(\tau)$ e forma de onda média, ou seja,

$$N = n \cdot \sum_{i=1}^{\tau} \sum_{t=1}^{\tau_i} [s_i(t) - s_m(t)]^2 \quad (56)$$

Assim, estima-se para um segmento do sinal de voz com tamanho igual à 50 períodos de frequência fundamental uma estimativa da relação entre a parte harmônica e a ruidosa. Para evitar suposições sobre qual o trecho mais estável do sinal de voz (onde se extrairá este índice acústico), utilizou-se de sobreposições de 10 períodos da fundamental, até a totalidade do sinal normalizado. A partir destas medidas ao longo do sinal de voz, aplicou-se as medidas estatísticas da seção 3.2.9 para avaliar o seu comportamento.

3.3.6 Hr

Outra avaliação da rouquidão na fala sustentada é o Hr (*Harmonic ratio*), apresentado por Hiraoka *et al* (1984). Esta medida acústica baseia-se na densidade de potência espectral para avaliar a relação entre as componentes harmônicas e as não harmônicas.

Primeiramente aplica-se um filtro FIR passa-baixa de ordem 100 com frequência de corte em 8 kHz sobre o sinal de voz. Calcula-se então, à cada 25,6 ms, um periodograma de 4096 pontos (com aplicação de uma janela de Hamming) sobre o qual será determinada a relação harmônica.

Uma modificação é realizada no procedimento apresentado pelos autores. Utilizou-se do método de estimação de frequência fundamental por cepstrum (Noll, 1967) ao invés de determiná-la a partir do periodograma (conforme citado pelos autores). A partir da F_0 pode-se determinar as componentes harmônicas por algoritmo de busca na região de existência destas, na curva de densidade de potência espectral. Esta região é definida em um intervalo de $\pm 1/3$ do período da fundamental, centrada em intervalos múltiplos de F_0 .

A medida de relação harmônica é então definida por

$$Hr = \left(\frac{\sum_{i>1} h_i}{P} \right) \quad (57)$$

onde P corresponde ao somatório dos máximos, harmônicos, na curva da densidade de potência espectral;

h_i é a magnitude das componentes harmônicas.

Os somatórios da eq. (57) são desenvolvidos até a frequência de 10 kHz. A partir das estimativas da relação harmônica para cada segmento de 4096 pontos, aplica-se as medidas estatísticas apresentadas na seção 3.2.9, de modo a analisar o comportamento da distribuição destes índices.

3.3.7 Ra

Para determinar a quantidade de ruído no sinal de voz, Kojima *et al* (1980) consideraram a utilização um trecho de tamanho igual à três períodos da fundamental do sinal de voz, e avaliaram a densidade espectral de potência do sinal resultante da repetição deste trecho.

A estimativa dos períodos da frequência fundamental foi feita através do algoritmo híbrido de duas fases (assíncrona e síncrona) descrito na seção 3.3.5. Novamente se emprega o método cepstrum de Noll (1967) para determinar o período de F_0 médio.

A partir destas estimativas (marcações dos cruzamentos de zero), construi-se um sinal com 4096 pontos empregando repetições de três períodos sucessivos da fundamental. O cálculo do índice acústico é realizado sobre o periodograma deste novo sinal, onde se estima o somatório quadrático das componentes harmônicas (H_i) e o somatório quadrático dos pontos de máximo na região entre as harmônicas (N_i). A determinação das componentes harmônicas sobre a curva de densidade espectral é feita através de busca na sua região de existência ($\pm 1/3$ do período da fundamental, centrada em intervalos múltiplos de F_0). Assim, à cada 325 ms do sinal de voz, determina-se uma avaliação parcial da medida, ou seja,

$$Ra = 10 \cdot \log \left(\frac{\sum_{i=1}^N H_i}{\sum_{i=1}^N N_i} \right) \quad (58)$$

onde N é o número de trechos de comprimento igual à três períodos da fundamental

Aplica-se este procedimento de estimativa sobre toda extensão do sinal de voz normalizado. Também foi testado o efeito de elevação de três para seis períodos na composição do sinal analisado espectralmente. Emprega-se então as medidas estatísticas definidas na seção 3.2.9 para determinar o comportamento da distribuição deste índice no sinal de voz.

3.3.8 NNE

Kasuya *et al* (1986) propõem estimar a quantidade de ruído, baseada nos intervalos entre as magnitudes harmônicas. Considerando que trechos do sinal de voz (utiliza-se janela de Hamming para reduzir os lóbulos espectrais laterais) são compostos por uma componente periódica e outra de ruído aditivo. Assim, no domínio da frequência, tem-se

$$S_m(\omega) = H_m(\omega) + N_m(\omega), \quad \omega = 0, 1, \dots, N-1 \quad (59)$$

onde m é o índice da janela de dados analisada

$S_m(\omega)$ é a componente harmônica

$N_m(\omega)$ é a componente ruidosa

N é a quantidade de pontos da transformada discreta de Fourier

Para avaliar a relação de energia entre as componentes do sinal de voz, utiliza-se a seguinte equação, correspondente ao índice acústico propriamente dito

$$NNE = 10 \cdot \log \frac{\frac{1}{L} \sum_{\omega=N_L}^{N_H} \sum_{m=1}^L |\hat{N}_m(\omega)|^2}{\frac{1}{L} \sum_{\omega=N_L}^{N_H} \sum_{m=1}^L |S_m(\omega)|^2} \quad (60)$$

onde $\hat{N}_m(\omega)$ é a estimativa espectral do ruído;

$S_m(\omega)$ é a densidade de potência espectral do sinal de voz;

L é a quantidade de janelas de dados que são utilizadas para análise;

N_L e N_H são os limites inferiores e superiores, em amostras, sobre o qual a medida é desenvolvida.

Para estimar o ruído, considera-se que entre os picos harmônicos, a única componente existente é ruidosa. Na região do pico, a existência das duas componentes, não é possível separar-se $N_m(\omega)$ de $S_m(\omega)$. Assim, utiliza-se do valor médio dos dois vales laterais de cada pico harmônico como sendo a componente ruidosa naquelas regiões, ou seja,

$$|\hat{N}_m(\omega)|^2 = \begin{cases} |S_m(\omega)|^2 & \omega \in D_j \\ \frac{\sum_{k \in D_j} |S_m(k)|^2}{2 \cdot N_j} + \frac{\sum_{k \in D_{j+1}} |S_m(k)|^2}{2 \cdot N_{j+1}} & \omega \in P_j \end{cases} \quad (61)$$

onde j indica uma região contendo um vale e um pico harmônico,

As regiões D_j e P_j são definidas com base na largura de banda da janela espectral (no caso, Hamming). Assim, estas são definidas por

$$\begin{aligned}
 P_j &= \left\{ j \mid j_i - \frac{2N}{M} \leq j \leq j_i + \frac{2N}{M} \right\} \\
 D_j &= \left\{ j \mid j_{i-1} + \frac{2N}{M} \leq j \leq j_i - \frac{2N}{M} \right\}
 \end{aligned} \tag{62}$$

onde M é a quantidade de amostras da m -ésima janela analisada

O algoritmo desenvolvido primeiramente determina o período médio de frequência fundamental através do método de autocorrelação. Extrai-se então um trecho de tamanho $7 T_p$ (período médio de F_0) e aplica-se a transformada discreta de Fourier com 1024 pontos. Sobre este resultado então, estima-se a posição de cada pico harmônico através de busca centrada nas frequências múltiplas da fundamental, permitindo então o cálculo de P_j e D_j . Se houver sobreposição entre larguras de bandas sucessivas, estas são desconsideradas na estimação da medida acústica.

Finalmente, aplica-se a eq. (60) para estimativa da medida NNE em determinados intervalos N_l e N_h amostrais na transformada discreta de Fourier. Os intervalos considerados (em termos de frequência) são 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz e 1000 à 5000 Hz.

3.3.9 N/S

Outra medida empregada na avaliação de rouquidão foi sugerida por Muta *et al* (1988). Seu cálculo baseia-se em estimativa síncrona da frequência fundamental e posterior avaliação da relação sinal ruído. O procedimento é realizado em três etapas distintas. A primeira etapa corresponde à extração da frequência fundamental do sinal.

Utilizando uma estimativa do período da fundamental (F_0) baseada no método cepstral (Noll, 1967), extrai-se um segmento com tamanho igual à 4 vezes àquele período estimado. Aplica-se uma janela de Hanning sobre o mesmo e determina-se a autocorrelação deste. Compara-se então a estimativa F_0 com a obtida da autocorrelação (F) e em caso de discrepância, considera-se $F_0 = F$, reexecutando esta etapa (desconsiderando apenas a fase cepstral) até que se obtenha um trecho contendo exatamente quatro períodos da fundamental. Como havia sinais muito ruidosos, definiu-se um número máximo de tentativas igual à 20.

A próxima etapa corresponde à análise síncrona do *pitch*. Considerando o sinal janelado da fase anterior, $s(n)$, determina-se a curva de densidade espectral de potência através de:

$$H(\omega) = \left| \sum_{n=1}^{M_R} s(n) \cdot e^{\frac{-j2\pi\omega n}{M_R}} \right|^2 \tag{63}$$

onde M_R é igual à $4F$.

Finalmente, avalia-se a relação sinal-ruído (NS) através da expressão

$$NS = 10 \cdot \log \frac{\sum_{\omega=3}^{4L+2} \min_{i=-1,0,1,2} H(4\omega + i)}{\sum_{\omega=3}^{4L+2} H(\omega)} \quad (64)$$

onde L é o número de harmônicos empregados no cálculo (no caso, $L=16$).

Antecedendo o método apresentado, aplica-se um filtro FIR passa-baixa de ordem 100, com frequência de corte em 4800 Hz. A cada estimativa da medida acústica, desloca-se a janela de análise em 25,6 ms. Ao conjunto de índices determinados sobre o sinal de voz, aplica-se os cálculos estatísticos apresentados na seção 3.2.9.

3.3.10 HNR Cepstral

Uma interessante técnica de avaliação do ruído no sinal de voz é através de *liftering* do cepstrum real. Krom (1993) especifica um método que foi empregado neste trabalho. Sabe-se que as componentes de baixa oscilação do sinal de voz (formantes) apresentam correspondência nos baixos *quefreny* do cepstrum, enquanto os harmônicos são evidenciados nos elevados *quefreny*. Como o ruído proveniente de ar turbulento nas cordas vocais tem banda larga, pode-se empregar um processo adaptativo de *liftering* para remover os picos em *quefreny* múltiplos da frequência fundamental, mantendo-se apenas a informação da estrutura formante e o ruído de turbulência. A medida acústica é então definida por

$$HNRCeps_{[f_i, f_h]} = 10 \cdot \log_{10} \left[\frac{\sum_{\omega=f_i}^{f_h} S^2(\omega)}{f_h - f_i + 1} \right] - 10 \cdot \log_{10} \left[\frac{\sum_{\omega=f_i}^{f_h} N^2(\omega)}{f_h - f_i + 1} \right] \quad (65)$$

onde f_i e f_h correspondem à frequência digital superior e inferior da banda espectral de análise,

$S(\omega)$ é a curva de densidade de potência espectral do trecho analisado do sinal de voz,

$N(\omega)$ é a curva de densidade de potência espectral do ruído existente do trecho analisado.

A estimativa do ruído no domínio cepstral é desenvolvida da seguinte forma. Determina-se o cepstrum real ($C(\tau)$) em uma janela com comprimento de 4096 amostras, de onde estima-se o período da frequência fundamental pelo método de Noll (1967).

Filtra-se adaptativamente este cepstrum real, anulando todas as magnitudes na região de existência dos picos cepstrais múltiplos da frequência fundamental. Mais especificamente obtém-se

$$C_{liftering}(\tau) = \begin{cases} 0, & n \cdot T - 0,5 \cdot b \leq \tau \leq n \cdot T + 0,5 \cdot b \\ C(\tau) & c.c. \end{cases} \quad (66)$$

onde n corresponde ao n -ésimo pico cepstral analisado ($n > 1$);

b é a largura de banda dos picos cepstrais;

T é o período da frequência fundamental.

A largura de banda é definida em cada pico periódico do cepstrum real nos pontos de alteração do sinal da derivada primeira desta curva, tanto à esquerda quanto à direita deste pico. Determina-se primeiramente uma aproximação da densidade de potência espectral do ruído através de

$$N_{ap}(\omega) = \Im\{C_{liftering}(\tau)\} \quad (67)$$

É necessário realizar uma compensação da linha de base do espectro do ruído estimado pois a estrutura harmônica do sinal de voz deve repousar sobre a abscissa (0 dB) na curva de densidade de potência espectral. Assim, a compensação pode ser definida por

$$Bd(\omega) = \left| \min[S(\omega) - N_{ap}(\omega)] \right| \quad (68)$$

Assim, a estimativa final da densidade espectral do ruído é dada por

$$N(\omega) = N_{ap}(\omega) - Bd(\omega) \quad (69)$$

A medida apresentada na eq. (69) é realizada sobre segmentos de 4096 amostras, à cada 25,6 ms, produzindo-se uma grande quantidade de amostras que são avaliadas através dos métodos estatísticos da seção 3.2.9. Os intervalos de frequência onde os cálculos foram realizados são 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz e 60 à 2000 Hz.

3.3.11 Primeiro e segundo harmônicos

As medidas de componentes harmônicas foram também utilizadas pois vozes soprosas apresentam modificações bem claras de magnitude nestas frequências. Para sua determinação, para determiná-los empregou-se um algoritmo de estimação de frequência fundamental por cepstrum (Noll, 1967).

A partir desta medida, realiza-se uma busca centrada em ambos os picos harmônicos (frequências múltiplas de F_0) na curva de densidade de potência espectral de 4096 pontos, sendo seus valores fornecidos em dB. A estimativa é realizada à cada 25,6 ms e sobre os valores obtidos aplica-se as avaliações estatísticas apresentadas na seção 3.2.9.

3.3.12 CPP

A amplitude do pico cepstral pode ser útil na avaliação de vozes soprosas. A medida *CPP* (*Cepstral peak prominence*) apresentada por Hillenbrand *et al* (1994) corresponde à diferença entre a amplitude do pico cepstral e o valor correspondente na regressão linear do trecho analisado (este trecho inicia-se a partir de 1 ms da curva de cepstrum real).

Estas estimativas são realizadas sobre janelas do sinal de voz de comprimento igual à 4096 amostras, com deslocamento de 25,6 ms.

3.3.13 P/A

Esta medida acústica corresponde à razão entre a máxima amplitude no segmento de voz analisado pela amplitude média da retificação em onda completa deste mesmo trecho. Esta medida (*Peak to amplitude*) foi sugerida por Hillenbrand *et al* (1994) para ser aplicada entre segmentos sucessivos de 10 msec. Neste trabalho, considerou-se uma janela de 4096 pontos, com deslocamento de 25,6 ms. Ao final das análises sobre todos os trechos do sinal de voz, são aplicadas as análises estatísticas da seção 3.2.9.

3.3.14 RPK

Baseado no conceito da autocorrelação, Hillenbrand *et al* (1994) sugerem a medida RPK (*Pearson r at autocorrelation peak*). O método de cálculo inicia pela estimação do período de frequência fundamental, obtido no trabalho através do método de cepstrum (Noll, 1967).

Utiliza-se então o trecho do sinal de voz analisado e sua versão deslocada um período da fundamental no cálculo do coeficiente de Pearson, ou seja,

$$r_p = \frac{\sum_{i=1}^{N-T} \{(s_i - \bar{s})(s_{i+T} - \bar{s}_T)\}}{\sqrt{\sum_{i=1}^{N-T} (s_i - \bar{s})^2} \sqrt{\sum_{i=1}^{N-T} (s_{i+T} - \bar{s}_T)^2}} \quad (70)$$

onde T corresponde ao período da fundamental estimado,

$\bar{s}$ e $\bar{s}_T$ correspondem respectivamente à média dos trechos analisado e deslocado um período de F_0 .

Os trechos analisados contém 4096 amostras, com deslocamento temporal de 25,6 ms. Aplica-se ao final do cálculo de todos os trechos disponíveis no sinal de voz, as análises estatísticas da seção 3.2.9, para avaliação do comportamento da distribuição dos valores obtidos.

3.4 Medidas no Sinal Residual

A terceira técnica para avaliação do trato vocal é através da análise do sinal residual (Davis, 1979). O fundamento desta metodologia para estimação do sinal glotal através do modelo inverso do trato vocal: assim, aplica-se o sinal de voz e obtém-se o sinal proveniente da glote.

O sinal residual, apesar de não ser o sinal glotal, corresponde ao resultado de uma filtragem inversa sobre o sinal de voz, onde o filtro construído é determinado considerando o modelo paramétrico do trato vocal. Este sinal corresponde ao sinal de excitação para produção de voz sintetizada.

Pesquisadores (Koike e Markel, 1975, Davis, 1979; Prosek *et al*, 1987) verificaram que características patológicas podem ser determinadas através do sinal residual, pois conforme a disfonia e sua extensão, são produzidas alterações temporais e espectrais neste sinal. Assim, desenvolveram um conjunto de medidas acústicas que permitam relacionar a voz com as disfonias.

Esta seção apresenta uma adaptação da técnica apresentada por esses pesquisadores, baseada na obtenção do sinal residual através de algoritmos adaptativos de modo a suportar as variações no tempo (mesmo que pequenas) do trato vocal.

3.4.1 Estimação Paramétrica na Análise de Voz

A transcrição do trato vocal em equações, realizada no capítulo 2, fundamenta a análise por estimativas paramétricas do sinal de voz. Assim, um modelo linear para a produção da fala é dado por:

$$H(z) = R(z) \cdot V(z) \cdot G(z) \cdot E(z) \quad (71)$$

onde estão representados a excitação produzida pelos pulmões (principalmente), por $E(z)$; a vocalização na laringe e cordas vocais, por $G(z)$, e assim, gerando o pulso glotal; a modulação das componente harmônicas do sinal glotal e assim, caracterização da estrutura formante pelo trato supraglotal, por $V(z)$; e os efeitos da radiação dos lábios (e outras perdas associadas à influências térmicas, por rugosidade e alteração das paredes do trato vocal) por $R(z)$.

Este modelo é composto então tanto de zeros quanto de pólos, uma vez que a influência dos zeros capacita a reprodução de sinais vocálicos nasalados (pelo acoplamento do trato nasal através da movimentação do véu palatino). Desta feita, seria necessário utilizar-se de modelos ARMA para descrever corretamente este sistema.

Entretanto, considerando que um modelo AR com um número infinito de pólos pode ser empregado para estimar espectralmente qualquer sistema, pelo teorema de Kolmogorov, (na prática, isto é impossível, pois os requisitos computacionais e a quantidade de amostras da série temporal são irreais), alguns pólos adicionais a um modelo AR permite a este representar o sistema de produção de

fala, corrigindo os erros advindos da falta de zeros (responsáveis pela atenuação as componentes espectrais do sinal de voz).

Assim, obtém-se um modelo AR a ser estimado que representa o espectro densidade de potências para um trecho de amostras, considerando-as um processo aleatório WSS:

$$H(z) = S(z) = \frac{\sigma}{\sum_{i=0}^P a_i z^{-i}} E(z) \quad (72)$$

onde P expressa a quantidade de pólos do sistema. Considerando que no modelo inicial ter-se-ia K formantes com frequências abaixo de 5000 Hz e comprimento de banda geralmente menor que 150 Hz, utiliza-se as três primeiras formantes, que são mais energéticas no espectro densidade de potência. Assim, P é expresso por:

$$P = 2K + m \quad (73)$$

uma vez que os pólos correspondentes aos formantes são conjugados, e o acréscimo de m pólos para correção dos erros devido à falta de zeros ($0 < m < 7$).

Na obtenção dos parâmetros a_i e σ , admite-se que o sinal de voz $s(n)$ é um processo AR com P pólos à serem determinados, e desta forma, o seu preditor é obtido por uma combinação linear de P amostras passadas:

$$\hat{s}[n] = -\sum_{i=1}^P a_i s[n-i] \quad (74)$$

Deseja-se então obter a_i 's que minimizem o erro médio quadrático entre a amostra predita e a amostra real, ou seja:

$$\rho = E[|s[n] - \hat{s}[n]|^2] \quad (75)$$

Através do princípio de ortogonalidade, em que o erro de estimativa é ortogonal à estimativa, obtém-se um sistema de equações não lineares, ou seja:

$$\begin{aligned} r_{ss}[k] &= -\sum_{i=1}^P a_i r_{ss}[k-i] & k = 1, 2, \dots, P \\ \sigma^2 &= \rho_{min} = r_{ss}[0] + \sum_{i=1}^P a_i r_{ss}[-i] \end{aligned} \quad (76)$$

onde $r_{ss}[k]$ é a autocorrelação do sinal $s(n)$ considerando um deslocamento de k amostras.

Estas são as equações de Yule-Walker para processos AR, ou equações de Wiener-Hopf, conforme a abordagem que se dê ao problema de predição linear. Diversos algoritmos são encontrados na literatura para a resolução deste problema, como o algoritmo de Levinson (normalmente empregado nas análises acústicas por sua eficiência na determinação dos coeficientes,

e pela recursividade que permite a estimação dos P parâmetros a partir dos $P-1$ prévios), método da covariância e covariância modificada, método de Burg, entre outros.

A caracterização deste modelo paramétrico para estimação do sistema de produção de fala é importante para a avaliação acústica da voz se o objetivo é avaliar o sinal glotal e desta forma, o comportamento (normal ou patológico) das estruturas da laringe. Retornando à eq. (72), pode-se descrever-la de forma a obter o modelo de excitação, ou seja:

$$E(z) = \frac{1}{\sigma} \sum_{i=0}^P a_i z^{-i} S(z) \quad (77)$$

que transcrito no domínio temporal, obtêm-se o gerador de pulsos de excitação, ou seja:

$$e(n) = \frac{1}{\sigma} \sum_{i=0}^P a_i s(n-i) \quad (78)$$

Este novo formato é conhecido por filtro preditor de erro, onde o sinal estimado de excitação é agora conhecido por erro de estimação. Considerando um estimador polarizado da função de autocorrelação r_{ss} na resolução das equações de Yule-Walker, uma das propriedades do filtro é a sua mínima fase, que garante a estabilidade de $e(n)$ uma vez que os zeros do modelo AR localizam-se dentro do círculo unitário, assegurando que o processo $s(n)$ é estacionário em sentido amplo (WSS).

O erro de predição (ou no caso dos sinais de voz, pulsos de estimação) caracteriza-se por aproximar-se muito de um ruído branco, devido a capacidade de maximização do “spectral flatness”. Uma vez que o sinal de excitação do modelo de produção de fala compõem-se de pulsos espaçados, estes serão obtidos na saída do filtro, com presença de ruído nos intervalos entre os pulsos.

O termo inverso advém do fato de estimar-se o espectro densidade de potência do trato supraglotal (com alta resolução) por meio de um modelo AR, e através da filtragem do sinal de voz usando o sistema inverso ao estimado, obtendo-se então os resíduos, ou seja, o sinal de excitação do sistema linear invariante no tempo.

É importante caracterizar este sinal de excitação, que é diferente do sinal glotal real. Isto advém das simplificações realizadas quando da caracterização do trato vocal, onde o efeito causado pelo subsistema que representa a laringe e cordas vocais produzirá o sinal glotal. Na determinação dos parâmetros do modelo AR, tanto $G(z)$, $R(z)$ e $V(z)$ estão aglutinadas, caracterizando o efeito conjunto destas estruturas na produção da voz.

A técnica de Davis (1979) baseia-se na modelagem apresentado nesta seção. Empregando o algoritmo de Levinson sobre o trecho mais estável do sinal de voz, determina-se um modelo AR, que aplicado sob a forma de filtro inverso fornece o sinal residual, ou de excitação o modelo paramétrico do trato vocal. Maiores detalhes deste trabalho são apresentados na seção 2.7.

3.4.2 Filtros Adaptativos

Para se estimar o sinal de excitação do modelo apresentado, algumas considerações devem ser feitas. A principal é a suposição de estacionariedade do sinal, uma vez que a estimativa paramétrica estimará um sistema linear invariante no tempo e desta forma exige o conhecimento de informações estatísticas de primeira e segunda ordem. Desta forma, é necessário que se desloque a janela temporal para avaliar amostras supostamente invariantes.

Apesar das análises serem realizadas sobre fonemas vocálicos sustentados, a natureza patológica das vozes de pacientes pode exigir técnicas de estimação residual que considerem a alteração na estrutura formante. Desta forma, o emprego de algoritmos adaptativos torna-se interessante.

Avaliando esta metodologia de cálculo de Davis (1979), percebe-se que os pontos críticos estão na exigência de estacionariedade do sinal pela definição matemática da estimativa do filtro autorregressivo (Deller Jr., Proakis e Hansen, 1993), impondo assim um trato vocal invariante no tempo por alguns milissegundos. Assim, movimentação da língua, por exemplo, mesmo que durante a fala sustentada, pode refletir no sinal residual como alguma disfonia ocorrendo ao nível glotal.

O segundo ponto importante é a definição do trecho inicial para estimação, ou seja, determinação do trecho mais estável do sinal de voz. As características de fluxo de ar iniciais combinadas com a constrição da laringe (“ataque vocal”) são diferentes na população (normal e patológica), exigindo assim que o início da estimativa seja feita de acordo com a pessoa.

São apresentados algoritmos adaptativos que consideram a não estacionariedade do sinal e estimam o sinal residual, além de medidas acústicas que representem este novo sinal, para posterior emprego em discriminador estatístico ou neural.

3.4.3 Estimativa usando Gradiente Descendente

A primeira abordagem neste sentido é através da utilização de filtros causais ótimos em relação ao erro médio quadrático (filtros de Wiener), para a produção do sinal residual que, como já citado, apresenta informações sobre a estrutura glotal. O modelo adotado para estes filtros preditivos é:

$$e(n) = s(n) - \mathbf{a}_n^T \mathbf{s}_{n-1} \quad (79)$$

$$\mathbf{s}_n = [s(n), s(n-1), \dots, s(n-P+1)]$$

Percebe-se que σ da eq. (78) foi ignorado temporariamente uma vez que não se tem conhecimento estatístico sobre o sinal. Considerando inicialmente a abordagem pela regra do gradiente, em que objetiva-se minimizar o erro médio quadrático entre a amostra real e a amostra obtida pelo preditor (operação de filtragem), tem-se a seguinte relação de recursão para atualização dos parâmetros do modelo paramétrico:

$$\mathbf{a}_{n+1} = \mathbf{a}_n + \tilde{\nu}[-\mathbf{g}] \quad (80)$$

onde $\tilde{\nu}$ é a taxa de convergência, para adaptação dos coeficientes do modelo e $[-\mathbf{g}]$ denota o gradiente negativo do erro, $E\{e^2(n)\}$, em relação à $\mathbf{a}_n$. Como não é de conhecimento esta última informação estatística, é necessário estimá-la, ou seja:

$$\hat{\mathbf{g}} = \nabla_{\mathbf{a}_n} \{e^2(n)\} = 2 \cdot e(n) \cdot \nabla_{\mathbf{a}_n} \{e(n)\} \quad (81)$$

Considerando agora a eq. (79) que rege a produção da fala, pode-se estimar o gradiente do erro com relação ao vetor de coeficientes $\mathbf{a}_n$, tem-se que:

$$\nabla_{\mathbf{a}_n} \{e(n)\} = -\mathbf{s}_{n-1} \quad (82)$$

E assim, da eq. (82) tem-se a recursão que adapta os parâmetros do filtro à cada nova amostra que é apresentada:

$$\mathbf{a}_{n+1} = \mathbf{a}_n + 2 \cdot \tilde{\nu} \cdot e(n) \cdot \mathbf{s}_{n-1} \quad (83)$$

Uma vez que se espera que os parâmetros estimados sejam estáveis e sua convergência siga para os valores estável, é necessário que a taxa de convergência esteja no intervalo:

$$0 < \tilde{\nu} < \frac{2}{\lambda_{\max}} \quad (84)$$

onde $\lambda_{\max}$ corresponde ao maior autovalor da matriz de autocorrelação $\mathbf{R}_{ss}$. Como novamente as informações estatísticas não estão disponíveis, utiliza-se o somatório dos autovalores desta matriz, que corresponde à potência média do sinal $s(n)$. Sua estimativa pode ser realizada através de:

$$\sigma_s^2(n) = \beta \cdot \sigma_s^2(n-1) + (1-\beta) \cdot s^2(n) \quad (85)$$

Como β corresponde à uma constante de tempo nesta recursão, e considerando N o número de amostras sobre o qual se trabalhará, tem-se que N é aproximadamente igual ao inverso de $(1-\beta)$. Assim, a recursão final para adaptação dos parâmetros do filtro preditor de erro é dada por:

$$\mathbf{a}_{n+1} = \mathbf{a}_n + \frac{1-\beta}{\sigma_s^2(n)} \cdot e(n) \cdot \mathbf{s}_{n-1} \quad (86)$$

Percebe-se agora que a estimativa da variância aparece como condição básica para convergência do procedimento, mesmo que fosse inicialmente desconsiderada. Se o valor fosse conhecido, no desenvolvimento analítico da expressão recursiva ter-se-ia a equivalência:

$$\sigma_s^2(n) = \sigma \quad (87)$$

Naturalmente então, a cada estimativa do modelo do trato vocal (caracterizado pelos parâmetros a), o sinal residual, caracterizando a excitação do sistema, é dado por $e(n)$.

3.4.4 Estimativa Recursiva por Mínimos Quadrados - RLS

Uma segunda abordagem ao problema de estimação é o emprego da técnica de estimação recursiva por mínimos quadrados. Considerando-se que a obtenção de uma estimativa de $s(n)$ pode ser obtida a partir de uma combinação linear de amostras anteriores, a solução direta para este problema é:

$$\begin{aligned}\hat{\mathbf{a}}_n &= [\mathbf{S}_{n-1}^T \mathbf{S}_{n-1}]^{-1} \cdot [\mathbf{S}_{n-1}^T \mathbf{S}_n] \\ \mathbf{S}_n &= \text{col}[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \\ \mathbf{s}_n &= [s(n-1), s(n-2), \dots, s(n-P-1)]\end{aligned}\tag{88}$$

Entretanto, devido ao não conhecimento das amostras, o emprego da recursividade se faz necessário, para a estimação adaptativa do erro de predição. Considerando-se a autocovariância das amostras passadas denotada por $\mathbf{P}_n$, tem-se que

$$\begin{aligned}\mathbf{P}_n &= [\mathbf{s}_{n-1}^T \mathbf{s}_{n-1}]^{-1} \\ \mathbf{P}_{n+1} &= [\mathbf{s}_n^T \mathbf{s}_n]^{-1} = [\mathbf{P}_n^{-1} + \mathbf{s}_{n+1}^T \mathbf{s}_{n+1}]^{-1}\end{aligned}\tag{89}$$

Empregando o lema de inversão de matrizes, obtém-se então a equação para atualização da matriz de covariância, ou seja:

$$\mathbf{P}_{n+1} = \mathbf{P}_n - \frac{\mathbf{P}_n \mathbf{s}_{n+1} \mathbf{s}_{n+1}^T \mathbf{P}_n}{1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1}}\tag{90}$$

com $\mathbf{P}_0 = \rho \mathbf{I}$, com a constante ρ normalmente elevada (da ordem de 1000) para acelerar a convergência do algoritmo.

Desta forma, retornando à eq. (88), obtém-se a forma recursiva de predição linear para o sinal de excitação do trato vocal, ou seja:

$$\hat{\mathbf{a}}_{n+1} = \hat{\mathbf{a}}_n + \kappa_{n+1} [s(n+1) - \mathbf{s}_{n+1}^T \hat{\mathbf{a}}_n]\tag{91}$$

onde

$$\begin{aligned}\delta_{n+1} &= 1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1} \\ \kappa_{n+1} &= \mathbf{P}_n \mathbf{s}_{n+1} \delta_{n+1}^{-1}\end{aligned}\tag{92}$$

Assim, o sinal de excitação do modelo de produção de fala é obtido à cada interação, através da eq (91), ou seja:

$$e_{n+1} = s(n+1) - \mathbf{s}_{n+1}^T \hat{\mathbf{a}}_n\tag{93}$$

3.4.5 Estimativa por Filtro de Kalman

A abordagem de estimação do sinal residual, pela determinação paramétrica do sistema, através de filtros de Kalman foi desenvolvida considerando a existência de ruído, possivelmente correlacionado, no sinal residual.

Considerando esta possibilidade, poder-se-ia pensar que uma filtragem mais robusta pudesse ser realizada de modo a se obter um sinal de excitação mais realístico, sem efeitos de mascaramento advindos do trato supraglotal (por ser estimado).

O desenvolvimento desta técnica é realizada em duas etapas. Em primeiro lugar deve-se estimar as matrizes de estado do sistema de produção de fala. Isto porque na segunda etapa, será estimado o resíduo propriamente dito através de filtro de Kalman e, a operação deste se dá através do conhecimento destas informações.

A estimativa do espaço de estados foi realizada a partir de um modelo ARX, utilizando o método dos mínimos quadrados. Assim, os coeficientes do sistema abaixo serão determinados:

$$\mathbf{A}(z^{-1})s_i = \mathbf{C}(z^{-1})e_j \quad i = 0, 1, \dots, P-1; j = 0, 1, \dots, N-1 \quad (94)$$

Conhecendo-se amostras passadas, busca-se determinar os coeficientes $\mathbf{A}$ e $\mathbf{C}$, representados por $\mathbf{a}_n$, a equação do erro de estimação é dada por:

$$\begin{aligned} \hat{e}(n) &= s(n) - \mathbf{x}_n \hat{\mathbf{a}}_n \\ \mathbf{x}_n &= [-s(n-1), \dots, -s(n-P-1), e(n-1), \dots, e(n-N-1)] \\ \hat{\mathbf{a}}_n &= [a_1, \dots, a_{P-1}, c_1, \dots, c_{N-1}]^T \end{aligned} \quad (95)$$

Minimizando o quadrado da equação acima, tem-se que os parâmetros do modelo ARX, considerando o conhecimento prévio de $s(n)$ e $e(n)$, são dados por:

$$\begin{aligned} \hat{\mathbf{a}}_n &= [\mathbf{X}_{n-1}^T \mathbf{X}_{n-1}]^{-1} \cdot [\mathbf{X}_{n-1}^T \mathbf{X}_n] \\ \mathbf{X}_n &= \text{col}[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \end{aligned} \quad (96)$$

Para uso destas equações de forma adaptativa (uma vez que o trato vocal é um sistema variante no tempo), e considerando o lema de inversão de matrizes, as equações que regem a identificação paramétrica é:

$$\begin{aligned} \hat{\mathbf{a}}_{n+1} &= \hat{\mathbf{a}}_n + \kappa_{n+1} [s(n+1) - \mathbf{x}_{n+1} \hat{\mathbf{a}}_n] \\ \kappa_{n+1} &= \frac{\mathbf{P}_n \hat{\mathbf{x}}_{n+1}^T}{1 - \hat{\mathbf{x}}_{n+1}^T \mathbf{P}_n \hat{\mathbf{x}}_{n+1}} \\ \mathbf{P}_{n+1} &= [\mathbf{I} - \kappa_{n+1} \hat{\mathbf{x}}_{n+1}] \mathbf{P}_n \\ \hat{\mathbf{x}}_n &= [-s(n-1), \dots, -s(n-P-1), -e(n-1), \dots, -e(n-P-1)] \cdot \hat{\mathbf{C}}(z^{-1}) \end{aligned} \quad (97)$$

com $\mathbf{P}_0 = \rho \mathbf{I}$, com ρ normalmente elevado para acelerar a convergência do algoritmo.

A cada iteração, as matrizes $\mathbf{A}$ e $\mathbf{C}$ são utilizadas no processo de filtragem propriamente dito, através do emprego de filtros de Kalman. Primeiramente é realizada uma transformação da equação à diferença eq. (94), com os parâmetros já estimados, para a representação em espaço de estados canônico, obtendo-se assim o seguinte sistema:

$$\begin{cases} \mathbf{z}_{n+1} = \mathbf{M}_n \mathbf{z}_n + \mathbf{N}_n w(n) \\ s(n) = \mathbf{H}_n \mathbf{z}_n + w(n) \end{cases} \quad (98)$$

onde w_n é um ruído branco com variância unitária.

A partir deste sistema (variante no tempo), determina-se então a melhor estimativa linear para o sinal de voz, baseado na consideração estocástica do sistema eq. (98). Determina-se primeiramente, no filtro de Kalman, a estimativa do espaço de estado à posteriori, através das equações abaixo:

$$\begin{aligned} \hat{\mathbf{z}}_{n+1|n} &= \mathbf{M}_n \hat{\mathbf{z}}_{n|n-1} + \kappa'_n [s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1}] \\ \hat{\mathbf{z}}_{0|-1} &= 0 \\ \kappa'_n &= \frac{\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n}{1 + \mathbf{H}_n \mathbf{P}'_n \mathbf{H}_n^T} \end{aligned} \quad (99)$$

A matriz de covariância $\mathbf{P}'_n$ é obtida, na forma recursiva, através da equação de Riccati, ou seja:

$$\begin{aligned} \mathbf{P}'_{n+1} &= \mathbf{M}_n \mathbf{P}'_n \mathbf{M}_n^T + \mathbf{N}_n \mathbf{N}_n^T - \kappa'_n [\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n] \\ \mathbf{P}'_0 &= \rho \mathbf{I} \end{aligned} \quad (100)$$

A partir destas iterações, a cada nova amostra no instante n , surge que o erro de estimação, que corresponde ao sinal de excitação do trato vocal, ou seja:

$$e'(n) = s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1} \quad (101)$$

Desta forma, tem-se a seqüência residual que espera-se corresponder à melhor estimativa do sinal de excitação do modelo de produção de fala, segundo a abordagem de Kalman.

3.4.6 Considerações sobre Convergência

Diferentemente da técnica convencional de produção do sinal residual, através da suposição de estacionariedade de uma janela de dados sobre a qual se aplica o método da autocorrelação, os métodos adaptativos necessitam de uma quantidade de amostras (tempo) para atingirem estabilidade assintótica dos coeficientes dos filtros.

Considerando-se que as medidas acústicas são calculadas sobre todo o sinal residual, estas sofrem influência direta desta característica dos filtros adaptativos pois as primeiras amostras de

resíduo ainda não condizem com o sinal de excitação do modelo do trato vocal. Assim, medidas errôneas podem ser geradas se o tempo de convergência não for levado em conta.

A primeira iniciativa neste sentido é desconsiderar o conjunto de amostras iniciais do sinal residual correspondente à um tempo mínimo de influência da convergência nas medidas acústicas. Para isto, deve-se levar em conta o padrão de voz que será analisado. Verificou-se que para vozes normais, os algoritmos adaptativos estudados tem um tempo de convergência inferior à um período de frequência fundamental, principalmente quando aplica-se os algoritmos por RLS e filtros de Kalman usando valores elevados para ρ (empregou-se $\rho=1000$).

Entretanto, vozes perceptualmente ásperas ou roucas apresentam grandes variações na frequência fundamental (conforme extensão da patologia) e foi necessário avaliar uma quantidade de amostras residuais que seriam desconsideradas na determinação dos parâmetros acústicos. Concluiu-se que desconsiderando o equivalente à um período de 50 Hz (20 ms) do sinal residual, que corresponde à uma voz extremamente grave é suficiente para os casos patológicos considerados.

Mesmo assim, é necessário empregar uma segunda técnica, conjuntamente à apresentada, para aumentar a robustez dos parâmetros acústicos. Isto porque o aumento do ruído de alta frequência e modificações na estrutura espectral (redução energética a partir do 2º formante) das vozes roucas (carcinomas e pacientes corpectomizados, por exemplo) retardam a convergência dos coeficientes dos filtros, principalmente no algoritmo de gradiente descendente.

É realizada uma estimação para os filtros que empregam o modelo AR (filtros de Wiener, nesta pesquisa), utilizando-se o método de autocorrelação sobre o conjunto das 1024 primeiras amostras do sinal de voz. Utilizando o algoritmo de Levinson na resolução das equações de Yule-Walker, obtém-se um conjunto de coeficientes para emprego nos filtros adaptativos, reduzindo-se assim o tempo de convergência destes. Note que as amostras utilizadas neste cálculo são reutilizadas pelos algoritmos adaptativos.

Apesar da possibilidade de redução da quantidade de amostras residuais descartadas no cálculo das medidas acústicas para 10 ms ou menos, de acordo com a patologia, manteve-se a primeira abordagem sem alteração, pois seria necessário avaliar para cada patologia a quantidade ótima de resíduos para desconsideração.

Para os coeficientes do modelo ARX do filtro de Kalman não foi realizada a sua estimação prévia, motivado pela sua velocidade de convergência, apesar da implementação não ser complexa.

3.4.7 Medidas do Sinal Residual

Considerando agora os três algoritmos implementados, é importante desenvolver um procedimento para se obter informações relevantes sobre as características do sinal de resíduo de estimação, de modo a relacionar este com as patologias, objetivo pelo qual buscou-se uma maior precisão na sua derivação.

A primeira medida considerada é SNR (relação sinal-ruído). Considerando que o sinal residual é o erro de estimativa entre as amostras estimadas para um modelo paramétrico e o sinal real, achou-se conveniente determinar esta medida pois a voz proveniente de algumas patologias, como cistos, pólipos e carcinomas caracterizam-se por soprosidade, e assim, ruído em alguns intervalos de frequência no espectro de potência.

O seu cálculo é feito sobre o trecho mais estável (considerando aspectos de convergência) e corresponde ao valor médio da soma do quadrado das amostras de resíduo.

A segunda medida é EX (Coeficiente de excesso) e corresponde matematicamente à medida de curtose da distribuição das amplitudes do sinal residual. Verifica-se em pessoas normais que a curva é leptocúrtica, enquanto que para padrões de voz soprosos, a distribuição é platicúrtica, pois o padrão de pulsos periódicos existente neste sinal é mais definido em vozes normais do que em patológicas. Este parâmetro é determinado pela razão entre o momento de quarta ordem pelo quadrado do momento de segunda ordem, dos resíduos, ou seja:

$$EX = \frac{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^4}{\left(\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^2 \right)^2} - 3 \quad (102)$$

onde N é a quantidade de resíduos, $\bar{s}$ é o valor médio dos resíduos neste segmento e s_i é o i -ésimo resíduo do trecho mais estável da filtragem inversa.

A determinação deste, e dos próximos parâmetros acústicos baseados em filtragem inversa, é realizada a cada trecho de 4096 amostras, sem sobreposição, para se determinar um valor mediano final com maior correlação entre a disfonia e o sinal de voz.

Para se avaliar o comportamento espectral do sinal de voz, definiu-se medidas para avaliação do nivelamento (*flattness*) do sinal de voz, pois sabe-se que quanto maior a quantidade de ruído, maior é o valor obtido para as medidas de nivelamento espectral, onde os limites teóricos são zero à menos infinito.

Primeiramente, SFR (*"spectral flattness" do ruído*) permite verificar o grau de mascaramento dos harmônicos por ruído (proveniente do fechamento incorreto da glote). É obtido através da razão entre a média geométrica pela média aritmética do sinal residual (em escala logarítmica), ou seja:

$$SFR = 10 \cdot \log \left(\frac{\frac{1}{N} \prod_{i=1}^N s_i}{\frac{1}{N} \sum_{i=1}^N s_i} \right) \quad (103)$$

onde N é a quantidade de resíduos e s_i é o i -ésimo resíduo do trecho mais estável da filtragem inversa.

Semelhante à esta medida, tem-se SFF (*"spectral flatness"* do filtro inverso), para avaliação do mascaramento das formantes por ruído. Sabe-se que soproidade é um fenômeno que acarreta um aumento da energia das componentes não harmônicas da voz e simultânea redução das amplitudes nas frequências formantes. Esta relação de energia é acompanhada por esta medida. Matematicamente é definida pela diferença entre a medida de nivelamento espectral do sinal de voz pela medida SFR.

Vozes ásperas apresentam elevado valor para SFR, pois perturbação na frequência fundamental e harmônicos caracteriza patologias como hipercinesia, avaliando o ruído proporcionado por estas alterações no espectro de frequências. Já vozes soprosas têm valores elevados para SFF pois uma quantidade de ruído elevado, proveniente da glote (adução incorreta), reduz a energia nas frequências formantes com aumento das amplitudes laterais. Juntamente com SNR, estas medidas tem por objetivo determinar a quantidade de ruído no espectro.

A medida **Jitter** mede a perturbação da frequência fundamental. Sabe-se que patologias como carcinomas, laringite crônica e paralisia unilateral, por exemplo, produzem sinais de voz com alta variação ciclo-a-ciclo do período de *pitch*, enquanto que outras como cistos ou nódulos tem oscilações reduzidas. Assim, é importante avaliar esta característica da voz.

Para pessoas normais, o sinal residual consiste de uma série de picos periódicos bem definidos, com baixa amplitude nas amostras laterais. Assim, pode-se determinar o período da frequência fundamental como diferença entre picos sucessivos. A estimativa destes períodos foi realizada de forma síncrona para garantir maior precisão, determinando-se as maiores diferenças absolutas de amplitudes (picos bem definidos), em sucessivas regiões de existência destes picos. O centro destas regiões foi calculado através de uma estimação assíncrona da frequência fundamental, utilizando o método da autocorrelação sobre segmentos de 4096 amostras, sem sobreposição. O intervalo de busca corresponde à $\frac{1}{2}$ período de F_0 estimado. Uma vez determinado precisamente o instante da ocorrência do pico, realiza-se uma interpolação parabólica para garantir precisão de 0,05% (Titze, Horii e Scherer, 1987).

O parâmetro Jitter é então definido como a mediana das perturbações de primeira ordem, ou seja, diferenças entre sucessivos períodos da fundamental. A mediana é empregada para garantir robustez na medição.

Definiu-se a medida **Pico** para avaliar a relação entre as patologias e as estimativas de picos periódicos desenvolvidas para o parâmetro anterior. Considerando a normalização do sinal de voz, espera-se que valores elevados para este parâmetro resultem na discriminação de pessoas normais, e proporcionalmente à diferença absoluta entre os períodos de pico (menor valor para este parâmetro), obtém identificação das patologias.

O cálculo desta medida baseia-se na mediana das máximas diferenças absolutas entre amplitudes, que caracteriza os picos periódicos.

Estas duas medidas acústicas (Jitter e Pico) foram desenvolvidas para realizar a mesma tarefa de APQ e PPQ de Koike (1973), Koike *et al* (1977) e Davis (1979), com menor exigência de processamento e empregando medida estatística robusta à oscilações aleatórias com baixa probabilidade de ocorrência.

A última medida desenvolvida também avalia a periodicidade do sinal residual. Através da determinação da mediana das amplitudes do segundo pico da autocorrelação normalizada do sinal residual (já calculada previamente, na estimação dos períodos de frequência fundamental), a medida *P/A* (*Pico na autocorrelação*) possibilita determinar o nível de perturbação na periodicidade do sinal de voz pois a redução na amplitude do segundo pico da autocorrelação e aumento das amplitudes na sua vizinhança. Patologias de origem soprosa (nódulos) ou rouca (paralisia unilateral) apresentam baixos valores para este parâmetro.

As medidas acústicas baseadas no sinal residual que foram apresentadas devem ser capazes de representar adequadamente o sinal residual para que durante o emprego de redes neurais artificiais para classificação e discriminação das patologias, estas estruturas possam “aprender” adequadamente este tipo de sinal visando reconhecimento do padrão vocálico das disfonias.

As medidas Jitter, Pico e *P/A* informam o nível de perturbação do sinal residual pela importância deste tipo de medida na avaliação de patologias (Baken, 1996). SFF, SFR e SNR por sua vez, avaliam a quantidade de ruído turbulento proveniente do inadequado funcionamento das pregas vocais na atividade de abertura e constrição da glote. O parâmetro EX permite avaliação tanto de perturbação da frequência fundamental quanto de ruído turbulento, pois a distribuição das amostras é afetada por estas duas características patológicas, e o grau de achatamento desta curva é um indício da extensão da disfonia.

3.5 Análise Estatística Empregada

Uma vez determinados os parâmetros acústicos do sinal de voz, é necessário avaliar a capacidade de discriminação das patologias. Para isto, foi empregado o teste estatístico Mann-Withney para determinar a diferença entre as distribuições dos parâmetros para cada par de disфонia (um total de 231 combinações, considerando as 22 patologias disponíveis).

Este procedimento de inferência estatística (através do uso deste teste não-paramétrico) deve-se principalmente ao desconhecimento da distribuição (gaussiana, gama, beta ou Weibul, por exemplo) que rege os valores obtidos para os índices acústicos. Isto encerra o uso do teste *t* de Student como alternativa comumente empregada. Teste não-paramétricos são, por definição, livres de distribuição, o que permite seu emprego nos valores obtidos para as medidas acústicas.

O processo de inferência estatística segue alguns passos, independente do tipo de teste que se realiza, fornecendo uma conclusão direta sobre uma determinada hipótese (pergunta que se faz sobre os dados). Os passos são:

1. Formulação da hipótese nula, a qual deseja-se rejeitar.
2. Definição do critério de teste.
3. Especificação do nível de significância (α).
4. Cálculo do teste sobre os dados.
5. Conclusão por comparação entre valor calculado e valor tabelado.

A hipótese nula corresponde a inexistência de diferença entre duas informações (distribuição, média, variância, por exemplo). Diretamente oposta à esta, obtém-se a hipótese alternativa. Para avaliar as medidas acústicas, utiliza-se a hipótese nula (H_0) de que as amostras deste parâmetro pertencem à mesma distribuição, enquanto que a hipótese alternativa (H_1) indica que ambas as disfonias são diferentes para tal índice.

Para definir o critério de teste é necessário considerar que tipo de informação deseja-se inferir estatisticamente e a região de rejeição. Para avaliar a separabilidade entre os pares de disfonias baseado nas medidas acústicas utilizou-se o teste Mann-Withney pois este considera que duas amostras independentes, uma vez que as estimativas de uma determinada medida acústica do sinal de voz para duas patologias são totalmente distintos. Como o teste estatístico neste estudo é realizado para determinar diferença entre distribuições, utiliza-se a região de dupla cauda para rejeição.

A probabilidade de erro na rejeição da hipótese nula, também chamada de nível de significância, foi definido em 5%. Sua escolha deve-se às patologias apresentarem similaridades nas medidas acústicas, e assim, um reduzido α significaria poucas rejeições no teste entre duas patologias. Não existe um procedimento determinístico para definir o nível de significância, sendo baseado em experiência e no conhecimento sobre a justaposição dos padrões acústicos das disfonias.

O teste Mann-Withney opera a partir da ordem (*ranking*) das amostras do conjunto de dados de duas disfonias analisadas, sendo um dos mais poderosos testes não-paramétricos e a mais útil alternativa para o teste t. O procedimento da avaliação da hipótese é feita através do cálculo do índice U de Mann-Withney (Siegel, 1956).

Primeiramente, ordena-se as amostras do parâmetro estudado (de ambas as patologias confrontadas, com quantidades N_1 e N_2) em um único conjunto numérico de ordem crescente; a seguir, aplica-se o método de graduação (*ranking*) neste, iniciando pelo valor 1, que é atribuído ao primeiro elemento ordenado, até o valor N_1+N_2 ao último elemento ordenado. Efetua-se uma correção (*ties correction*) para as amostras repetidas, atribuindo-se à estas a média de suas graduações.

Para obter a medida U do teste, aplica-se a seguinte expressão:

$$U = N_1 \cdot N_2 + \frac{N_1 \cdot (N_1 + 1)}{2} - R_1 \quad (104)$$

ou

$$U = N_1 \cdot N_2 + \frac{N_2 \cdot (N_2 + 1)}{2} - R_2 \quad (105)$$

onde R_1 e R_2 correspondem ao somatório das graduações para o conjunto de tamanho N_1 e N_2 , respectivamente.

Determinado o valor U, este é confrontado com o valor crítico tabelado correspondente, que define o limite em que duas disfonias são similares ou diferentes para uma medida acústica, de acordo com a quantidade de elementos do maior conjunto (N_1 ou N_2) entre as duas patologias avaliadas. Tabelas são disponibilizadas para quantidades inferiores à 20. Acima deste patamar, emprega-se uma aproximação da distribuição normal (Siegel, 1956), obtendo-se a variável z normalizada:

$$z = \frac{U - \frac{N_1 \cdot N_2}{2}}{\sqrt{\frac{N_1 \cdot N_2 \cdot (N_1 + N_2 + 1)}{12}}} \quad (106)$$

Conforme o nível de significância (no caso, 5%) e a relação de ordem entre os valores de U (calculado e tabelado) do teste, rejeita-se ou não a hipótese nula.

Para exemplificar o procedimento, serão considerados um conjunto de valores obtidos para uma medida acústica sobre 9 vozes digitalizadas, analisando-se a separabilidade entre duas patologias, obtendo-se assim a tab. 3.

Tabela 3 - Exemplo para teste estatístico de Mann-Withney

Patologia A	Patologia B
20	50
21	52
22	53
23	54
51	

Primeiramente é necessário realizar a graduação (*ranking*) das medidas acústicas. Para tanto, obtém-se a tab. 4 descrevendo este processo (ordenação de todos os valores e atribuição de um ordinal).

Tabela 4 - Exemplo de graduação (*ranking*) para teste estatístico

Patologia A	Graduação	Patologia B	Graduação
20	1	50	5
21	2	52	7
22	3	53	8
23	4	54	9
51	6		
$R_1=16$		$R_2=29$	

Note que são ordenados todas as medidas independente do tipo de patologia à que pertence. Após a obtenção do somatório das graduações para cada uma das disfonias, aplica-se a eq. 104 para obtenção do teste estatístico U para os dados. Para o exemplo apresentado, $U=1$.

Pelo teste de hipótese estabelecido, em que deseja-se avaliar se as duas patologias são separáveis, considerando-se esta medida acústica específica, que corresponde à rejeição da hipótese nula, ou seja:

H_0 : patologias são semelhantes

H_1 : patologias são diferentes.

Para um nível de significância de 5%, obteve-se o valor tabelado é igual à 3. Logo, por ser o valor tabelado superior ao valor obtido pelos dados, conclui-se estatisticamente que as patologias são diferentes, pela rejeição de H_0 .

Este exemplo mostra o processo que realiza-se para avaliação da separabilidade entre duas categorias distintas (ou patologias, para o presente estudo) através do teste de Mann-Withney.

3.5.1 Procedimento para avaliação da discriminação

Foram definidas as combinações, dois a dois, de todas as patologias (disponíveis durante a gravação de voz), utilizando os valores obtidos para cada medida acústica realizada. Aplicou-se o teste não-paramétrico citado e definiu-se para cada par de disфония se as distribuições do parâmetro acústico eram diferentes (rejeição de H_0) ou similares (não rejeição da hipótese nula).

Assim, determinou-se para cada índice acústico quantos pares de disfonias eram separáveis, obtendo-se então o **número de discriminações**. Quando uma disfonia é discriminável em 22 combinações, empregando-se a mesma medida acústica, pode-se afirmar que este parâmetro da voz identifica esta disfonia, estatisticamente.

Entretanto, é necessário considerar o grau de separabilidade entre as patologias para cada medida acústica. Para isto, fez-se uso da diferença entre os valores de U (calculado e tabelado) apenas para as situações de rejeição do teste de hipótese. Define-se então o **nível de discriminação** de uma medida acústica em relação às disfonias pela seguinte relação.

$$\text{Nível de discriminação} = \sum_{\forall \text{ par de disfonias}} |U_{\text{calculado}} - U_{\text{tabelado}}| \quad (107)$$

A aplicação do módulo se deve que em alguns casos (quando a quantidade de amostras para a medida acústica de uma das disfonias é superior à 20), a rejeição ocorre quando o $U_{\text{calculado}}$ é inferior ao U_{tabelado} , exigindo assim a utilização do valor absoluto da diferença.

Como última medida de separabilidade entre anormalidades da fala, determinou-se a **média das discriminações** como resultado da razão entre o nível e o número de discriminações.

Nas avaliações das medidas acústicas em relação à identificação das patologias, estas três variáveis são consideradas, priorizando pelo número de diferenciações e nível de separabilidade entre as distribuições, pois é importante que o parâmetro acústico simultaneamente consiga distinguir um grande número de patologias, com maior afastamento estatístico entre as suas distribuições (maior nível de discriminação).

O emprego de teste de hipótese ao invés de discriminadores estatísticos (métodos bayesianos, por exemplo), é a facilidade de implementação (apesar de exigir cálculos exaustivos pois é necessário realizar as combinações, dois a dois, das patologias) e permitir verificar o grau de separabilidade entre cada patologia.

Baseado nas avaliações estatísticas sobre as medidas estatísticas, são escolhidos os parâmetros de maior poder de distinção entre disfonias para serem empregados no classificador neural.

3.6 Redes Neurais Artificiais

Inspirado em modelos biológicos das células nervosas, as redes neurais artificiais apresentam um conjunto de características que conseguem resolver problemas complexos e de natureza não-linear, apesar dos modelos tradicionais não apresentarem plausibilidade física.

O reconhecimento de padrões audiovisuais é uma tarefa que o cérebro humano realiza com extrema competência. Através do trabalho paralelo de diversas células (de processamento simples), pode-se identificar símbolos (imagens, trechos de séries temporais, sons) utilizando apenas uma pequena quantidade de dados mesmo que corrompidos por ruído ou inconsistente dos demais valores.

Baseado na capacidade de identificar padrões, utilizou-se das redes neurais com o objetivo de identificar as patologias do trato vocal baseado em um conjunto de medidas acústicas. Estes parâmetros são escolhidos com base na capacidade estatística de discriminar um maior conjunto de disfonias.

Uma rede neural artificial é definida por um conjunto de componentes, buscando-se ao máximo uma representação conexionista. Segundo Rumelhart e McClelland (1986), esta estrutura de processamento deve conter:

1. Um conjunto de unidades de processamento
2. Um conjunto de estados de ativação
3. Uma função de saída para cada unidade
4. Um padrão de conexão entre as unidades
5. Uma regra de propagação
6. Uma regra de ativação das unidades
7. Uma regra de aprendizagem
8. Um meio onde é empregada

Uma célula, ou nó, da rede neural desempenha um processamento simples onde coleta informações de entrada, aplica um determinado processamento sobre estes que indicará o estado de ativação da mesma. Finalmente, este elemento produz um valor de saída que se propagam para outras células.

Entretanto, é necessário definir uma estrutura de encadeamento destas unidades básicas por onde as informações serão transmitidas. Através do estabelecimento de conexões entre as células (definido-se ponderações - pesos - que serão aplicadas sobre os valores transmitidos entre unidades) define-se o conhecimento propriamente dito da rede neural. Estes pesos definem ligações inibitórias e excitatórias entre as células, que biologicamente correspondem às ligações sinápticas entre neurônios reais.

Para que esta estrutura construída tenha “conhecimento”, é necessário definir um mecanismo para reorganizar as conexões, ou seja, alterar as sinapses entre as células. Utilizando regras de

aprendizado, geralmente baseadas no princípio de Hebb (1949), ou seja, em que conexões ativas são reforçadas enquanto outras devem ser enfraquecidas, de modo que a rede possa responder adequadamente às entradas de acordo com os padrões adquiridos durante seu treinamento.

Na prática, define-se uma rede neural através de um conjunto de nós, divididos em N camadas. A primeira camada é denominada camada de entrada, enquanto que a última chama-se camada de saída. Entre estas estão as camadas escondidas, pois somente recebem e transmitem valores para outras camadas. As ligações, genericamente, são definidas entre todos os nós; de acordo com o tipo de representação neural, são atribuídos valor zero para alguns pesos das conexões, impedindo ligações entre nós de diferentes camadas, realimentação nas células, por exemplo. Estes critérios de ligação variam de acordo com as características que se deseja obter no conhecimento a ser armazenado, além de aspectos matemáticos como estabilidade, mínimos locais e globais.

A operação das redes neurais segue um fluxo bem definido de passos, podendo ser dividido em duas categorias: a fase de aprendizado e a fase de teste. A diferença entre estas duas fases é a atualização dos pesos “sinápticos” conforme um determinado critério matemático (gradiente do erro entre saída desejada e saída obtida, recompensa e penalização, reforço e inibição).

As etapas comuns entre o aprendizado e o teste são:

1. Apresentação de um vetor de dados às unidades da camada de entrada
2. Propagação das saídas de uma camada para a próxima, por toda a rede
3. Atualização dos estados de ativação das células
4. Atualização das saídas das células

Com estas características em mente, diversos tipos de redes neurais foram e são concebidos, para solucionar gargalos como evitar que na superfície da função de erro de saída atinja-se mínimos locais, aumentar a quantidade de padrões que a rede neural consiga aprender, melhorar a capacidade de generalização e abstração, acelerar o tempo de treinamento.

A rede neural utilizada neste trabalho baseia-se no modelo proposto por Widrow e Hoff (1960), onde utiliza-se a diferença entre as saídas desejada e obtida na modificação dos pesos, sendo do tipo *feedforward*, ou seja, apresenta-se um vetor de entrada, que é propagado para as demais camadas da rede, até atingir a última, onde obtém-se finalmente um vetor de saída. Considerando finalmente o emprego de múltiplas camadas, adota-se o modelo Madaline, que apresenta as seguintes características:

A regra de propagação de valores entre as camadas é definida por:

$$NET_j = \sum_{i=1}^{N_i} \omega_{ij} \cdot o_i \quad (108)$$

onde N_i é o número de nós da camada anterior

ω_{ij} é o peso na conexão entre a célula i da camada anterior e a célula j da camada atual
 o_i é a entrada no j -ésimo célula.

Para cada célula j da camada atual, calcula-se a soma ponderada de todas as suas entradas, sendo que este resultado deve determinar a ativação ou não da célula, provocando assim a emissão de um valor na saída da mesma, obtendo-se então o vetor de saída da rede, ou seja:

$$y_j = f(NE T_j) \quad (109)$$

Diversas funções de saídas são propostas na literatura, geralmente sendo empregada as sigmoidais logísticas (o segundo termo se deve à variação entre zero e a unidade), pois posteriormente será necessário determinar a derivada da função de saída. Pela continuidade destas funções, evita-se problemas com pontos de singularidade. Propagando-se os valores de saída entre as camadas, obtém-se ao final a saída da rede neural.

É necessário então comparar o vetor de saída obtido (y) com os resultados desejados (t) que estão associados ao vetor de entrada. Baseado na diferença entre estes, deve-se atualizar os pesos de forma que a rede assimile novas características, ou seja, aprenda. Utilizando o critério de aprendizado por *backpropagation*, deseja-se definir um conjunto de pesos, $\{\omega_{ij}\}$, que minimize a seguinte função de erro:

$$E = \frac{1}{2} \sum_j (t_j - y_j)^2 \quad (110)$$

A variação deste erro em função dos pesos é definida por:

$$\frac{\partial E}{\partial \omega_{ij}} = - \frac{\partial E}{\partial NE T_j} \cdot y_j \quad (111)$$

Para as conexões entre a camada de saída e a antecessora, a variação entre o valor médio quadrático do erro e o valor total propagado para a j -ésima célula é dada por:

$$\frac{\partial E}{\partial NE T_j} = -f'(NE T_j) \cdot (t_j - y_j) \quad (112)$$

Enquanto que para as demais conexões:

$$\frac{\partial E}{\partial NE T_j} = -f'(NE T_j) \cdot \sum_k \left(\frac{\partial E}{\partial NE T_k} \cdot \omega_{jk} \right) \quad (113)$$

onde os elementos de índice k são provenientes da camada precedente; é esta característica que fornece o nome da técnica de aprendizado.

Assim, as atualizações dos pesos podem ser definidas por:

$$\omega_{ij}(t+1) = \omega_{ij}(t) + \eta \cdot \frac{\partial E}{\partial NE T_j} \cdot y_j \quad (114)$$

onde η é a taxa de aprendizado.

O processo de treinamento é desenvolvido realiza-se sobre um conjunto de medidas de modo que o erro médio por padrão (apresentado à rede neural) seja inferior à um valor preestabelecido. É importante que os pares entrada/saída de treinamento sejam apresentados de forma aleatória, para evitar que esta assimile características de ordem entre os padrões.

Outra necessidade é inicializar adequadamente os pesos. Normalmente opta-se por utilizar valores aleatórios (entre 0 e 1), mas pode-se utilizar pesos já empregados anteriormente, para refinamento de um processo anterior de aprendizado.

Assim, o algoritmo final que define a rede neural do tipo feedforward, com treinamento em backpropagation tem os seguintes passos:

1. Iniciação dos pesos
2. Apresentação do p -ésimo padrão à ser treinado (vetores de entrada e saída desejada)
3. Propaga-se a entrada até a última camada, obtendo-se um vetor de saída (etapa feedforward)
4. Adapta-se os pesos a partir da última camada (etapa backpropagation).
5. Verifica-se o critério de parada.
- 5.1. Se este critério não é atingido, define-se novo padrão ($p=p+1$) e reexecuta sequência a partir do passo 2.

Este procedimento permite que a rede abstraia as características dos padrões apresentados, representando-os na forma de pesos (sinapses). Ainda mais, como o processamento é realizado de forma paralela entre as células, a informação adquirida fica distribuída, caracterizando uma tolerância à falhas.

Outra característica importante é que a rede neural está realizando uma “clusterização” ou agrupamento das informações semelhantes, de forma não-linear (conforme o número de camadas). Isto é importante pois as medidas de voz são extraídas de um sinal proveniente de um sistema de produção de fala complexo e altamente não-linear. Assim, a classificação dos padrões patológicos através desta abordagem torna-se mais preciso.

3.6.1 Considerações práticas

A utilização de redes neurais para classificação e discriminação dos padrões acústicos, e indiretamente, das disfonias tem algumas peculiaridades que devem ser consideradas.

Um conjunto de medidas acústicas, selecionadas através de análise estatística, formam o vetor de entrada que é apresentado à rede neural. Como a quantidade de medidas extraídas do sinal de voz é extremamente elevado, foram realizadas avaliações de quais conjuntos de índices acústicos fornecem melhor resultado na discriminação de disfonias.

Entretanto, deve-se considerar a diferença entre intervalos de variação para cada parâmetro acústico. Algumas medidas têm magnitudes extremamente baixas, como as medidas de perturbação em contraposição à outras como avaliação de curtose. Para solucionar esta dificuldade determinou-se

para cada medida acústica o maior valor absoluto, e este valor (com um incremento de 2%) utilizado para escalonar as estimativas obtidas para aquele parâmetro, ou seja:

$$\text{novo valor}_i = \frac{\text{valor}_i}{1,02 \cdot \max\{\text{valor}_i\}} \quad (115)$$

Assim, o vetor de entrada efetivo da rede neural contém apenas valores normalizados entre -1 e +1, o que auxilia na manutenção da estabilidade do algoritmo de treinamento. O percentual de excesso (2%) empregado opera como margem de segurança para que durante a fase de teste, evite-se a exigência de que a rede realize algum tipo de extrapolação, atividade em que não obter-se-ia resultados adequados.

O modo de representação das disfonias deve propiciar a identificação correta das patologias, além de permitir que a rede forneça a probabilidade de que a voz analisada corresponda à uma específica anormalidade da laringe. Deste modo, definiu-se que cada nó de saída da rede neural fornece a probabilidade de que as medidas acústicas da entrada correspondam à um tipo de voz específico. Tem-se então um conjunto de 22 células na camada de saída (incluindo as 21 disfonias e o padrão de controle, ou voz normal).

O suporte para que se possa considerar as saídas da rede neural utilizada como probabilidade à posteriori de correlação entre as medidas acústicas e as disfonias é baseado no princípio de entropia relativa (Haykin, 1994). Apesar de nenhuma mudança ser necessária no algoritmo de treinamento apresentado (backpropagation), apresenta-se a teoria

Considerando a probabilidade da i -ésima célula da camada de saída para um padrão de treinamento α denotada por $P_{i|\alpha}$, enquanto que o valor real de probabilidade é dado por $Q_{i|\alpha}$. Pode-se definir a entropia relativa da rede neural através da expressão:

$$H_{P||Q} = \sum_{\alpha} P_{\alpha} \sum_i \left[Q_{i|\alpha} \cdot \ln \left(\frac{Q_{i|\alpha}}{P_{i|\alpha}} \right) + (1 - Q_{i|\alpha}) \cdot \ln \left(\frac{1 - Q_{i|\alpha}}{1 - P_{i|\alpha}} \right) \right] \quad (116)$$

onde P_{α} é a probabilidade a priori da ocorrência da patologia para o padrão apresentado.

O objetivo é minimizar a entropia relativa (ou critério de Kullback-Leibler) de tal forma que $P_{i|\alpha}$ seja o mais próximo de $Q_{i|\alpha}$, ou seja, obter-se um valor de $H_{P||Q}=0$. Aplicando-se o método do gradiente descendente, obtém-se a eq. (114), que define o método de adaptação dos pesos nas conexões da rede neural.

A saída da rede neural, então, compõem-se de um vetor de 22 linhas em que cada posição está associada à uma patologia. Para o treinamento, apresenta-se a tab. 5 onde se relaciona as patologias (além do padrão de voz normal) com os valores esperados nos nós de saída.

Tabela 5 - Relação entre os tipos de voz e o vetor de saída esperado

Patologia	Vetor de saída correspondente
Carcinoma	{1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
Cisto	{0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
Corpectomia Parcial	{0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
Corpectomia Total	{0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
Disartria	{0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
Disfonia Funcional	{0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
Edema	{0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0}
Edema de Reinke	{0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0}
Hiperemia	{0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0}
Espessamento de Pregas Vocais	{0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0}
Fenda Fusiforme	{0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0}
Granuloma	{0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0}
Hipercinesia	{0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0}
Hipocinesia	{0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0}
Laringite Crônica	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0}
Lesão das Cordas Vocais	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0}
Leucoplasia	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0}
Nódulo	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0}
Normal	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0}
Outras Disfonias	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}
Paralisia Unilateral	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}
Pólipo	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}

Pode-se perceber que cada disfonia apresenta somente um valor unitário dentro do vetor. Alguns sinais de voz entretanto foram diagnosticados com mais de uma anormalidade. Assim, para estes casos, os vetores de saída esperados eram providos de mais de um valor unitário, correspondentes às patologias que produziram aquele tipo de fala.

Para o treinamento, foram apresentados todos os 229 padrões de entrada e saída de forma aleatória (um ciclo completo), pois caso contrário a rede neural armazenaria nas ligações sinápticas a relação sequencial estes padrões. O critério de parada é que o erro médio quadrático por padrão fosse inferior à um erro de tolerância. Posteriormente será especificado este valor.

Para permitir que a rede neural não encontre mínimos locais muito rapidamente, foi aplicado ruído gaussiano com média zero, apenas alterando-se a variância, proporcionalmente ao número de ciclos executados, e principalmente, à magnitude das medidas acústicas.

Além de permitir que a convergência encontre mínimos globais, a aplicação de uma característica aleatória à entrada possibilita que os agrupamentos patológicos (clusters) tenham uma extensão maior, elevando a capacidade de generalização da rede neural, ao mesmo tempo, reduzir a influência o problema da baixa quantidade de padrões por patologia na discriminação.

No controle da convergência, variou-se a taxa de aprendizado manualmente. Considerou-se o emprego do parâmetro de momento (α) como fator de aceleração na busca dos pesos ótimos que minimizassem o erro médio quadrático. Entretanto, resultados práticos mostravam que de acordo

com o valor empregado, o treinamento tornava-se instável. Utilizando valores baixos (da ordem de 0,01), encontrava-se muitas regiões planas que tornavam ainda mais lento o processo de aprendizado. Assim, ignorou-se o uso do termo de momento.

Estruturalmente a maioria das redes neurais consideradas possuíam, além das camadas de entrada e saída, duas camadas escondidas, cuja quantidade de células variou conforme o treinamento realizado. Constatou-se em testes preliminares que as características de não-linearidade e fundamentalmente a proximidade estatística das distribuições dos dados exigiam mais de uma camada intermediária. Geometricamente a separabilidade das disfonias somente é obtida pelo emprego de curvas côncavas. Somente as redes neurais em que se empregou um grande número de medidas acústicas (vetor de entrada), que analisam as duas principais características encontradas em patologias, permitiram o uso de somente uma camada intermediária.

O número de nós por camada intermediária mostrou-se importante pois uma pequena quantidade destes não é capaz de discriminar adequadamente algumas disfonias (principalmente aquelas que são muito semelhantes acusticamente). Simultaneamente, uma grande quantidade de células rapidamente convergia, alcançando facilmente o critério de parada estabelecido. Entretanto, durante a fase de teste, a quantidade de erros era elevada pois acredita-se que a rede acabava por criar agrupamentos de apenas um ou dois exemplos por patologia, resultando em um baixo poder de generalização. Assim, várias configurações foram testadas, sendo que os resultados são apresentados no capítulo 4.

4. Resultados

Neste capítulo são apresentados os resultados da discriminação patológica, baseado no emprego do teste de hipótese formulado na seção 3.5. A escolha dos índices acústicos com melhor potencial na identificação de disfonias para o sistema de classificação empregando redes neurais artificiais motiva tal abordagem, evitando-se assim a realização de um conjunto inviável de combinações possíveis das medidas acústicas.

Posteriormente, neste mesmo capítulo, apresenta-se os resultados da classificação empregando a técnica de redes neurais, sendo que os testes são realizados sobre um conjunto de vozes onde não se realizou o tratamento estatístico descrito na seção 3.5 e nem o treinamento neural.

Algumas conclusões já são apresentadas durante este capítulo, devido a grande quantidade de informação que foi produzida, facilitando a compreensão dos resultados obtidos.

4.1 Medidas de Perturbação

Para avaliar o potencial discriminador das medidas de perturbação, foi empregado o teste Mann-Withney, anteriormente apresentado, sobre todos os pares de disfonias (um total de 231 pares) para estes índices acústicos.

Os parâmetros de perturbação utilizados foram avaliados para os fonemas gravados (/a/, /e/ e /i/) tanto para perturbações em amplitude e período da frequência fundamental. Considerando primeiramente as medidas de *jitter*, obteve-se um conjunto de tabelas para verificação da capacidade de distinguir disfonias através destas medidas acústicas.

4.1.1 Perturbações de Período da Frequência Fundamental

Devido ao grande número de índices acústicos obtidos para as perturbações (68), estas foram agrupadas segundo o subsinal em que eram aplicadas. Isto também permite avaliar a utilidade dos sinais de perturbação na discriminação.

Tabela 6 - Análise estatística das perturbações de ordem zero (*jitter*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. Clássica	97	33,8	0,348453608
A regr. Robusto	122	82,8097	0,678768033
Assimetria	87	14,00968	0,161030805
B regr. Clássica	96	36,8	0,383333333
B regr. Robusto	102	49,8	0,488235294
Curtose	98	60,10968	0,613364082
Desvio absoluto	116	131,07479	1,129955086
Desvio regr. clássica	124	124,233918	1,001886435
Desvio padrão	118	137,73395	1,167236864
Desvio regr. robusta	126	124,89125	0,991200397
Mediana	111	41,0097	0,369456757
Média	99	1	0,01010101

Utilizando as medidas de perturbação de ordem zero, pode-se perceber que algumas medidas apresentam baixo rendimento, principalmente média, coeficientes A e B da regressão linear e assimetria da distribuição. Isto deve-se ao fato de que a não aplicação da diferença absoluta e sim a diferença simples quando estes índices são determinados está influenciando esta baixa capacidade de discriminação. Pode-se perceber que as medidas de desvio (que empregam valores absolutos ou quadráticos no sinal de perturbação) apresentam resultados superiores. Deve-se considerar simultaneamente o número, nível e média de discriminação pois o coeficiente A por regressão robusta, apesar de ser capaz de separar 122 pares de disfonias, não têm um baixo nível de discriminação se comparado ao desvio padrão das perturbações de ordem zero.

Constata-se que os parâmetros que obtiveram melhor resultado na distinção entre disfonias foram as medidas de dispersão, ou seja, desvio da regressão robusta, desvio da regressão clássica, desvio padrão e desvio absoluto. Isto corrobora o trabalho de Pinto e Titze (1990), onde afirma que medidas de desvio (absoluto e padrão) são úteis na avaliação do trato vocal.

Tabela 7 - Análise estatística das perturbações de ordem 1 (*jitter*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média de Discrim.
A regr. clássica	107	32,93816	0,307833271
A regr. Robusto	111	34,80968	0,313600721
Assimetria	63	7,34267	0,116550317
B regr. Clássica	109	30,93816	0,28383633
B regr. Robusto	110	42,63928	0,387629818
Curtose	107	32,42958	0,303080187
Desvio absoluto	129	97,72239	0,757537907
Desvio regr. clássica	120	110,56165	0,921347083
Desvio padrão	127	94,7224	0,745845669
Desvio regr. robusta	119	110,56165	0,929089496
Mediana	125	20,74227	0,16593816
Média	82	1,82958	0,022311951

Considerando a tab. 7, percebe-se uma redução na distância entre as distribuições (conforme o nível e a média das discriminações). Como o sinal de perturbação de primeira ordem corresponde à derivada primeira dos períodos da fundamental, é possível que as oscilações rápidas e aleatórias sejam similares entre as disfonias. Percebe-se novamente das medidas estatísticas de dispersão são as de melhor capacidade de discriminação entre os pares disfônicos, apesar de resultados um pouco inferiores que os obtidos na tab. 6, ou seja, desvio padrão, desvio absoluto, desvio da regressão robusta e da regressão clássica.

Tabela 8 - Análise estatística das perturbações de ordem 2 (*jitter*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	122	19,77185	0,162064344
A regr. robusto	94	6,00968	0,063932766
Assimetria	77	4	0,051948052
B regr. clássica	131	33,57185	0,256273664
B regr. robusto	91	0	0
Curtose	104	35,42958	0,340669038
Desvio absoluto	137	92,72239	0,676805766
Desvio regr. clássica	123	107,55197	0,87440626
Desvio padrão	128	91,72239	0,716581172
Desvio regr. robusta	123	107,55197	0,87440626
Mediana	119	5,82958	0,048988067
Média	112	31,83928	0,284279286

A perturbação de segunda ordem (derivada segunda das medidas de período da frequência fundamental) foram capazes de distinguir um maior número de pares de disfonias, apesar de uma certa proximidade das distribuições dos índices acústicos para tais patologias.

Aparentemente, ocorre uma redução no intervalo de número de discriminações (apesar do parâmetro acústico conseguir separar apenas 77 pares de disfonias), enquanto que o nível de discriminação da maioria das medidas acústicas atingiu patamares mínimos de distância entre as distribuições.

Mantendo as mesmas características dos experimentos anteriores, as medidas de dispersão (desvios absoluto, padrão, de regressão clássica e regressão robusta) mantêm melhor qualidade de separação entre as disfonias considerando esta ordem de perturbação.

Tabela 9 - Análise estatística das perturbações de ordem 3 (*jitter*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	90	21,0097	0,233441111
A regr. robusto	105	22,62958	0,21551981
Assimetria	93	15,00183	0,16131
B regr. clássica	105	25,66886	0,244465333
B regr. robusto	129	39,90861	0,30936907
Curtose	106	39,92958	0,376694151
Desvio absoluto	135	94,72239	0,701647333
Desvio regr. clássica	125	107,55197	0,86041576
Desvio padrão	127	90,72239	0,714349528
Desvio regr. robusta	125	107,55197	0,86041576
Mediana	109	29,8	0,273394495
Média	111	24,57185	0,221368018

Seguindo a tendência das tabelas anteriores, as medidas de perturbação de ordem 3 reduzem os limites do número de discriminações. Juntamente com este, tanto o nível e média de separabilidade também são restringidos.

Mesmo assim, considerando a quantidade de pares discrimináveis e a distância entre as distribuições dos parâmetros acústicos para as disfonias, as medidas de dispersão estatística (desvio absoluto, desvio padrão e desvios de regressão clássica e regressão robusta) apresentam melhor capacidade de identificação de patologias.

Tabela 10 - Análise estatística das perturbações de ordem 4 (*jitter*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. Clássica	111	16,050881	0,144602532
A regr. Robusto	94	0	0
Assimetria	102	13,34267	0,13081049
B regr. Clássica	124	45,35088	0,365732903
B regr. Robusto	94	1	0,010638298
Curtose	108	43,92958	0,40675537
Desvio absoluto	139	90,72239	0,652679065
Desvio regr. clássica	125	108,55197	0,86841576
Desvio padrão	127	91,72239	0,722223543
Desvio regr. robusta	125	108,55197	0,86841576
Mediana	127	11,28874	0,088887717
Média	86	2	0,023255814

Considerando a tab. 10, pode-se perceber que a tendência de redução do intervalo de número de discriminações entre as medidas acústicas é mantida.

Já o nível de separabilidade para a maior parte dos parâmetros do sinal de voz é extremamente baixo, devido a similaridade entre as patologias para a derivada de quarta ordem dos períodos da fundamental. Pode-se perceber que as medidas de desvio da regressão clássica e robusta apresentam similar capacidade de discriminação, apesar dos coeficientes destas estimativas mostrarem desigual distinção entre os pares patológicos. Uma vez que na determinação de desvio,

considera-se o quadrado da diferença entre amostras sucessivas da perturbação, deve ocorrer uma compensação para que as avaliações estatísticas do teste de hipótese sejam tão próximas.

Através destas tabelas, pode-se perceber que as medidas de desvio foram as que melhor identificaram as disfonias, fato confirmado pelos prognósticos de Pinto e Titze (1990).

Tabela 11 - Análise estatística das medidas de perturbação (*jitter*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
DPF	114	11,65916	0,102273333
JF	142	88,55198	0,623605493
JR	140	88,55198	0,632514143
Média móvel (ordem 11)	134	111,91401	0,835179179
Média móvel (ordem 5)	142	97,56165893	0,687053936
PFF	105	51,1224	0,48688
PVI	137	167,19308	1,220387445
Série Temporal	115	57,96167	0,504014522
Variância	118	137,73395	1,167236864

Considerando as medidas da tab. 11, onde se enquadram parâmetros acústicos encontrados na literatura, verifica-se que a pior discriminação fica a cargo de PFF (*perturbation frequency factor*), devido ao limiar escolhido para esta como nível de perturbação normal, ou seja, 0,5 ms. Verifica-se também que a medida direcional de perturbação (DPF) produz valores muito similares entre as disfonias, comprovado pelo nível de discriminação (o mais baixo nesta tabela).

A medida de melhor performance (incluindo número, nível e média de separabilidade) é PVI (*period variation index*), obtida a partir do coeficiente de correlação das estimativas de período da frequência fundamental. As medidas de perturbação que contabilizam os efeitos da elevação ou redução da F0 foram as que apresentaram melhores resultados na distinção entre os pares disfônicos. Assim, tem-se que as medidas PVI, médias móveis de 5 e 11 coeficientes, JR e JF mostram-se úteis para identificação de anormalidades nas pregas vocais.

Utilizando o fonema /e/ na avaliação das vozes. Uma vez que as pregas vocais apresentam uma estabilidade melhor durante a vocalização deste fonema sustentado, os resultados na discriminação têm menor discrepância que os apresentados para /a/.

Tabela 12 - Análise estatística das perturbações de ordem 0 (*jitter*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	94	24,298411	0,258493734
A regr. robusto	93	51,4111	0,552807527
Assimetria	88	3,00967893	0,034200897
B regr. clássica	94	30,64896	0,326052766
B regr. robusto	100	30,84896	0,3084896
Curtose	98	60,1	0,613265306
Desvio absoluto	119	138,95084	1,167654118
Desvio regr. clássica	134	107,400261	0,801494485
Desvio padrão	127	126,95081	0,999612677
Desvio regr. robusta	134	105,12123	0,784486791
Mediana	94	8,95194893	0,095233499
Média	93	13,33926	0,143432903

Comparando estes resultados com os obtidos para o fonema /a/ verifica-se novamente a capacidade das medidas de dispersão em distinguir os pares de disfonias, com um maior número de discriminações, apesar de uma redução no nível de separabilidade, motivada pela maior estabilidade das estruturas de produção da fala durante a geração deste fonema. Isto provoca então uma redução na dispersão dos valores do nível de discriminação, pela proximidade das medidas acústicas para as disfonias.

As outras medidas de avaliação da curva de distribuição (assimetria e curtose) apresentam baixa eficiência na identificação das patologias, com resultados similares para ambos os fonemas.

Tabela 13 - Análise estatística das perturbações de ordem 1 (*jitter*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	95	24,83928	0,261466105
A regr. robusto	105	21,8	0,207619048
Assimetria	102	30,66886	0,300675098
B regr. clássica	110	31,46886	0,286080545
B regr. robusto	108	24,74227	0,229095093
Curtose	90	31,8	0,353333333
Desvio absoluto	121	61,93928	0,511894876
Desvio regr. clássica	120	79,94896	0,666241333
Desvio padrão	122	77,94896	0,638925902
Desvio regr. robusta	120	79,94896	0,666241333
Mediana	114	19,6	0,171929825
Média	103	20,981928	0,203708039

A derivada das estimações de períodos sucessivos da frequência fundamental apresentam resultados inferiores quando comparados aos obtidos quando se considera o fonema /a/. Apesar da quantidade de pares separáveis ser similar para ambos, o nível de discriminação apresenta um limite superior menor que o obtido para /a/. Além disto, a dispersão desta medida estatística (nível de separabilidade) também é reduzido.

Comparativamente às perturbações de ordem zero, constata-se que as distribuições dos valores para os pares discrimináveis tendem a se aproximar uma das outras, devido à similaridade dos padrões de alteração na vibração das pregas vocais.

Tabela 14 - Análise estatística das perturbações de ordem 2 (*jitter*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	99	4,17225	0,042143939
A regr. robusto	129	50,021301	0,387762023
Assimetria	80	3,612692	0,04515865
B regr. clássica	102	9,34267	0,091594804
B regr. robusto	95	0	0
Curtose	97	33,8	0,348453608
Desvio absoluto	124	61,93928	0,499510323
Desvio regr. clássica	120	75,94896	0,632908
Desvio padrão	124	71,93928	0,580155484
Desvio regr. robusta	120	75,94896	0,632908
Mediana	111	39,0097	0,351438739
Média	107	11,91269	0,111333551

Novamente é visível a redução da distância entre as distribuições das amostras de medidas acústicas, devido a diminuição do nível e média de discriminações. Apesar de algumas medidas acústicas mostrarem-se adequadas para a distinção entre pares de disfonias, como o caso do coeficiente A da regressão robusta, notadamente os índices desvio padrão, absoluto, desvio da regressão clássica e da regressão robusta demonstram ser mais eficientes para identificar tipos de vozes.

Os resultados obtidos para as perturbações de segunda ordem não diferem significativamente daqueles obtidos para as perturbações de primeira ordem.

Tabela 15 - Análise estatística das perturbações de ordem 3 (*jitter*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	105	2	0,019047619
A regr. robusto	111	0	0
Assimetria	94	18,0097	0,191592553
B regr. clássica	123	20,91269	0,17002187
B regr. robusto	124	18,77186	0,151385968
Curtose	95	32,8	0,345263158
Desvio absoluto	123	60,93928	0,495441301
Desvio regr. clássica	121	74,94896	0,619412893
Desvio padrão	125	70,93928	0,56751424
Desvio regr. robusta	121	74,94896	0,619412893
Mediana	123	18,12237	0,147336341
Média	107	18,45539	0,17248028

Considerando a tab. 15, pode-se notar que a quantidade de pares discrimináveis aumentou, ao mesmo tempo que as medidas de desvio apresentam uma maior preponderância em relação aos demais parâmetros acústicos, quando se leva em conta o nível e a média de distinções.

Pode-se perceber que as medidas estatísticas, com exceção das avaliações de dispersão, apresentam grande variabilidade quando são comparados os resultados entre as ordens de perturbação e, agora, entre fonemas. Isto indica que medidas de tendência central aplicadas sobre as perturbações (é importante lembrar que mantém-se o sinal das diferenças sua estimação) resultam em um baixo grau de discriminação das patologias, pois as distribuições possuem um elevado nível de sobreposição.

Tabela 16 - Análise estatística das perturbações de ordem 4 (*jitter*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	122	13,563968	0,111180066
A regr. robusto	96	4,342668	0,045236125
Assimetria	79	0	0
B regr. clássica	120	7,221301	0,060177508
B regr. robusto	114	16,3427	0,143357018
Curtose	95	31,8	0,334736842
Desvio absoluto	122	58,93928	0,483108852
Desvio regr. clássica	121	75,94896	0,627677355
Desvio padrão	125	70,93928	0,56751424
Desvio regr. robusta	121	75,94896	0,627677355
Mediana	113	0,82958	0,007341416
Média	114	8,11269	0,071163947

Seguindo a tendência apresentada pelas demais ordens de perturbação para o fonema /e/, o número de discriminações apresenta uma baixa dispersão, além de mostrar que a elevação da ordem da derivada não melhora a identificação de patologias, diferente de quando se considera o uso do fonema /a/.

Tanto a média quanto o nível de separabilidade apresentam redução de performance, confirmando a redução de eficiência das medidas obtidas para as perturbações de período da fundamental. Provavelmente, o maior estiramento das pregas vocais durante a fonação do /e/ contribua para a criação de condições de uma vibração mais estável da mucosa, mesmo em casos disfônicos.

Tabela 17 - Análise estatística das medidas de perturbação (*jitter*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
DPF	113	5,65916	0,050081062
JF	113	55,93928	0,495037876
JR	119	59,93928	0,503691429
Média móvel (ordem 11)	120	69,14898	0,5762415
Média móvel (ordem 5)	123	61,43928	0,499506341
PFF	88	26,0097	0,295564773
PVI	131	187,35225	1,430169847
Série Temporal	121	145,51296	1,202586446
Variância	127	126,95081	0,999612677

Na tab. 17, verifica-se que as medidas PVI, avaliação por série temporal e variância dos períodos da fundamental apresentam maior distância entre os pares de tipos de voz quando comparado às medidas acústicas realizadas até agora. Isto indica que tais parâmetros são mais confiáveis pois a probabilidade de erro na identificação patológica é menor.

Outra alteração percebida em relação ao fonema /a/ é a redução na diferença entre distribuições dos valores para as medidas de média móvel, além da redução na quantidade de pares patológicos distinguíveis.

De modo geral, pode-se perceber que o arranjo da estrutura glotal durante a fonação sustentada do fonema /e/ produz períodos sucessivos similares entre as patologias, principalmente se for considerado que algumas vozes patológicas, como as produzidas por fenda fusiforme, são identificadas quando se considera avaliação do nível de ruído turbulento (alta frequência) na glote, e desta forma acabam sendo confundidas com vozes produzidas por pessoas normais.

Tabela 18 - Análise estatística das medidas de perturbação de ordem 0 (*jitter*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	100	36,466032	0,36466032
A regr. robusto	102	33,15604	0,325059216
Assimetria	102	19,40386	0,190233922
B regr. clássica	104	51,96603	0,499673365
B regr. robusto	116	17,65827	0,152226466
Curtose	127	12,033952	0,094755528
Desvio absoluto	99	108,0949556	1,091868238
Desvio regr. clássica	96	67,0667356	0,698611829
Desvio padrão	95	90,0386	0,947774737
Desvio regr. robusta	96	65,0667356	0,677778496
Mediana	111	64,75799	0,583405315
Média	96	2,02818	0,021126875

O fonema /i/ é o que proporciona maior esforço das estruturas de adução das pregas vocais. Conforme indicado por Horii (1969), com o aumento da frequência fundamental, ocorre uma redução das perturbações de frequência fundamental. Esta estabilidade artificial (superior à provocada pela fonação de /e/) faz com que as vozes dos pacientes assemelhem-se à das pessoas normais, em termos de anormalidades nos períodos ciclo-a-ciclo.

Isto pode ser visto pelos resultados da perturbação de primeira ordem, apresentada na tab. 18. Os parâmetros de desvio, que outrora mostravam resultados em número de discriminações, são os de menor resultado, apesar da proximidade com as demais medidas acústicas. O que ocorre é uma grande redução do intervalo de variação desta medida estatística de separabilidade dos pares disfônicos. Entretanto, quando é avaliada a distância entre as distribuições dos valores produzidos pelos índices acústicos, verifica-se que a tendência observada para os fonemas /a/ e /e/ permanece, apesar do limite superior das medições de nível e média de separabilidade ser menor que naquelas vocalizações.

Tabela 19 - Análise estatística das medidas de perturbação de ordem 1 (*jitter*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	116	38,86877	0,335075603
A regr. robusto	95	14,91069	0,156954632
Assimetria	102	3,05636	0,029964314
B regr. clássica	115	33,49885	0,291294348
B regr. robusto	94	4,40386	0,046849574
Curtose	111	28,724	0,258774775
Desvio absoluto	111	65,66288	0,591557477
Desvio regr. clássica	99	55,0667376	0,556229673
Desvio padrão	96	58,6628756	0,611071621
Desvio regr. robusta	99	55,0667376	0,556229673
Mediana	111	9,4320376	0,084973312
Média	102	22,963843	0,225135716

A tendência de equiparação do número de discriminações observada na tab. 19 é mantida nas medidas acústicas para perturbações de primeira ordem dos períodos da fundamental. Os resultados das três avaliações para os testes de hipótese são bastante semelhantes, principalmente na quantidade de pares em que é possível distinguir as patologias.

Novamente, as medidas de desvio (absoluto, padrão, da regressão clássica e da regressão robusta) apresentam uma performance total superior as demais medidas. A maioria dos índices acústicos apresentam redução no nível de separação entre as distribuições, com exceção de média e coeficiente A da regressão clássica, o mesmo ocorrendo para a média das separações (exceto para curtose e média).

Tabela 20 - Análise estatística das medidas de perturbação de ordem 2 (*jitter*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	96	0	0
A regr. robusto	124	20,65828	0,166599032
Assimetria	116	27,40386	0,236240172
B regr. clássica	95	1,0281756	0,010822901
B regr. robusto	128	16,6864456	0,130362856
Curtose	110	27,724	0,252036364
Desvio absoluto	117	62,169686	0,531364838
Desvio regr. clássica	98	54,0667376	0,551701404
Desvio padrão	99	59,6628756	0,602655309
Desvio regr. robusta	98	54,0667376	0,551701404
Mediana	114	7,53497	0,066096228
Média	105	6,05636	0,057679619

Considerando esta nova ordem de perturbação, verifica-se que as medidas de regressão robusta (coeficiente A e B) conseguem distinguir um maior número de pares, apesar de não superarem as medidas de dispersão, em nível de discriminação.

Outra evidência é que o número de pares onde as patologias são identificáveis é maior do que o obtido para as demais ordens de perturbação no fonema /i/, apesar de simultânea redução da distância entre as distribuições dos valores das medidas (“distância entre as disfonias”).

Tabela 21 - Análise estatística das medidas de perturbação de ordem 3 (*jitter*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	103	8,05636	0,078217087
A regr. robusto	112	4,02818	0,035965893
Assimetria	75	0	0
B regr. clássica	110	8,05636	0,073239636
B regr. robusto	110	3,02818	0,027528909
Curtose	109	28,35409	0,260129266
Desvio absoluto	117	61,169686	0,522817829
Desvio regr. clássica	98	56,0667376	0,572109567
Desvio padrão	99	60,0667376	0,606734723
Desvio regr. robusta	98	56,0667376	0,572109567
Mediana	109	19,35411	0,177560642
Média	86	0	0

Com o aumento da ordem da derivada dos períodos da fundamental, as medidas acústicas para o fonema /i/, percebe-se uma redução nos níveis e médias de discriminação, devido a maior estabilidade da fonação durante a pronúncia do fonema /i/.

É interessante verificar que as medidas de desvio, que mantiveram a mesma quantidade de discriminações, mesmo com a elevação da ordem para a perturbação (este fenômeno é naturalmente estendido para as médias de separabilidade). Comprova-se que estes parâmetros são muito úteis para avaliação de disfonias, mesmo considerando que outros índices são necessários para identificar todas as patologias.

Tabela 22 - Análise estatística das medidas de perturbação de ordem 4 (*jitter*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	89	5,72402	0,064314831
A regr. robusto	112	11,53497	0,102990804
Assimetria	96	15	0,15625
B regr. clássica	92	7,72402	0,083956739
B regr. robusto	126	9,53497	0,075674365
Curtose	108	28,35409	0,26253787
Desvio absoluto	117	60,169686	0,514270821
Desvio regr. clássica	99	57,0667356	0,576431673
Desvio padrão	100	60,0667376	0,600667376
Desvio regr. robusta	99	57,0667356	0,576431673
Mediana	109	0	0
Média	120	14,4884	0,120736667

Na tab. 22 fica evidenciada, comparando juntamente as perturbações de ordem zero até três para o fonema /i/, que as medidas desvio absoluto, desvio padrão, desvio de regressão clássica e de regressão robusta mantiveram as três avaliações do teste de Mann-Whitney, para verificação da

separabilidade das disfonias, praticamente nos mesmos valores, ou seja, as distribuições para as patologias separáveis e o número de pares discrimináveis não se alteraram com o aumento da ordem.

Considerando os três fonemas analisados, as medidas de dispersão utilizadas são significativamente capazes de identificar um número maior de disfonias com um maior grau de certeza, uma vez que as distribuições dos valores para cada patologia aparenta ser superior, em relação aos demais parâmetros considerados. Utilizar diferença quadrática ou absoluta tem uma influência marcante no pré-diagnóstico automático de anomalias na laringe e pregas vocais.

Tabela 23 - Análise estatística das medidas de perturbação (*jitter*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
DPF	100	5,02818	0,0502818
JF	128	97,79976	0,764060625
JR	123	94,203622	0,765883106
Média móvel (ordem 11)	118	110,20362	0,933928983
Média móvel (ordem 5)	121	102,79977	0,849584876
PFF	76	28,63468	0,376772105
PVI	99	129,23184	1,305372121
Série Temporal	122	42,56893	0,348925656
Variância	95	90,0386	0,947774737

Considerando as medidas clássicas, na tab. 23, percebe-se mais uma vez que as medidas que ponderam a elevação da frequência fundamental na estimativa das alterações dos períodos sucessivos de F0 apresentaram melhores resultados, principalmente JF e JR. As avaliações por médias móveis também mostram-se eficientes na identificação de patologias, garantindo inclusive boa separabilidade das distribuições dos valores para cada disfonia.

O índice acústico PVI demonstrou uma maior performance na distinção dos pares patológicos, fato que é confirmado para as demais vocalizações, podendo ser considerado independente para os três fonemas analisados, apesar de melhores resultados terem sido obtidos para o fonema /e/ (com grau de tensão intermediário nas pregas vocais em relação à /a/ e /i/).

O fonema /a/, de modo geral, apresentou melhores resultados na identificação de patologias, pelo método estatístico empregado, enquanto que avaliando-se as ordens de perturbação, a terceira derivada dos períodos sucessivos da frequência fundamental credita-se a maior eficiência, quando são aplicadas as técnicas estatísticas apresentadas.

Entretanto, foram os índices acústicos “clássicos” que discriminam com maior segurança (nível de separabilidade) as patologias estudadas. Verifica-se que o emprego do fonema /a/ ou /e/ para o cálculo deste parâmetros do sinal de voz indicam mais precisamente a anormalidade das estruturas de vocalização.

4.1.2 Perturbações da Amplitude da Frequência Fundamental

A utilização de medidas de variação de amplitude para discriminação de patologias da laringe e cordas vocais é apresentada nesta seção. Foram considerados os fonemas /a/, /e/ e /i/, sobre os quais se determinou os índices acústicos. O procedimento estatístico é apresentado na seção 3.2.9.

A primeira consideração importante é que medidas de avaliação de anomalias nas amplitudes de frequência fundamental produzem resultados bastante inferiores se comparadas às perturbações de período de F_0 . O modo como são estimadas as amplitudes ciclo-a-ciclo, é um fator importante pois foi empregado o pico de menor amplitude (que tem um grau de relação com o fechamento da glote) em um período da fundamental como estimativa de amplitude.

Outro fator importante é a flutuação de longo período encontrada nas vozes (patológicas e normais), proporcionado pela perda de energia acústica no fluxo de ar proveniente dos pulmões pela glote. Apesar do emprego de mecanismos eletrônicos de controle automático de ganho, pode-se notar pelas tabelas que seguem, a baixa relevância destas medidas quando comparadas aos parâmetros de alterações nos períodos sucessivos de frequência fundamental.

Tabela 24 - Análise estatística das perturbações de ordem 0 (*shimmer*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	84	20	0,238095238
A regr. robusto	79	3,00968	0,038097215
Assimetria	68	4,34267	0,063862794
B regr. clássica	80	16	0,2
B regr. robusto	69	0	0
Curtose	100	57,39121	0,5739121
Desvio absoluto	85	0	0
Desvio regr. clássica	88	0	0
Desvio padrão	80	0	0
Desvio regr. robusta	91	0	0
Mediana	85	4	0,047058824
Média	70	6	0,085714286

Considerando as medidas de perturbação de ordem zero, o índice acústico curtose proporcionou uma maior discriminação dos pares de patologias, permitindo que 43,29% das confrontações entre disfonias pudessem resultar em identificação das anomalias da laringe, com uma boa separação das distribuições (caracterizado pelos valores obtidos para nível e média da discriminação).

As demais medidas não acompanharam a performance do índice curtose. Os resultados nulos obtidos para a maioria dos parâmetros nas avaliações estatísticas de nível e média da discriminação indicam a grande proximidade das distribuições dos valores em relação aos pares de patologia distinguíveis. Levando-se em conta que utilizou-se um erro de rejeição na rejeição da hipótese nula de 5%, as medidas aplicadas sobre as perturbações de ordem zero.

Tabela 25 - Análise estatística das perturbações de ordem 1 (*shimmer*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	73	1	0,01369863
A regr. robusto	92	0	0
Assimetria	99	3,8	0,038383838
B regr. clássica	67	3	0,044776119
B regr. robusto	90	5,00968	0,055663111
Curtose	104	48,73207	0,468577596
Desvio absoluto	94	1	0,010638298
Desvio regr. clássica	92	0	0
Desvio padrão	84	0,82958	0,009875952
Desvio regr. robusta	92	0	0
Mediana	77	0	0
Média	85	17,80968	0,209525647

Avaliando estatisticamente as medidas aplicadas sobre as perturbações de primeira ordem, novamente a medida curtose consegue distinguir um maior número de pares de patologias, garantindo conjuntamente uma separabilidade das distribuições das disfonias.

Outra característica importante verificada com a elevação da ordem é o aumento do número de pares distinguíveis, apesar da redução na distância entre as distribuições dos valores de cada medida para os tipos de voz estudados. Ainda mais, a quantidade de valores nulos nas avaliações estatísticas de nível de média de separabilidade das disfonias pressupõem que as diferença entre medidas sucessivas de amplitude (perturbações de primeira ordem) fornecem melhores resultados na identificação de patologias do que a simples diferença a *i*-ésima amplitude da fundamental e o valor médio destas amplitudes.

Tabela 26 - Análise estatística das perturbações de ordem 2 (*shimmer*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	93	2,82958	0,030425591
A regr. robusto	73	0,279031	0,003822342
Assimetria	75	2	0,026666667
B regr. clássica	95	7,65916	0,080622737
B regr. robusto	68	0	0
Curtose	105	49,56165	0,472015714
Desvio absoluto	95	1	0,010526316
Desvio regr. clássica	93	0	0
Desvio padrão	85	0,82958	0,009759765
Desvio regr. robusta	93	0	0
Mediana	69	8,6	0,124637681
Média	82	5,83925893	0,071210475

O emprego da derivada de segunda ordem na determinação de medidas estatísticas que possibilitem a identificação de tipos de voz é apresenta na tab. 26. A medida de curtose consegue discriminar um número maior de pares patológicos, além de garantir uma boa distância entre as suas distribuições.

Os demais parâmetros acústicos mantiveram baixa capacidade de discriminação de patologias, apesar das avaliações do teste de hipótese experimentarem, em média, um aumento numérico, indicando que a elevação da ordem da perturbação para estimativa dos índices acústicos melhora a identificação das disfonias.

Tabela 27 - Análise estatística das perturbações de ordem 3 (*shimmer*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	82	3,82958	0,046702195
A regr. robusto	47	5,34267	0,11367383
Assimetria	101	8,00968	0,079303762
B regr. clássica	85	4,82958	0,056818588
B regr. robusto	45	8,34267	0,185392667
Curtose	107	53,56165	0,500576168
Desvio absoluto	93	1	0,010752688
Desvio regr. clássica	93	0	0
Desvio padrão	83	0,82958	0,00999494
Desvio regr. robusta	93	0	0
Mediana	84	0	0
Média	99	4,82958	0,048783636

Na tab. 27, pode-se perceber que os níveis e médias de separabilidade das distribuições de cada medida acústica para as disfonias elevaram-se, assim como o número médio de discriminações (apesar de alguns índices reduzirem sua capacidade de identificação de disfonias).

A medida de curtose sobre a perturbação de terceira ordem permite a melhor discriminação dos pares patológicos. É interessante notar que as estimativas dos coeficientes de regressão robusta operaram no sentido inverso das demais medidas uma vez que com o aumento da ordem da derivada sobre as estimativas de amplitude da fundamental, reduziram sua capacidade de identificação de disfonia.

Tabela 28 - Análise estatística das perturbações de ordem 4 (*shimmer*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	99	1	0,01010101
A regr. robusto	79	7	0,088607595
Assimetria	60	3	0,05
B regr. clássica	102	4,82958	0,047348824
B regr. robusto	80	2	0,025
Curtose	106	59,56165	0,561902358
Desvio absoluto	93	1	0,010752688
Desvio regr. clássica	91	0	0
Desvio padrão	83	0,82958	0,00999494
Desvio regr. robusta	91	0	0
Mediana	87	4,94227	0,056807701
Média	86	2,82958	0,032902093

Considerando a avaliação das disfonias através das perturbações de quarta ordem, verifica-se que não é possível estabelecer um padrão consistente de relação linear entre a ordem nas

perturbações e a capacidade de discriminação das disfonias e distanciamento entre as distribuições dos valores para cada patologia.

A única medida que permitiu identificar adequadamente as estruturas de vocalização foi a medida de curtose. O uso de técnicas de estatísticas sobre as perturbações das amplitudes ciclo-a-ciclo do fonema /a/ mostram que estas são significativamente menos eficazes que o uso das medidas de alterações da frequência fundamental.

Tabela 29 - Análise estatística das medidas de perturbação (*shimmer*) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
AVI	77	0,112692	0,001463532
DPF	106	8,65916	0,081690189
JF	63	0,82958	0,013167937
JR	67	0,82958	0,012381791
Média Móvel (ordem 11)	78	49,3427	0,632598718
Série Temporal	82	0	0
Variância	80	0	0

A medida de perturbação direcional (DPF) apresentou um número significativamente maior que os demais parâmetros clássicos, inclusive obtendo performance equivalente as medidas de curtose aplicadas sobre as estimações de perturbação de ordem superior.

Entretanto, o índice de melhor capacidade de separação entre as distribuições de seus valores nas disfonias foi a média móvel de ordem 11. As demais medidas apresentaram baixo rendimento na discriminação patológica. Aparentemente, os efeitos de perda energética no fluxo de ar produzem amplitudes aleatórias da fundamental que são muito semelhantes para os tipos de voz analisados.

Provavelmente as disfonias que caracterizam-se pela percepção de sopro são imperceptíveis para as medidas de perturbação de amplitudes sucessivas e suas derivadas, sendo confundidas com vozes normais na identificação automática.

A performance produzida pela utilização dos fonemas /e/ para aplicação das medidas de amplitude ciclo-a-ciclo é apresentada nas tabelas abaixo.

Tabela 30 - Análise estatística das perturbações de ordem 0 (*shimmer*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	103	1,5	0,014563107
A regr. robusto	98	0,82958	0,008465102
Assimetria	95	1	0,010526316
B regr. clássica	117	1,5	0,012820513
B regr. robusto	101	0,5	0,004950495
Curtose	113	53,55195	0,473911062
Desvio absoluto	92	14	0,152173913
Desvio regr. clássica	78	0	0
Desvio padrão	78	5	0,064102564
Desvio regr. robusta	78	0	0
Mediana	84	7,34267	0,087412738
Média	81	2,82958	0,034933086

Comparativamente às medidas realizadas sobre o fonema /a/, as medidas estatísticas realizadas sobre o sinal de perturbação de ordem zero conseguiram discriminar um número maior de disfonias, apesar de um menor nível de separabilidade entre os tipos de voz.

Apesar da medida de coeficiente da regressão linear conseguir distinguir as patologias uma maior quantidade de pares, foi o parâmetro curtose que oferece maior “segurança” na identificação das patologias.

Tabela 31 - Análise estatística das perturbações de ordem 1 (*shimmer*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	103	7	0,067961165
A regr. robusto	103	1	0,009708738
Assimetria	85	0	0
B regr. clássica	108	6	0,055555556
B regr. robusto	104	6,8	0,065384615
Curtose	109	59,72237	0,547911651
Desvio absoluto	95	6	0,063157895
Desvio regr. clássica	80	0	0
Desvio padrão	80	2	0,025
Desvio regr. robusta	80	0	0
Mediana	92	6,3	0,068478261
Média	107	5	0,046728972

Com a elevação da ordem da perturbação, o número de discriminações apresentou um crescimento médio um pouco superior ao mostrado na tab. 30, o mesmo ocorrendo para as medidas de separabilidade da distribuição das disfonias para os índices acústicos aplicados.

Pode-se notar que a medida de curtose (melhor parâmetro no grupo de estimativas estatísticas sobre a derivada de primeira ordem das amplitudes ciclo-a-ciclo) permitiu uma maior separabilidade das disfonias, apesar da diferença de 4 pares de disfonias distinguíveis a menos em relação à mesma medida empregada sobre perturbações de ordem zero.

Tabela 32 - Análise estatística das perturbações de ordem 2 (*shimmer*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	89	1	0,011235955
A regr. robusto	109	0	0
Assimetria	101	2	0,01980198
B regr. clássica	96	1	0,010416667
B regr. robusto	98	0	0
Curtose	108	56,72237	0,52520713
Desvio absoluto	95	5	0,052631579
Desvio regr. clássica	77	0	0
Desvio padrão	80	3	0,0375
Desvio regr. robusta	78	0	0
Mediana	123	20,83926	0,169424878
Média	111	6	0,054054054

O número de pares discrimináveis para as perturbações de segunda ordem foi superior ao obtido nos casos anteriores, para o fonema /e/, apesar da pequena redução da distância entre as distribuições das disfonias.

Verifica-se novamente que a medida curtose provê o melhor resultado (considerando as três avaliações dos testes de hipótese), mas a medida Mediana mostra um maior número de discriminações, com uma boa separabilidade entre as disfonias identificáveis. Cabe lembrar que este é um fato isolado para as apreciações de amplitudes sucessivas da frequência fundamental, mas que merece avaliação.

Tabela 33 - Análise estatística das perturbações de ordem 3 (*shimmer*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	96	1	0,010416667
A regr. robusto	89	3	0,033707865
Assimetria	100	10,8	0,108
B regr. clássica	105	0	0
B regr. robusto	97	0	0
Curtose	107	55,72237	0,520769813
Desvio absoluto	93	5	0,053763441
Desvio regr. clássica	78	0	0
Desvio padrão	81	3	0,037037037
Desvio regr. robusta	79	0	0
Mediana	107	11,34267	0,106006262
Média	101	0	0

O padrão de incremento da capacidade de identificar disfonias não é mais mantido a partir da terceira derivada das amplitudes da fundamental, conforme comprovado na tab. 33. Apesar das medidas Curtose e Mediana serem as de melhor capacidade de discriminação dos pares disfônicos (considerando juntamente os níveis e média de separabilidade).

Outro fato importante é a quantidade de valores nulos nas medidas de distância entre distribuição das disfonias, em cinco medidas estatísticas. Isto indica que estes parâmetros, apesar do número de discriminações de pares patológicos, produzem valores muito similares para as patologias, e provavelmente induzem à erros na identificação das patologias.

Tabela 34 - Análise estatística das perturbações de ordem 4 (*shimmer*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	102	0	0
A regr. robusto	102	9	0,088235294
Assimetria	102	1	0,009803922
B regr. clássica	107	6,8	0,063551402
B regr. robusto	99	6	0,060606061
Curtose	113	55,55195	0,491610177
Desvio absoluto	93	5	0,053763441
Desvio regr. clássica	79	0	0
Desvio padrão	82	3	0,036585366
Desvio regr. robusta	80	0	0
Mediana	102	16,00968	0,156957647
Média	103	5	0,048543689

Apesar da ampliação do número de discriminações na maioria das medidas acústicas apresentadas na tab. 34, não ocorreu uma melhoria significativa na separação entre as distribuições das disfonias, resultando na manutenção dos resultados obtidos nas perturbações de terceira ordem.

Pode-se verificar que a medida de mediana, que nas perturbações de ordem zero e de primeira ordem se mostrou pouco eficiente para identificar patologias, demonstra uma boa capacidade de identificação de patologias a partir da segunda ordem na derivada das amplitudes da fundamental.

Comparativamente aos resultados obtidos para o fonema /a/, utilizando-se uma vocalização que exija um pouco mais de tensão das pregas vocais produz melhor identificação de disfonias pois, aparentemente, a voz das pessoas normais torna mais densa as distribuições dos valores dos índices acústicos. A medida que melhor discrimina patologias quando se emprega medidas de perturbação de ordem superior foi o parâmetro curtose.

Tabela 35 - Análise estatística das medidas de perturbação (*shimmer*) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
AVI	83	7	0,084337349
DPF	97	13,74227	0,141672887
JF	94	4	0,042553191
JR	95	0	0
Média Móvel (ordem 11)	87	0	0
Série Temporal	113	30,81938	0,272737876
Variância	78	5	0,064102564

Empregando as medidas clássicas sobre o fonema /e/, os resultados mostrados pela tab. 35 demonstram uma elevação no número de pares de patologia discrimináveis. Entretanto, os níveis de separabilidade reduziram-se.

O parâmetro acústico DPF, que para o fonema /a/ foi o mais eficiente na identificação de patologias, posiciona-se abaixo do índice série temporal, o que mostra que a eliminação das correlações entre as amplitudes auxilia na identificação de patologias da laringe, apesar de ser uma

constatação isolada, que pode ser apenas obtida para o fonema /e/. A medida AVI, que encontra em seu similar PVI (para períodos sucessivos) um excelente índice para discriminação de disfonias, apresenta uma baixa performance.

São apresentados agora os resultados obtidos da aplicação de medidas acústicas sobre o fonema /i/ para avaliação das alterações nas amplitudes de frequência fundamental.

Tabela 36 - Análise estatística das perturbações de ordem 0 (*shimmer*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	77	0	0
A regr. robusto	83	1,02818	0,012387711
Assimetria	91	2,02818	0,022287692
B regr. clássica	84	0	0
B regr. robusto	87	2,0281756	0,023312363
Curtose	112	32,03399	0,286017768
Desvio absoluto	105	14,5	0,138095238
Desvio regr. clássica	92	2	0,02173913
Desvio padrão	98	4	0,040816327
Desvio regr. robusta	92	2	0,02173913
Mediana	104	8,43204	0,081077308
Média	69	0	0

O número de pares disfônicos discriminados é inferior ao obtido no fonema /e/, apesar de superior ao fonema /a/. Uma redução do nível de separabilidade pode ser constatado nos resultados obtidos na tab. 36. Isto pode significar que o aumento da tensão nas pregas vocais cria uma condição de estabilidade “artificial” da movimentação das estruturas de vocalização, produzindo amplitudes muito similares entre os tipos de voz.

A medida curtose apresenta melhores resultados na distinção de doenças da laringe e pregas vocais (nos três quesitos de análise do teste de hipótese), seguido pela medida de desvio absoluto.

Tabela 37 - Análise estatística das perturbações de ordem 1 (*shimmer*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	118	1,02818	0,00871339
A regr. robusto	111	17,30557	0,155906036
Assimetria	83	2	0,024096386
B regr. clássica	106	1,02818	0,009699811
B regr. robusto	104	8,90171	0,085593365
Curtose	111	44,06217	0,396956486
Desvio absoluto	100	16,52818	0,1652818
Desvio regr. clássica	96	6	0,0625
Desvio padrão	93	10	0,107526882
Desvio regr. robusta	96	6	0,0625
Mediana	115	7,6582656	0,066593614
Média	102	0	0

A maior tensão das pregas vocais na geração do fonema /i/ provocou um aumento da capacidade das medidas de perturbação de primeira ordem quanto à identificação de patologias (distinção dos pares patológicos por rejeição do teste de hipótese).

Apesar de um grande número de índices acústicos possibilitarem discriminar mais de 100 pares disfônicos, verifica-se que apenas os parâmetros curtose, coeficiente A da regressão robusta e desvio absoluto permitiram uma boa separabilidade das distribuições dos tipos de voz..

Tabela 38 - Análise estatística das perturbações de ordem 2 (*shimmer*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	96	0,63009	0,006563438
A regr. robusto	105	21,01359	0,200129429
Assimetria	84	2,02818	0,024145
B regr. clássica	103	6,6582656	0,064643355
B regr. robusto	113	17,5068	0,154927434
Curtose	113	46,06217	0,407629823
Desvio absoluto	102	15,52818	0,152237059
Desvio regr. clássica	96	6	0,0625
Desvio padrão	94	10	0,106382979
Desvio regr. robusta	96	6	0,0625
Mediana	114	41,5068	0,364094737
Média	110	2,0563556	0,018694142

A perturbação de segunda ordem das amplitudes da fundamental permitiram que duas medidas acústicas apresentassem equivalência na distinção de patologias. Os índices curtose e mediana foram capazes de discriminar disfonias em níveis e média de discriminação superiores às das outras medidas.

O coeficiente A da regressão, apesar da redução do número de pares onde as patologias foram identificadas, ampliou sua margem de segurança na discriminação de disfonias, pelo aumento dos dois avaliadores do teste de hipótese.

Um aumento na distância entre as distribuições permite uma redução na probabilidade de erro quando da avaliação automática das patologias, o que possibilita dizer que os parâmetros acústicos aplicados sobre a derivada de segunda ordem das amplitudes sucessivas da fundamental promovem uma eficiente discriminação das disfonias.

Tabela 39 - Análise estatística das perturbações de ordem 3 (*shimmer*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	103	3,6864456	0,035790734
A regr. robusto	98	6,4298856	0,065611078
Assimetria	99	6,534966	0,066009758
B regr. clássica	117	8,68645	0,074243162
B regr. robusto	100	1,02818	0,0102818
Curtose	114	47,06217	0,412826053
Desvio absoluto	104	16,52818	0,158924808
Desvio regr. clássica	96	6	0,0625
Desvio padrão	96	10,0281756	0,104460163
Desvio regr. robusta	96	6	0,0625
Mediana	84	5,534966	0,065892452
Média	82	0	0

De modo análogo ao ocorrido no fonema /e/, a tendência de melhora na performance das medidas acústicas sobre as perturbações de terceira ordem não é mantida. Mesmo assim, a medida curtose praticamente apresenta os mesmos resultados obtidos na tab. 37.

A medida de desvio absoluto, ao longo das perturbações de amplitude também mantém a mesma eficiência, mostrando ser independente da ordem da derivada, enquanto que as demais medidas acústicas mostram um comportamento aleatório.

Comparando estas medidas com as obtidas para os demais fonemas, pode-se perceber que a vogal sustentada /i/ apresenta melhores resultados na utilização de medidas de amplitude, provavelmente porque as distribuições são leptocúrticas.

Tabela 40 - Análise estatística das perturbações de ordem 4 (*shimmer*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	85	8,63009	0,101530471
A regr. robusto	69	1	0,014492754
Assimetria	86	2,02818	0,023583488
B regr. clássica	105	10,13688	0,096541714
B regr. robusto	74	0	0
Curtose	116	47,06217	0,405708362
Desvio absoluto	104	16,52818	0,158924808
Desvio regr. linear	96	6	0,0625
Desvio padrão	98	10,0281756	0,102328322
Desvio regr. robusta	96	6	0,0625
Mediana	95	20,5068	0,215861053
Média	107	9,68645	0,09052757

A tab. 41 mostra uma redução na quantidade de pares de disfonias discrimináveis, apesar da maior distância entre disfonias (considerando os valores obtidos para as medidas acústicas). A medida de desvio absoluto também apresentou bons resultados quando aplicada sobre as perturbações consideradas.

Pode-se perceber que o parâmetro acústico mediana apresentou grandes oscilações, com melhor resultado para a segunda e quarta ordem da derivada das amplitudes da fundamental.

Pode-se perceber que as perturbações de amplitude oferecem baixa capacidade de discriminação de patologias se comparadas às avaliações realizadas sobre os períodos ciclo-a-ciclo da fundamental. Verifica-se que o intervalo de variação do número de discriminações é baixo em relação às anomalias de frequência fundamental.

Tabela 41 - Análise estatística das medidas de perturbação (*shimmer*) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
AVI	102	10	0,098039216
DPF	106	5,932042	0,05596266
JF	101	5,02818	0,04978396
JR	105	12,02818	0,114554095
Média Móvel (ordem 11)	107	4,02818	0,037646542
Série Temporal	109	28,88033	0,264957156
Variância	98	4	0,040816327

Finalizando as análises de discriminação dos pares de tipos de voz, aplicando-se as medidas clássicas sobre as perturbações de amplitude do fonema /i/, constata-se que a medida de série temporal proporciona bons resultados na identificação das disfonias (o mesmo ocorrendo para o fonema /e/).

Um resultado interessante é o proporcionado pelo parâmetro JR, mas deve ser encarado apenas como um acontecimento isolado pois somente neste fonema é que permitiu um certo distanciamento das distribuições, superior inclusive à média móvel e ao parâmetro DPF.

Constata-se que os fonemas /e/ e /i/ para estimação de medidas acústicas ligadas à perturbação de amplitude da frequência fundamental proporcionam melhor performance. Se for comparado este resultado com o obtido para medidas de *jitter*, pode-se perceber que a avaliação das perturbações na vocalização (principalmente doenças com características de aspereza e rouquidão) é mais eficiente se realizada através de estimativas de perturbação de períodos sucessivos de F_0 . Isto se deve a consistência na estimação destes eventos no sinal de voz.

Verificou-se que a medida curtose permite uma boa identificação de patologias, fato não encontrado na literatura, além da independência de fonema. As variações da distribuição das perturbações de amplitude devem começar em curvas leptocúrticas (vozes normais) até as platicúrticas (disfonias).

As medidas clássicas provaram ser dependentes do tipo de vogal sustentada empregada. Verifica-se que para o fonema /a/, a medida DPF apresenta os melhores resultados de discriminação, enquanto que para as demais vogais sustentadas (e estudadas), a medida realizada através de séries temporais mostrou-se mais eficiente. Esta variabilidade pode ser percebida nas tabelas obtidas para as medidas de *shimmer*.

Fica claro que é necessário considerar juntamente à estas medidas de perturbação da vibração das pregas vocais, parâmetros acústicos que considerem a turbulência de ar através da glote, pois o máximo número de pares discrimináveis foi 61,47% de pares disfônicos para as medidas de média móvel (ordem 5) e JF (*jitter factor*), ambas sobre o fonema /a/.

4.2 Medidas de Ruído

Um conjunto de índices acústicos foi empregado para quantificar o ruído no sinal de voz, proveniente da inadequada adução das pregas vocais. Para avaliar a capacidade destas medidas em discriminar patologias como cistos, nódulos e pólipos, por exemplo, utilizou-se a metodologia estatística descrita na seção 3.5, sendo considerado inclusive os efeitos produzidos pelo emprego de diferentes fonemas (/a/, /e/ e /i/).

A primeira análise é realizada sobre as medidas de energia espectral (em determinadas bandas) e magnitude do primeiro e segundo harmônicos, sendo descrita na seção 4.2.1. Os índices acústicos obtidos através de técnicas cepstrais são apresentados na seção 4.2.2. As avaliações de aspereza baseado na curva densidade espectral de potência pelo método apresentado por Sansone e Emanuel (1970) seguem na seção 4.2.3, e finalmente são apresentados os resultados da discriminação para os demais parâmetros acústicos que em sua maioria são desenvolvidos sobre a curva de densidade espectral de potência.

4.2.1 Avaliação das medidas harmônicas

As magnitudes da primeira e segunda componentes harmônicas são importantes na avaliação da condição soprosa da voz de pacientes e, desta forma, avaliar o escape de ar glotal durante a fase de fechamento das pregas vocais.

Tabela 42 - Análise estatística do primeiro harmônico para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	116	49,95197	0,430620431
A regr. robusto	109	31,12237	0,28552633
Assimetria	107	10	0,093457944
B regr. clássica	124	83,73816	0,675307742
B regr. robusto	129	52,73816	0,408822946
Curtose	110	62,20858	0,565532545
Desvio absoluto	131	144,84409	1,105680076
Desvio regr. clássica	129	159,68337	1,237855581
Desvio padrão	135	134,67367	0,997582741
Desvio regr. robusta	127	159,68337	1,25734937
Mediana	100	41,1224	0,411224
Média	106	49,1224	0,463418868

Pode-se perceber que as medidas de dispersão (principalmente desvio clássico e robusto das regressão linear) da primeira harmônica conseguem discriminar um maior número de disfonias, com um nível de segurança garantido pela distância entre as distribuições das disfonias no espaço unidimensional (nível e média das distribuições).

Tabela 43 - Análise estatística do segundo harmônico para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	124	42,401421	0,341946944
A regr. robusto	121	39,12239	0,323325537
Assimetria	100	5,5	0,055
B regr. clássica	113	15,83928	0,140170619
B regr. robusto	122	44,11828	0,361625246
Curtose	115	39,12958	0,340257217
Desvio absoluto	128	121,84409	0,951906953
Desvio regr. clássica	121	151,19304	1,249529256
Desvio padrão	125	130,84409	1,04675272
Desvio regr. robusta	124	151,19304	1,21929871
Mediana	123	41,28873	0,335680732
Média	125	61,40139	0,49121112

Considerando agora o segundo harmônico, nota-se primeiramente que sua capacidade de discriminação é levemente inferior em relação ao primeiro harmônico. As medidas de dispersão (desvio absoluto e desvio padrão) conseguiram discriminar um maior número de patologias, apesar da manutenção das distâncias entre as distribuições das disfonias em relação à tab. 42.

Verifica-se que as medidas estatísticas aplicadas sobre as estimativas do segundo harmônico apresentam uma uniformidade de valores, além do incremento discriminatório de medidas de tendência central (média e mediana).

Tabela 44 - Análise estatística do primeiro harmônico para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	109	30,95196	0,283962936
A regr. robusto	121	35,95196	0,297123636
Assimetria	79	7,96162893	0,100780113
B regr. clássica	76	36,08873	0,474851711
B regr. robusto	76	23,8097	0,313285526
Curtose	87	15,279	0,17562069
Desvio absoluto	138	76,331	0,553123188
Desvio regr. clássica	133	119,00326	0,894761353
Desvio padrão	133	91,33101	0,686699323
Desvio regr. robusta	135	125,83284	0,932095111
Mediana	118	34,42239	0,291715169
Média	117	46,42239	0,396772564

Com o leve aumento da tensão nas pregas vocais e modificação na cavidade supraglotal (para produção do fonema /e/), verifica-se um incremento no número de patologias identificadas quando se

emprega medidas de dispersão do primeiro harmônico do sinal de voz. Entretanto, as distribuições das disfonias (para os valores dos parâmetros acústicos) estão mais próximas.

Tabela 45 - Análise estatística do segundo harmônico para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	113	20,3097	0,179731858
A regr. robusto	115	21,5097	0,18704087
Assimetria	83	9,80968	0,118188916
B regr. clássica	108	34,6314	0,320661111
B regr. robusto	95	10,00968	0,105365053
Curtose	86	3,65916	0,042548372
Desvio absoluto	137	89,57474	0,653830219
Desvio regr. clássica	131	144,02264	1,099409466
Desvio padrão	138	136,1834	0,986836232
Desvio regr. robusta	134	145,02264	1,082258507
Mediana	110	20,83928	0,189448
Média	113	32,13928	0,284418407

Comparando os resultados obtidos entre os fonemas /a/ e /e/, tem-se o indicativo que a elevação da tensão nas pregas vocais permite um aumento no número de tipos de voz discrimináveis. Contudo, isto implica na redução da distância entre as distribuições motivado pela maior similaridade entre as disfonias para estas medidas acústicas (ou avaliações estatísticas da magnitude dos harmônicos).

Provavelmente, as exigências mecânicas sobre as pregas vocais impliquem em um fechamento maior das pregas vocais para a produção do fonema /e/, em relação ao fonema /a/. As disfonias como nódulos, cistos e pólipos, por exemplo, caracterizadas por um crescimento orgânico nas paredes da glote são melhor identificáveis pois qualquer anomalia na mucosa fica mais evidente nestas condições de maior frequência de vibração.

Tabela 46 - Análise estatística do primeiro harmônico para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	102	16,41065	0,160888725
A regr. robusto	97	11,00679	0,113472062
Assimetria	100	3,5	0,035
B regr. clássica	110	36,3776976	0,330706342
B regr. robusto	98	63,34953	0,646423776
Curtose	112	17,752196	0,15850175
Desvio absoluto	142	126,70363	0,892279085
Desvio regr. clássica	142	117,3561356	0,826451659
Desvio padrão	154	131,703642	0,855218455
Desvio regr. robusta	140	118,32796	0,845199714
Mediana	103	43,4107	0,421463107
Média	107	52,4107	0,489819626

Aumentando a tensão sobre as pregas vocais ainda mais, confirma-se as evidências encontradas para os fonemas anteriormente analisados. Nota-se que os parâmetros acústicos baseado

em medidas de dispersão apresentaram melhores resultados em relação às demais medidas, inclusive quando compara-se a performance dos demais fonemas.

A distância das distribuições das disfonias apresenta um resultado inferior às obtidas nas tabs. 42 e 44, confirmando as expectativas de que a elevação da tensão nas pregas vocais faz com que sejam produzidas estimativas harmônicas muito semelhante entre as patologias analisadas.

Tabela 47 - Análise estatística do segundo harmônico para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	125	55,5068	0,4440544
A regr. robusto	132	55,5068	0,420506061
Assimetria	89	13	0,146067416
B regr. clássica	93	4,02818	0,043313763
B regr. robusto	97	2,02818	0,020909072
Curtose	90	15,53497	0,172610778
Desvio absoluto	126	81,3146	0,645353968
Desvio regr. clássica	127	111,23185	0,875841339
Desvio padrão	135	81,69689	0,605162148
Desvio regr. robusta	125	112,23185	0,8978548
Mediana	107	53,5068	0,500063551
Média	109	53,5068	0,490888073

Considerando as avaliações estatísticas sobre o segundo harmônico da curva de densidade espectral de potência, quando determinadas sobre o fonema /i/, produzem valores que caracterizam distribuições das patologias muito próximas. Isto pode ser verificado pelas avaliações do nível e média de discriminação dos testes de hipótese realizados. Entretanto, a tendência de aumento da quantidade de pares patológicos discrimináveis não foi significativamente superior quando comparado ao fonema /e/.

Percebe-se também que as medidas de dispersão da magnitude do primeiro e do segundo harmônico permitem uma melhor distinção das disfonias em comparação aos demais índices. Conforme a extensão da patologia, tem-se uma maior ou menor variabilidade na estrutura harmônica ao longo do sinal de voz.

Patologias com características de soprosidade propiciam modificações importantes na característica harmônica, fato que pode ser comprovado pelos resultados apresentados, onde a capacidade individual de discriminação é considerada. Aparentemente, o emprego de fonemas em que uma maior tensão é aplicada sobre as pregas vocais possibilita uma melhor identificação de disfonias onde a glote apresenta adução incompleta, seja por crescimento orgânico (nódulos e pólipos) ou por mal funcionamento das estruturas de vocalização (fenda fusiforme).

4.2.2 Medidas de energia espectral

A avaliação da energia espectral também foi considerada, assim como alterações entre trechos sucessivos do sinal de voz sobre os quais determinava-se a curva de densidade espectral de potência.

Verifica-se que alterações patológicas achatam esta curva espectral, tornando-a semelhante ao espectro do ruído branco. O aumento da magnitude espectral nas altas frequências e simultânea redução energética nas frequências formantes são evidências de anormalidades no sinal glotal.

Através da determinação da energia (valor médio quadrático das magnitudes) em determinadas bandas de frequência, pode-se discriminar disfonias com características perceptuais de rouquidão e soproidade.

Tabela 48 - Análise estatística da energia espectral (60 à 400 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	108	37,12239	0,343725833
A regr. robusto	110	35,12239	0,319294455
Assimetria	79	16,18332986	0,204852277
B regr. clássica	119	46,73816	0,392757647
B regr. robusto	114	54,73816	0,480159298
Curtose	119	138,3425	1,162542017
Desvio absoluto	138	126,7314	0,918343478
Desvio regr. clássica	127	153,68335	1,210105118
Desvio padrão	139	132,56098	0,953676115
Desvio regr. robusta	128	161,69303	1,263226797
Mediana	103	46,12239	0,447790194
Média	106	45,12239	0,425682925

As medidas estatísticas sobre as estimativas de energia entre as frequências 60 e 400 Hz mostram que as disfonias são melhor discrimináveis quando se quantifica a dispersão destas estimativas (principalmente a medida de desvio padrão). Paralelamente, as distâncias entre as distribuições das patologias indicam que estes índices acústicos devem prover uma quantidade menor de erros durante a discriminação.

Outra medida que demonstra a variabilidade das estimações sucessivas de energia ao longo do sinal de voz é a curtose, devido ao nível e média de discriminação bem próximos aos cálculos de desvios.

Tabela 49 - Análise estatística da energia espectral (400 à 2000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	125	78,11936	0,62495488
A regr. robusto	128	78,11936	0,6103075
Assimetria	86	9,631381	0,111992802
B regr. clássica	91	3,82958	0,042083297
B regr. robusto	99	4,82958	0,048783636
Curtose	96	57,71829	0,601232188
Desvio absoluto	103	5,00968	0,04863767
Desvio regr. clássica	127	154,51293	1,216637244
Desvio padrão	112	52,06101	0,464830446
Desvio regr. robusta	127	156,51293	1,232385276
Mediana	108	75,44896	0,698601481
Média	110	73,44896	0,667717818

A mudança de banda de frequência (400 à 2000 Hz) mostra que a capacidade de discriminação das disfonias é inferior ao apresentado na tab. 48. Percebe-se que algumas medidas de dispersão mantiveram a mesma capacidade de identificar patologias (inclusive considerando as distâncias entre as distribuições das disfonias), enquanto que com outras (desvio absoluto e padrão) não se obteve a mesma performance.

Considerando que as medidas de coeficiente angular (A) das regressões lineares (tanto clássica quanto robusta) conseguiram discriminar um número maior de disfonias, pode-se afirmar que existe uma tendência de redução da energia, principalmente nesta banda de frequência, ao longo do sinal de voz. Isto explica a diferença encontrada entre as medidas de dispersão.

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	108	33,81936	0,313142222
A regr. robusto	103	36,81936	0,357469515
Assimetria	98	0	0
B regr. clássica	86	5,83926	0,067898372
B regr. robusto	84	6,65916	0,079275714
Curtose	83	32,77225	0,394846386
Desvio absoluto	107	22,00186	0,20562486
Desvio regr. clássica	124	142,33283	1,147845403
Desvio padrão	109	44,62121893	0,409368981
Desvio regr. robusta	127	142,33283	1,120730945
Mediana	105	34,81938	0,331613143
Média	100	30,81936	0,3081936

Tabela 50 - Análise estatística da energia espectral (2000 à 5000 Hz) para fonema /a/

A energia espectral entre 2000 e 5000 Hz mostra que a distância entre as curvas de distribuição das disfonias para estas medidas reduz-se à medida que se considera bandas mais altas. Isto mostra que densidade espectral de potência é semelhante entre as disfonias, apesar do aumento de ruído provocado pelas disfonias.

Mesmo assim, as medidas acústicas de desvio das regressões lineares pelo método clássico e método robusto mantêm praticamente o mesmo número de pares patológicos distinguíveis.

Tabela 51 - Análise estatística da energia espectral (5000 à 8000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	113	15,62958	0,138314867
A regr. robusto	112	19,62958	0,175264107
Assimetria	95	1	0,010526316
B regr. clássica	119	10,66884	0,089654118
B regr. robusto	96	1,82958	0,019058125
Curtose	85	42,25916	0,497166588
Desvio absoluto	133	0	0
Desvio regr. clássica	135	97,07316	0,719060444
Desvio padrão	132	22,97852	0,174079697
Desvio regr. robusta	134	95,07316	0,709501194
Mediana	111	29,63926	0,26702036
Média	112	24,63926	0,219993393

A última faixa de frequência considerada (entre 5000 e 8000 Hz) confirma a proximidade entre as distribuições das disfonias, evidenciado pelo nível de média das discriminações. Entretanto, verifica-se que o número de pares de tipos de voz através das medidas acústicas calculadas é superior aos resultados obtidos para as bandas 400 à 2000 Hz e 2000 à 5000 Hz.

Isto confirma as evidências espectrais de que a elevação da energia espectral nas altas frequências está relacionado com a extensão das anomalias nas estruturas de vocalização. As melhores discriminações para o fonema /a/ através desta metodologia foram obtidos quando considerou-se a faixa entre 60 à 400 Hz, provavelmente pela maior disparidade entre as vozes disfônicas e as normais.

Tabela 52 - Análise estatística da energia espectral (60 à 400 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	95	24,12237	0,253919684
A regr. robusto	100	29,12237	0,2912237
Assimetria	99	58,0014	0,585872727
B regr. clássica	83	28,8097	0,347104819
B regr. robusto	84	30,63928	0,364753333
Curtose	119	117,34406	0,986084538
Desvio absoluto	121	72,16056	0,596368264
Desvio regr. clássica	114	114,02266	1,000198772
Desvio padrão	117	97,50325	0,833361111
Desvio regr. robusta	112	115,02266	1,026988036
Mediana	102	36,12239	0,354141078
Média	105	38,12239	0,363070381

Aplicando a mesma metodologia (energia espectral) sobre fonema /e/, constata-se a baixa capacidade dos tipos de voz em comparação aos resultados obtidos para o fonema /a/. A medida de curtose apresenta o melhor resultado na identificação das disfonias, mantendo resultados semelhantes aos obtidos para o fonema anterior.

Considerando todas as medidas, pode-se verificar que o aumento da tensão nas pregas vocais e a característica de ressonância da estrutura supraglotal afetam significativamente a separabilidade das patologias.

Tabela 53 - Análise estatística da energia espectral (400 à 2000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	96	24,6	0,25625
A regr. robusto	97	28,60968	0,294945155
Assimetria	83	1	0,012048193
B regr. clássica	74	9,00968	0,121752432
B regr. robusto	68	7,00968	0,103083529
Curtose	102	31,46884	0,308518039
Desvio absoluto	96	29,81936	0,310618333
Desvio regr. clássica	117	112,73395	0,963538034
Desvio padrão	105	67,61938	0,643994095
Desvio regr. robusta	117	113,73395	0,972085043
Mediana	92	11,8	0,12826087
Média	91	22,6	0,248351648

A tab. 53, permite que se verifique uma redução na capacidade de discriminação dos pares disfônicos. As medidas de desvio da energia ao longo do sinal de voz, das regressões lineares (clássica e robusta) realizaram melhor uso da energia estimada nesta banda de frequência para o fonema /e/.

Apesar disto, o emprego deste fonema apresentam baixa performance na separação das disfonias, seja em número de discriminações quanto em distância das distribuições das patologias (nível e média de discriminações).

Tabela 54 - Análise estatística da energia espectral (2000 à 5000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	95	34,42958	0,362416632
A regr. robusto	99	35,42958	0,357874545
Assimetria	93	3	0,032258065
B regr. clássica	103	5,65916	0,054943301
B regr. robusto	94	3,65916	0,038927234
Curtose	125	81,49358	0,65194864
Desvio absoluto	109	5,82958	0,053482385
Desvio regr. clássica	122	155,85224	1,277477377
Desvio padrão	121	24,09461893	0,199129082
Desvio regr. robusta	122	151,85224	1,244690492
Mediana	100	16,8	0,168
Média	102	19,8	0,194117647

Utilizando a energia espectral entre 2000 e 5000 Hz, pode-se verificar uma melhora, estatística, na capacidade de discriminação das patologias empregando-se os índices acústicos apresentados, com algumas medidas mostrando serem superiores aos resultados obtidos no fonema /a/. Verifica-se inclusive que dentre as estimativas de energia espectral realizadas sobre o fonema /e/ para as quatro bandas consideradas, esta foi a que melhor discriminou as disfonias (considerando simultaneamente o número e o nível/média das discriminações).

As medidas de desvio das regressões lineares (tanto clássica como robusta) em relação as medidas de dispersão (desvio absoluto e padrão) mostram que a energia ao longo do sinal de voz (5 segundos) apresenta uma tendência linear (de queda, motivada pela perda de energia cinética do fluxo de ar através das pregas vocais).

Tabela 55 - Análise estatística da energia espectral (5000 à 8000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	114	39,78155	0,348960965
A regr. robusto	113	40,79123	0,360984336
Assimetria	100	5,12237	0,0512237
B regr. clássica	108	4,80968	0,044534074
B regr. robusto	116	10,91269	0,094074914
Curtose	123	65,38153893	0,531557227
Desvio absoluto	112	1,82958	0,016335536
Desvio regr. clássica	115	86,51291	0,752286174
Desvio padrão	127	37,54227	0,295608425
Desvio regr. robusta	116	94,51294	0,814766724
Mediana	116	44,95198	0,387517069
Média	116	45,95198	0,396137759

A ultima banda considerada (entre 5000 à 8000 Hz) caracteriza-se por medidas acústicas uniformes, ou seja, o número de discriminações os cálculos estatísticos desenvolvidos é muito semelhante. Além disto, o índice acústico Desvio Padrão consegue discriminar mais patologias em comparação aos demais parâmetros acústicos utilizados para o fonema /e/.

Verifica-se que, no caso do fonema /a/, as melhores bandas de frequência para aplicação das estimativas de energia para discriminação de disfonias foram 60 à 400 Hz e 400 à 2000 Hz, e para o fonema /e/, a faixa entre 2000 e 5000 Hz apresenta melhores resultados de discriminação. Pode-se explicar isto pela disposição de energia na curva de densidade espectral de potência, mas precisamente pela posição das frequências formantes e suas respectivas bandas de energia. O fonema /a/ possui energia espectral concentrada entre 500 e 1500 Hz, onde se localizam o primeiro e segundo formantes. Já o fonema /e/ tem sua energia dispersa no espectro, onde o primeiro formante localiza-se em 500 Hz e os demais podem ser localizados a partir de 2000 Hz.

Estas evidências podem indicar que a avaliação de energia na curva de densidade espectral de potência deve ser realizada em locais onde exista uma clara diferença (região das formantes) entre vozes patológicas e normais, uma vez que a tendência de nivelamento do espectro conforme a extensão da disfonia é mais visível nestas bandas espectrais.

Tabela 56 - Análise estatística da energia espectral (60 à 400 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	92	10,50679	0,114204239
A regr. robusto	92	6,50679	0,070725978
Assimetria	97	55,96706	0,57698
B regr. clássica	90	24,0349756	0,267055284
B regr. robusto	92	35,43884	0,385204783
Curtose	114	45,96706	0,403219825
Desvio absoluto	125	60,41066	0,48328528
Desvio regr. clássica	144	110,7317956	0,768970803
Desvio padrão	137	89,67546	0,654565401
Desvio regr. robusta	144	110,7317956	0,768970803
Mediana	107	40,43888	0,377933458
Média	109	44,4388756	0,407696106

O fonema /i/ possibilita as menores distâncias entre distribuições de patologias considerando os fonemas estudados. Apesar disto, algumas medidas acústicas, principalmente os desvios de regressão linear (clássica e robusta), permitiram um elevado número de pares disfônicos separáveis, superior aos resultados encontrados para os fonemas /a/ e /e/. Mesmo assim, se for considerado o valor médio da quantidade de discriminações dos cálculos estatísticos realizados para as estimativas de energia nesta faixa de frequência, verifica-se que o fonema /i/ apresenta os menores resultados para a identificação de disfonias.

Constata-se que o fonema /a/ é o que apresenta melhores resultados (tanto em número de pares separáveis quanto em distância das distribuições das patologias), provavelmente pela proximidade das formantes nesta faixa de frequência.

Tabela 57 - Análise estatística da energia espectral (400 à 2000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	123	6,72402	0,054666829
A regr. robusto	123	11,72402	0,095317236
Assimetria	83	4,50679	0,054298675
B regr. clássica	108	2,0281756	0,018779404
B regr. robusto	102	2,02818	0,019884118
Curtose	112	14,53498	0,129776607
Desvio absoluto	112	79,94469	0,713791875
Desvio regr. clássica	132	96,00105	0,727280682
Desvio padrão	118	92,94469	0,787666864
Desvio regr. robusta	134	96,00105	0,716425746
Mediana	105	4,72402	0,044990667
Média	107	1	0,009345794

Considerando a banda espectral entre 400 e 2000 Hz, pode-se perceber que existe uma tendência de queda no nível e média de discriminações com o aumento da tensão sobre as pregas vocais, indicando uma maior proximidade das distribuições das patologias. Isto pode ser verificado em todas as avaliações de energia espectral.

As medidas que ofereceram melhor discriminação para este fonema foram os desvios das regressões lineares (clássica e robusta), seguidos das medidas de desvio padrão e desvio absoluto. As demais medidas acústicas (baseadas em métodos estatísticos sobre as estimativas de energia ao longo do sinal de voz) apresentam baixa diferenciação entre as distribuições das disfonias, mostrando serem pouco confiáveis na identificação de patologias.

Tabela 58 - Análise estatística da energia espectral (2000 à 5000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	92	28,9039	0,314172826
A regr. robusto	85	15,90386	0,187104235
Assimetria	109	23,5068	0,215658716
B regr. clássica	92	1	0,010869565
B regr. robusto	95	2	0,021052632
Curtose	119	27,96702	0,235016975
Desvio absoluto	115	6,87353	0,059769826
Desvio regr. clássica	146	92,32795	0,632383219
Desvio padrão	134	51,64507	0,38541097
Desvio regr. robusta	147	91,82795	0,624679932
Mediana	94	19,4039	0,206424468
Média	91	21,4039	0,235207692

A tab. 58 mostra que as medidas de dispersão (principalmente os desvios das regressões lineares clássica e robusta) permitem separar um número elevado de pares patológicos, suplantando todas medidas acústicas realizadas sobre as estimativas de energia espectral, mesmo em outras faixas de frequência. Isto pode ser motivado pelo posicionamento das formantes (segunda e terceira) bem próximas da banda aqui considerada, seguido do aumento da energia das altas frequências.

Tabela 59 - Análise estatística da energia espectral (5000 à 8000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	127	39,53498	0,311299055
A regr. robusto	124	39,53498	0,318830484
Assimetria	93	1	0,010752688
B regr. clássica	116	14,8085556	0,127659962
B regr. robusto	116	12,78038	0,11017569
Curtose	118	36,59711	0,310145
Desvio absoluto	100	8,50679	0,0850679
Desvio regr. clássica	129	32,56893	0,252472326
Desvio padrão	110	14,5068	0,13188
Desvio regr. robusta	130	29,56893	0,227453308
Mediana	108	30,53498	0,282731296
Média	115	34,53498	0,300304174

A última banda analisada para o fonema /i/ mostra uma uniformidade (baixa variabilidade) no número de discriminações entre as medidas acústicas realizadas (com base na energia espectral), juntamente com os menores níveis e médias de separabilidade.

Isto ocorre devido a distribuição da energia na curva de densidade espectral de potência. Como a maioria das patologias é caracterizada por um aumento de sopro, existe uma maior concentração de energia proveniente de ruído nas altas frequências. Como a concentração de energia harmônica para o fonema /i/ inicia-se a partir de 2000 Hz, a semelhança entre as curvas de densidade espectral de potência de pessoas normais e pacientes disfônicos torna difícil a discriminação de patologias utilizando medidas de energia sobre esta banda espectral para este fonema.

A utilização de medidas de energia sobre determinadas faixas de frequência apresenta bons resultados na distinção de patologias, principalmente pelo fato da grande maioria das disfonias estudadas serem identificáveis perceptualmente como soprosas.

Mais ainda, tais avaliações devem ser realizadas nas regiões espectrais de maior energia, para que a distinção entre pessoas normais e disfônicas seja feita com uma confiabilidade maior. Isto também proporciona um intervalo de variação maior das medidas acústicas.

Particularmente, as medidas de desvio das regressões lineares clássica e robusta possibilitam que um maior número de disfonias sejam identificadas, indicando que deve-se considerar uma perda na magnitude espectral devido ao decréscimo de energia cinética do fluxo de ar pulmonar através da glote. Tal constatação proporciona que estes índices de dispersão são mais adequados para discriminação patológica.

4.2.3 Medidas de Energia Espectral - HFPR

As medidas realizadas nesta seção são baseadas no trabalho de Shoji *et al* (1992-1), onde determina-se a curva média da densidade espectral de potência e avalia-se então a energia em intervalos de frequência bem definidos.

Para determinar a capacidade de discriminação destas medidas, foi empregado o teste de hipótese definido na seção 3.5, e seus resultados são apresentados na tabela abaixo.

Tabela 60 - Análise estatística da medida HFPR

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Fonema /a/			
60 à 400 Hz	90	32,72236893	0,363581877
400 à 2000 Hz	78	30,6	0,392307692
2000 à 5000 Hz	115	53,13206893	0,462017991
5000 à 8000 Hz	118	46,75197	0,396203136
acima de 6000 Hz	114	50,63928	0,444204211
Fonema /e/			
60 à 400 Hz	90	25,11269	0,279029889
400 à 2000 Hz	78	19,8	0,253846154
2000 à 5000 Hz	102	27,8097	0,272644118
5000 à 8000 Hz	120	65,55198	0,5462665
acima de 6000 Hz	118	44,781552	0,379504678
Fonema /i/			
60 à 400 Hz	73	6,230806	0,085353507
400 à 2000 Hz	77	0,724016	0,009402805
2000 à 5000 Hz	99	36,91065	0,372834848
5000 à 8000 Hz	111	37,53498	0,338152973
acima de 6000 Hz	114	44,53498	0,390657719

Estas bandas espectrais foram consideradas, para verificar os efeitos da alteração de fonemas (configuração do trato supraglotal e nível de tensão nas pregas vocais). Apesar da quantidade de discriminações obtida pelos três fonemas ser bastante similar, o nível de discriminação sofre influência das características de cada vogal. Pode-se constatar que os fonemas /a/ e /e/ fornecem os melhores resultados.

Considerando as bandas espectrais empregadas, verifica-se que as duas primeiras (entre 60 e 400 Hz e entre 400 e 2000 Hz) fornecem resultados pouco eficazes na discriminação patológica pois a maioria das disfonias analisadas são, perceptualmente, roucas ou soprosas. Assim, espera-se que a energia de alta frequência aumente pela presença de ruído turbulento no sinal de voz. Como pode-se verificar, os intervalos acima de 5000 Hz fornecem os melhores resultados.

A banda espectral acima de 6000 Hz, sugerida por Shoji *et al* (1992-1), mostra-se eficiente quando utiliza-se o fonema /a/, enquanto que a banda entre 5000 e 8000 Hz produz melhores resultados se aplicada às vogais /e/ e /i/. Isto indica que a avaliação da energia de alta frequência na curva de densidade espectral de potência fornece informações úteis para a identificação de patologias.

4.2.4 Medidas de Soprosidade - BR

Empregando a técnica de estimação de soprosidade sugerida por Fukazawa, El-Assooty e Honjo (1988), determinou-se a eficácia do parâmetro Br (*Breathiness Ratio*) na identificação de patologias. Os resultados são apresentados na tabela abaixo, onde são considerados também o emprego dos fonemas /a/, /e/ e /i/.

Tabela 61 - Análise estatística da medida BR

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Fonema /a/	113	54,44896	0,481849204
Fonema /e/	129	40,64896	0,315108217
Fonema /i/	125	40,8405456	0,326724365

Claramente percebe-se que este parâmetro acústico, quando aplicado sobre o fonema /e/, fornece um maior número de disfonias identificadas, apesar do fonema /a/ garantir a maior distância entre as distribuições de patologias. Se comparado às técnicas de estimação de energia espectral, a extrema simplicidade de cálculo juntamente com a distinção de um bom número de pares patológicos viabiliza o seu emprego na identificação de tipos de voz.

Pela tab. 61, apesar das diferenças constatadas, pode-se considerar que os resultados são praticamente independentes dos fonemas considerados, mostrando-se na prática que esta medida acústica independe da tensão aplicada sobre as pregas vocais e alterações na configuração das estruturas supraglotais, particularmente da língua.

4.2.5 Medidas de Aspreza - SNL

Seguindo a técnica proposta por Sansone e Emanuel (1970), estimou-se a medida SNL a partir da energia mínima nos trechos com 100 Hz de largura, em cinco diferentes intervalos de frequência, para verificar a sua capacidade de discriminação das disfonias analisadas. Os cálculos foram desenvolvidos sobre os fonemas /a/, /e/ e /i/. O princípio envolvido é de que o aumento de ruído proveniente da vozes com características ásperas produzem elevação na magnitude das componentes não harmônicas.

Tabela 62 - Análise estatística da energia de ruído espectral (100 à 2600 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	123	68,75197	0,558959106
A regr. robusto	121	69,75197	0,576462562
Assimetria	128	52,27854	0,408426094
B regr. clássica	104	6,65916	0,064030385
B regr. robusto	76	2,82958	0,037231316
Curtose	124	56,27852	0,453859032
Desvio absoluto	87	4,32958	0,049765287
Desvio regr. clássica	101	41,20861	0,40800604
Desvio padrão	105	8,12958	0,077424571
Desvio regr. robusta	100	39,20861	0,3920861
Mediana	127	57,75197	0,454739921
Média	130	57,58155	0,442935

Pela tab. 62 pode-se notar que os parâmetros acústicos média e mediana da estimativa de nível de ruído espectral (SNL) discriminam o maior número de patologias. Comparando-se com outras medidas de ruído já apresentadas em seções anteriores, a distância entre as distribuições de disfonias

é pequena, mostrando que para o fonema /a/ nesta banda de análise, outros índices também apresentam resultados semelhantes aos já citados, ou seja, coeficiente angular (A) de regressão linear clássica e robusta, assimetria e curtose. Os demais parâmetros acústicos demonstram produzir valores que dificultam a discriminação patológica.

Tabela 63 - Análise estatística da energia de ruído espectral (2600 à 5100 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	115	29,83928	0,259472
A regr. robusto	114	26,83928	0,235432281
Assimetria	116	17,00968	0,146635172
B regr. clássica	114	18,66884	0,163761754
B regr. robusto	87	13,83926	0,159071954
Curtose	112	45,2882	0,404358929
Desvio absoluto	83	0,82958	0,00999494
Desvio regr. clássica	107	57,33098	0,535803551
Desvio padrão	99	6,32958	0,063935152
Desvio regr. robusta	107	58,33098	0,545149346
Mediana	126	32,83928	0,260629206
Média	127	33,83928	0,266451024

Utilizando outra banda de frequência para analisar o ruído espectral, pode-se perceber uma perda significativa na capacidade de discriminação das disfonias pelos cálculos estatísticos desenvolvidos sobre as estimativas de energia de ruído.

As medidas acústicas média e mediana apresentadas conseguem distinguir uma maior quantidade de pares patológicos apesar da redução na distância das distribuições de disfonias. Ambas medidas de desvio das regressões lineares (clássica e robusta) elevaram o nível e média das discriminações, mostrando que existe uma tendência linear de alteração energética ao longo do sinal de voz (5 segundos).

Como o fonema /a/ é caracterizado por uma concentração de energia abaixo dos 2000 Hz, pode-se esperar que ocorra uma diminuição na capacidade de identificação de disfonias pois a existência de uma quantidade natural de ruído dificulta a separação entre vozes normais e patológicas, além do intervalo de variação entre estes dois tipos de voz (que permite a discriminação das disfonias) ficar reduzida.

Tabela 64 - Análise estatística da energia de ruído espectral (5100 à 8000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	115	26,751952	0,23262567
A regr. robusto	114	23,751952	0,208350456
Assimetria	117	26,81936	0,229225299
B regr. clássica	116	6,65916	0,057406552
B regr. robusto	105	5,82958	0,05551981
Curtose	121	48,1081	0,397587603
Desvio absoluto	103	3,5	0,033980583
Desvio regr. clássica	111	67,67365	0,609672523
Desvio padrão	122	26,25194893	0,215179909
Desvio regr. robusta	110	66,67365	0,606124091
Mediana	111	34,63926	0,312065405
Média	113	36,63928	0,324241416

Algumas alterações são sentidas com a utilização de uma faixa de altas frequências: o nível e a média das discriminações aumenta (com algumas exceções) enquanto que o número de pares distinguíveis em alguns parâmetros estatísticos, particularmente as medidas de curtose e desvios padrão e absoluto.

Considerando que não existem formantes significativas nesta região espectral (a terceira formante localiza-se próxima de 2500 Hz), esperar-se-ia dificuldades na identificação de disfonias. Entretanto, considerando que muitas patologias são caracterizadas pela elevação da energia de alta frequência (ruído proveniente de escapes glotais de ar), a variação desta energia ao longo do sinal de voz apresenta correlação com as disfonias.

Tabela 65 - Análise estatística da energia de ruído espectral (100 à 5100 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	132	55,58155	0,421072348
A regr. robusto	130	56,75197	0,436553615
Assimetria	108	37,27854	0,345171667
B regr. clássica	103	6,65916	0,064652039
B regr. robusto	89	9,62958	0,108197528
Curtose	113	60,27854	0,533438407
Desvio absoluto	81	1	0,012345679
Desvio regr. clássica	102	40,20861	0,394202059
Desvio padrão	109	9,32958	0,085592477
Desvio regr. robusta	101	43,20861	0,42780802
Mediana	137	52,78155	0,385266788
Média	140	53,78155	0,384153929

Considerando uma faixa de frequências maior na estimação da energia de ruído, percebe-se a influência da energia na região das formantes na discriminação patológica. As medidas acústicas média e mediana identificam um maior número de disfonias, suplantando inclusive os resultados obtidos na faixa entre 100 e 2600 Hz.

Entretanto, os efeitos de redução da distância entre as distribuições de disfonias proporcionados pelas não-formantes no intervalo entre 2600 e 5100 Hz ficam evidenciados nos resultados obtidos. A medida de curtose, entretanto, apresentou um maior distanciamento das distribuições de disfonias.

Tabela 66 - Análise estatística da energia de ruído espectral (100 à 8000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	127	64,75197	0,509858031
A regr. robusto	129	65,75197	0,509705194
Assimetria	113	47,27853893	0,41839415
B regr. clássica	108	6,65916	0,061658889
B regr. robusto	97	2,82958	0,029170928
Curtose	119	66,27854	0,556962521
Desvio absoluto	100	3,32958	0,0332958
Desvio regr. clássica	108	46,33098	0,428990556
Desvio padrão	109	13,12958	0,120454862
Desvio regr. robusta	108	47,33098	0,438249815
Mediana	128	66,58156	0,520168438
Média	128	67,58155	0,527980859

A utilização de quase toda a banda de frequência disponível (frequência de Nyquist para os sinais de voz estudados é de 11025 Hz), demonstra que a eficiência da identificação das disfonias não melhora significativamente, considerando o número de discriminações.

Verifica-se no entanto, que a confiabilidade nos pares separáveis é maior, apenas sendo superada pelo procedimento matemático realizado sobre a faixa de frequências entre 100 e 2600 Hz. As medidas que melhor realizaram a discriminação na tab. 66 foram as medidas de tendência central (Mediana e média) e o coeficiente angular (A) das regressões lineares (clássica e robusta). Cabe ressaltar que a medida de curtose mostrou uma elevação sistemática da distância das distribuições de patologias nos trechos espectrais apresentados.

Tabela 67 - Análise estatística da energia de ruído espectral (100 à 2600 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	113	37,83928	0,334860885
A regr. robusto	113	38,83928	0,343710442
Assimetria	85	86,85377	1,021809059
B regr. clássica	99	11,83926	0,119588485
B regr. robusto	89	10,27903	0,115494719
Curtose	90	98,18335	1,090926111
Desvio absoluto	88	8,95195	0,101726705
Desvio regr. clássica	98	57,66056	0,588373061
Desvio padrão	105	15,37185	0,146398571
Desvio regr. robusta	101	58,66056	0,580797624
Mediana	119	44,95197	0,377747647
Média	117	44,95197	0,384204872

A utilização do fonema /e/ (em que as pregas vocais estão com um nível de tensão superior ao do fonema /a/) reduz a performance de discriminação de patologias, com exceção das medidas de assimetria e curtose, que permitem a maior separação das distribuições das disfonias.

Os parâmetros acústicos Mediana e média produziram o maior número de discriminações dentre as estimativas realizadas sobre a faixa de frequência de 100 à 2600 Hz. As medidas de desvio das regressões lineares (clássica e robusta) aumentaram suas capacidades de discriminação em relação ao fonema /a/.

Tabela 68 - Análise estatística da energia de ruído espectral (2600 à 5100 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	107	35,5097	0,331866355
A regr. robusto	103	36,5097	0,354463107
Assimetria	86	5,579031	0,064872453
B regr. clássica	124	7,65916	0,061767419
B regr. robusto	94	0	0
Curtose	100	25,57903	0,2557903
Desvio absoluto	127	0	0
Desvio regr. clássica	107	64,37185	0,601606075
Desvio padrão	119	10,8	0,090756303
Desvio regr. robusta	109	61,37185	0,563044495
Mediana	125	44,5097	0,3560776
Média	129	45,5097	0,352788372

Na tab. 68, pode-se verificar o aumento no número de discriminações conseguido pelas avaliações estatísticas realizadas sobre a energia nesta banda de frequência. Aparentemente, as medidas de tendência central (média e mediana) mantiveram a distância das distribuições de patologias, mesmo com o aumento no número de pares distinguíveis. Este mesmo efeito pode ser notado nas duas medidas de desvio das regressões lineares. Os índices de simetria e curtose entretanto mostram-se ineficientes com a nova banda de frequência para o fonema /a/.

Tabela 69 - Análise estatística da energia de ruído espectral (5100 à 8000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	115	51,55197	0,448278
A regr. robusto	115	43,751972	0,38045193
Assimetria	89	34,93926	0,392575955
B regr. clássica	152	27,95195	0,183894408
B regr. robusto	142	16,66884	0,117386197
Curtose	108	55,93926	0,517956111
Desvio absoluto	143	19,95916	0,139574545
Desvio regr. clássica	107	57,55195	0,537868692
Desvio padrão	139	52,05195	0,37447446
Desvio regr. robusta	105	59,55195	0,567161429
Mediana	107	47,75197	0,446280093
Média	113	52,75197	0,466831593

Nas altas frequências, pode-se perceber uma melhora na performance discriminatória para as medidas realizadas sobre o fonema /e/, inclusive superando os resultados obtidos para o fonema /a/, apesar das medidas de desvio de regressão linear não seguirem esta tendência.

As medidas de dispersão, ou seja, desvio padrão e absoluto conseguem discriminar um bom número de disfonias, assim como o coeficiente B das regressão linear (clássica e robusta). Pode-se verificar que a uniformidade das distâncias das distribuições de patologias, o que permite considerar que ocorrem muitas oscilações durante o sinal de voz analisado.

Tabela 70 - Análise estatística da energia de ruído espectral (100 à 5100 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	104	36,122392	0,347330692
A regr. robusto	109	41,12239	0,377269633
Assimetria	91	30,73138	0,337707473
B regr. clássica	116	7,65916	0,066027241
B regr. robusto	91	0	0
Curtose	107	51,56096	0,481878131
Desvio absoluto	110	0	0
Desvio regr. clássica	113	62,84065893	0,556112026
Desvio padrão	122	30,37185	0,24894959
Desvio regr. robusta	113	64,67024	0,572303009
Mediana	117	49,95197	0,426939915
Média	116	50,95197	0,439241121

Utilizando uma faixa de frequências maior (englobando os três primeiros formantes da vogal /e/), as medidas que melhor identificam as patologias são as de tendência central, ou seja, média e mediana, juntamente com as avaliações de desvio de regressão linear (clássica e robusta).

Apesar disto, comparando os resultados obtidos com o fonema /a/, a performance degrada-se com a elevação da tensão na prega vocal, provavelmente pela maior “estabilidade” do sistema de vocalização.

Tabela 71 - Análise estatística da energia de ruído espectral (100 à 8000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	112	46,95197	0,419214018
A regr. robusto	111	47,95197	0,43199973
Assimetria	93	42,20861	0,453856022
B regr. clássica	129	5,82958	0,045190543
B regr. robusto	112	1,5	0,013392857
Curtose	109	60,380858	0,553952826
Desvio absoluto	127	8,82958	0,069524252
Desvio regr. clássica	115	59,39121	0,516445304
Desvio padrão	133	45,54227	0,342423083
Desvio regr. robusta	117	62,39121	0,533258205
Mediana	119	51,95198	0,436571261
Média	118	56,75198	0,480948983

Ampliando-se ainda mais a faixa espectral analisada, os cálculos estatísticos sobre as estimativas de energia de ruído conseguem melhor discriminar as disfonias, apesar dos resultados ainda serem inferiores aos obtidos para o fonema /a/. A melhora intra-segmentos deve-se a incorporação do trecho de alta frequência, que mostrou-se bem eficiente, pelos resultados mostrados na tab. 71.

Entretanto, pode-se considerar que o aumento da tensão nas pregas vocais provoque uma condição de estabilidade na produção do sinal de excitação da cavidade supraglotal que padroniza a energia espectral do ruído (em pessoas normais e disfônicas). A variabilidade das medidas que sustentaram melhores resultados na discriminação patológica para este fonema corrobora com este fato. Apesar disto, verifica-se novamente que as medidas de desvio das regressões lineares são mais compatíveis para discriminação dos tipos de voz..

Tabela 72 - Análise estatística da energia de ruído espectral (100 à 2600 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	117	50,53498	0,431922906
A regr. robusto	112	50,53498	0,451205179
Assimetria	104	44,53498	0,428220962
B regr. clássica	109	5,0563556	0,046388583
B regr. robusto	97	0	0
Curtose	107	37,5349756	0,350794164
Desvio absoluto	90	5,63009	0,062556556
Desvio regr. clássica	116	22,49417	0,193915259
Desvio padrão	114	15,6582756	0,137353295
Desvio regr. robusta	113	21,49417	0,190213894
Mediana	122	47,53498	0,389630984
Média	121	47,53498	0,392851074

Efetuando-se os cálculos sobre a vogal /i/, a quantidade de disfonias discrimináveis é maior quando se emprega as medidas de tendência central (nesta faixa de frequência), ou seja, média e mediana. O coeficiente angular (A) das regressões lineares (clássica e robusta) também apresenta bons resultados na identificação das patologias.

Entretanto, considerando os demais fonemas utilizados, a vogal /a/ mostra-se mais eficiente para a identificação de patologias, provavelmente pela concentração de energia espectral nesta faixa de análise.

Tabela 73 - Análise estatística da energia de ruído espectral (2600 à 5100 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	107	31,0068	0,289783178
A regr. robusto	112	31,0068	0,276846429
Assimetria	117	35,5068	0,303476923
B regr. clássica	100	3,02818	0,0302818
B regr. robusto	101	2,02818	0,02008099
Curtose	116	31,5068	0,271610345
Desvio absoluto	100	0	0
Desvio regr. clássica	112	35,59035	0,317770982
Desvio padrão	107	9,90386	0,092559439
Desvio regr. robusta	110	35,59035	0,323548636
Mediana	119	23,13689	0,194427647
Média	118	22,13689	0,187600763

Comparando-se os resultados da tab. 68 com os obtidos para esta faixa de frequência, verifica-se a perda de performance na discriminação patológica. Apenas as duas medidas de desvio das regressões lineares mantiveram similar capacidade de separação de pares disfônicos em relação à banda espectral anterior.

Aparentemente, esta faixa de frequência mostra-se ineficiente na avaliação da condição patológica do paciente, principalmente no requisito de distância entre distribuições dos tipos de voz, pois este item garante a confiabilidade das identificação das patologias.

Tabela 74 - Análise estatística da energia de ruído espectral (5100 à 8000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	117	52,96706	0,452709915
A regr. robusto	120	53,96706	0,4497255
Assimetria	99	37,5068	0,378856566
B regr. clássica	97	30,78036	0,317323299
B regr. robusto	105	34,7803512	0,33124144
Curtose	102	32,5068	0,318694118
Desvio absoluto	97	30,5068	0,314503093
Desvio regr. clássica	131	53,23182	0,406349771
Desvio padrão	121	32,53498	0,268884132
Desvio regr. robusta	132	53,231822	0,403271379
Mediana	111	45,59711	0,410784775
Média	111	50,59711	0,45582982

Notadamente os níveis e médias das discriminações para as medidas acústicas consideradas nas altas frequências garante uma maior confiabilidade na identificação dos tipos de voz. Os desvios das regressões lineares mostram-se mais capazes de identificar as disfonias, superando inclusive todos os índices acústicos calculados para todas as faixas de frequência analisadas sobre o fonema /i/.

Devido a uniformidade das avaliações do teste de hipótese (número, nível e média de discriminação), pode-se concluir que a variação dos parâmetros acústicos sobre esta faixa de

frequência é elevada (fato confirmado para os fonemas /a/ e /e/), sendo explicado pelo ruído de alta frequência encontrado na maioria das vozes disfônicas, em relação às normais.

Tabela 75 - Análise estatística da energia de ruído espectral (100 à 5100 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	116	58,1650656	0,501422979
A regr. robusto	111	58,16507	0,52400964
Assimetria	106	46,53498	0,439009245
B regr. clássica	97	6,05636	0,062436701
B regr. robusto	99	0	0
Curtose	113	39,408513	0,348747903
Desvio absoluto	84	1	0,011904762
Desvio regr. clássica	115	39,494212	0,34342793
Desvio padrão	111	22,56217	0,203262793
Desvio regr. robusta	115	37,09035	0,322524783
Mediana	124	59,19325	0,477364919
Média	123	56,19325	0,456855691

Confirmando os resultados apresentados pelos fonemas /a/ e /e/, a média e a mediana (tendência central) da energia espectral do ruído apresentam boa performance na identificação das patologias.

O emprego de uma banda com 5000 Hz de largura mostra que o emprego de medidas de energia mínima em segmentos de 100 Hz de comprimento é eficiente para a região espectral onde encontram-se as três formantes, ou seja, onde se concentra maior parte da energia para os fonemas estudados.

As estimativas do coeficiente angular (A) das regressões lineares (tanto clássica como robusta) também mostram bons resultados na discriminação patológica, semelhante aos resultados obtidos para o fonema /a/.

Tabela 76 - Análise estatística da energia de ruído espectral (100 à 8000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	114	52,1932456	0,457835488
A regr. robusto	118	53,1932456	0,450790217
Assimetria	103	48,5349756	0,471213355
B regr. clássica	84	17,56315	0,209085119
B regr. robusto	90	0,0281756	0,000313062
Curtose	110	49,8123656	0,452839687
Desvio absoluto	77	6	0,077922078
Desvio regr. clássica	121	52,60393	0,434743223
Desvio padrão	109	18,56217	0,170295138
Desvio regr. robusta	120	50,60393	0,421699417
Mediana	118	54,19325	0,459264831
Média	120	57,19325	0,476610417

Com o aumento da tensão na prega vocal, a distância entre as distribuições das disfonias reduzem-se, motivadas pela maior estabilidade das estruturas de vocalização. Simultaneamente, a

quantidade de tipos de voz discrimináveis também decresce. Excetuando-se as medidas do coeficiente B das regressões lineares e desvios absoluto e padrão, as demais medidas fornecem resultados similares na separação dos pares de tipos de voz.

Com estas análises, constata-se que as bandas que fornecem medidas com maior poder de discriminação são, nesta ordem, entre 100 e 2600 Hz e entre 5100 e 8000 Hz. A primeira deve-se a concentração das frequências formantes nesse intervalo, o que permite uma clara diferenciação entre vozes normais e as disfônicas. Isto corrobora as evidências de Sansone e Emanuel (1970) e outros pesquisadores, apesar de nesses estudos serem consideradas apenas patologias com características ásperas. A gama de tipos de voz estudadas neste trabalho envolve a maioria das disfonias conhecidas.

Este fato confirma também os resultados obtidos para a faixa de altas frequências, pois as vozes soprosas (em patologias como cistos, nódulos e pólipos) produzem ruído nesta região espectral, além de modificações nas componentes harmônicas.

Apesar destas evidências, verifica-se que a performance destas medidas é baixo se forem consideradas as análises realizadas sobre o primeiro e segundo harmônicos e a energia de segmentos da curva de densidade espectral de potência.

4.2.6 Medidas baseadas em Cepstrum

Nesta seção é apresentado algumas estimativas de relação sinal ruído baseadas nas estimativas cepstrais do sinal de voz. O aparecimento de um picos em *quefrecys* periódicas é útil na discriminação patológica, pois a sua magnitude é proporcional à extensão patológica e consequentemente, ao ruído provocado por perturbações da frequência fundamental e fluxo de ar turbulento na glote.

Além disto, utilizou-se a técnica de Krom (1993) para estimação da relação entre componentes harmônicas e ruído, pela filtragem dos picos de alta *quefrecy* e estimativa da quantidade de ruído remanescente. Os cálculos foram desenvolvidos sobre um conjunto de 5 bandas espectrais diferentes.

Tabela 77 - Análise estatística do pico cepstral proeminente (CPP) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	131	129,72801	0,990290153
A regr. robusto	126	122,72799	0,974031667
Assimetria	96	11,122372	0,115858042
B regr. clássica	85	0	0
B regr. robusto	93	0	0
Curtose	101	101,05128	1,000507723
Desvio absoluto	131	68,44896	0,522511145
Desvio regr. clássica	134	34,63928	0,25850209
Desvio padrão	117	37,01937893	0,316404948
Desvio regr. robusta	135	33,63928	0,249179852
Mediana	140	127,73767	0,912411929
Média	137	137,73767	1,005384453

As medidas de tendência central realizadas sobre o pico cepstral (CPP) mostram-se mais eficientes na discriminação patológica, com um número elevado de pares patológicos separáveis, além da confiabilidade ser garantida pelo nível e média de discriminação.

O coeficiente angular (A) das regressões lineares também apresenta resultado similar, representando o coeficiente angular de reta das estimativas de CPP ao longo do sinal de voz, variando conforme a disfonia. Apesar das medidas de desvio das regressões lineares conseguirem identificar um número de disfonias equivalente as demais avaliações estatísticas realizadas sobre os picos cepstrais, as distribuições das disfonias estão próximas.

Tabela 78 - Análise estatística da relação cepstral HNR (60 à 400 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	94	39,12239	0,416195638
A regr. robusto	80	34,12239	0,426529875
Assimetria	85	2,82958	0,033289176
B regr. clássica	121	112,0013389	0,9256309
B regr. robusto	115	110,0013589	0,956533556
Curtose	95	64,88494	0,682999368
Desvio absoluto	128	23,84896	0,18632
Desvio regr. clássica	143	94,02263	0,657500909
Desvio padrão	140	50,28824	0,359201714
Desvio regr. robusta	143	88,02263	0,615542867
Mediana	86	29,95197	0,348278721
Média	87	33,95197	0,390252529

Realizando a estimativa de energia através da supressão dos picos cepstrais correspondentes à estrutura harmônica do sinal de voz e realizando-se as avaliações estatísticas da energia de ruído sobre a faixa entre 60 e 400 Hz, verifica-se que as medidas de dispersão (desvio padrão, absoluto e das regressões lineares) mostram-se mais eficientes na identificação das disfonias, permitindo afirmar que a variação da energia ao longo do sinal permite avaliar as patologias.

Percebe-se principalmente que os desvios das regressões lineares (clássica e robusta) são as mais eficientes medidas para avaliação de anormalidades nas estruturas de vocalização, juntamente com o coeficiente B desta regressão.

Tabela 79 - Análise estatística da relação cepstral HNR (400 à 2000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	135	41,42238	0,306832444
A regr. robusto	144	42,62238	0,29598875
Assimetria	86	10,112692	0,117589442
B regr. clássica	92	6,62958	0,072060652
B regr. robusto	107	11,62958	0,108687664
Curtose	98	48,56774	0,495589184
Desvio absoluto	108	26,92238	0,249281296
Desvio regr. clássica	119	98,6737	0,829190756
Desvio padrão	112	63,05198	0,562964107
Desvio regr. robusta	117	98,6737	0,843364957
Mediana	135	24,33926	0,180290815
Média	131	35,33926	0,269765344

Considerando uma outra banda de frequência, onde localizam-se as duas primeiras frequências formantes para o fonema /a/, novas evidências são apresentadas. Primeiro, a perda de performance do coeficiente B das regressões lineares e das medidas de dispersão, com elevação da capacidade de discriminação das medidas de tendência central e coeficiente angular (A) das regressões lineares.

Tabela 80 - Análise estatística da relação cepstral HNR (2000 à 5000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	124	5,65916	0,045638387
A regr. robusto	118	2,82958	0,023979492
Assimetria	59	0	0
B regr. clássica	105	10,66884	0,101608
B regr. robusto	101	9,83926	0,097418416
Curtose	111	95,17362	0,85742
Desvio absoluto	106	2	0,018867925
Desvio regr. clássica	91	97,88086	1,075613846
Desvio padrão	113	73,890571	0,653898858
Desvio regr. robusta	93	97,88086	1,052482366
Mediana	109	6,65916	0,061093211
Média	106	4,65916	0,04395434

Nesta nova banda espectral, constata-se que a distância das distribuições das disfonias, comparada com as demais faixas de frequência, é pequena. Mesmo assim, os parâmetros de desvio das regressões lineares, apesar da menor quantidade de disfonias discrimináveis, mantém a mesma confiabilidade na identificação das disfonias, pois o nível de separabilidade dos pares patológicos é significativamente o mesmo.

Outra medida que apresenta bons resultados é a curtose das estimativas de relação harmônico-ruído ao longo do sinal de voz, com resultados similares aos obtidos para os índices de dispersão.

Tabela 81 - Análise estatística da relação cepstral HNR (5000 à 8000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	92	6,77185	0,073607065
A regr. robusto	84	5,77185	0,0687125
Assimetria	106	3,62958	0,034241321
B regr. clássica	97	11,66884	0,12029732
B regr. robusto	105	18,67852	0,177890667
Curtose	124	76,84062893	0,619682491
Desvio absoluto	105	10,65916	0,10151581
Desvio regr. clássica	108	97,48043	0,902596574
Desvio padrão	121	78,37741893	0,647747264
Desvio regr. robusta	107	95,31001	0,890747757
Mediana	102	7,11269	0,069732255
Média	108	9,11269	0,084376759

Calculando a energia nas altas frequências, confirma-se novamente a variação ao longo do sinal de voz da estimativa de energia de ruído, pois as medidas de desvio das regressões lineares são capazes de discriminar praticamente o mesmo número de disfonias, mantendo a mesma distância entre suas distribuições.

A capacidade destes parâmetros acústicos na identificação de patologias mostra a existência de similaridade entre os tipos de voz, motivada pela semelhança entre as curvas de densidade espectral de potência obtidas após o processamento matemático. A baixa amplitude dos picos cepstrais nas altas *quefrequency* confunde-se com as componentes não harmônicas (*rharmonics*) nesta faixa de frequência.

Tabela 82 - Análise estatística da relação cepstral HNR (60 à 2000 Hz) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	105	45,12239	0,429737048
A regr. robusto	95	46,95197	0,494231263
Assimetria	81	2,82958	0,034933086
B regr. clássica	116	109,99168	0,948204138
B regr. robusto	109	111,00136	1,018361101
Curtose	94	65,71452	0,699090638
Desvio absoluto	133	25,84896	0,194353083
Desvio regr. clássica	142	94,19305	0,663331338
Desvio padrão	136	58,28824	0,42859
Desvio regr. robusta	142	90,19305	0,635162324
Mediana	92	24,95195	0,271216848
Média	94	29,95195	0,318637766

Considerando a banda espectral sugerida por Krom (1993), constata-se que as medidas de dispersão apresentam melhor capacidade de discriminação, juntamente com o coeficiente B das regressões lineares. Estes resultados sugerem que a estimativa para o fonema /a/ deve ser realizada

sobre regiões espectrais onde predomina a maior parte da energia acústica, ou seja, onde os principais formantes estão localizados. Isto fica evidenciado pela baixa performance das estimativas quando se considera a energia nas frequências acima de 2000 Hz.

Outra confirmação é que as medidas acústicas de desvio de regressão linear mantêm a mesma capacidade nas cinco bandas espectrais consideradas. Isto se confirma pela variabilidade das estimativas de relação harmônico-ruído permitir discriminar as anormalidades nas estruturas de vocalização.

A utilização da magnitude do pico cepstral na avaliação patológica para o fonema /a/ mostra-se eficiente, fornecendo resultados muito mais confiáveis que as estimativas de relação harmônico-ruído.

Tabela 83 - Análise estatística do pico cepstral proeminente (CPP) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	135	93,72795893	0,694281177
A regr. robusto	135	97,72795893	0,723910807
Assimetria	70	2,00968	0,028709714
B regr. clássica	91	13,00183	0,142877253
B regr. robusto	84	3,82958	0,045590238
Curtose	105	74,72799	0,711695143
Desvio absoluto	127	61,26886	0,482431969
Desvio regr. clássica	88	16,66884	0,189418636
Desvio padrão	131	45,66886	0,348617252
Desvio regr. robusta	91	14,66884	0,161196044
Mediana	137	103,06095	0,752269708
Média	137	99,06095	0,723072628

Extraíndo-se as medidas de pico proeminente do cepstrum ao longo do fonema /e/, pode-se notar uma redução da capacidade de discriminação com o aumento da tensão na prega vocal, pela redução da distância entre as distribuições das disfonias no espaço unidimensional. As medidas de tendência central (média e mediana) conseguem identificar um maior número de tipos de voz, seguidas pelo coeficiente angular (A) estimados para as regressões lineares, verificando-se então a manutenção dos resultados (ainda que inferiores) obtidos no fonema /a/.

Tabela 84 - Análise estatística da relação cepstral HNR (60 à 400 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	96	22,83928	0,237909167
A regr. robusto	98	22,83928	0,233053878
Assimetria	92	9,00968	0,097931304
B regr. clássica	91	2	0,021978022
B regr. robusto	91	1	0,010989011
Curtose	112	53,75198	0,479928393
Desvio absoluto	106	5,942272	0,05605917
Desvio regr. clássica	103	55,391212	0,537778757
Desvio padrão	92	1,82958	0,019886739
Desvio regr. robusta	102	49,391212	0,484227569
Mediana	99	36,83928	0,372113939
Média	100	37,83928	0,3783928

Os efeitos de perda de performance pela utilização de técnicas cepstrais sobre condições de maior tensão da prega vocal também aparecem na estimativa de relação entre componentes harmônicas e não harmônicas, principalmente nas medidas de coeficiente B (angular) da regressão.

Os índices de desvio de regressão linear (clássica e robusta) apresentam-se como os melhores parâmetros para identificação das disfonias nesta faixa de frequência para o fonema /e/. Verifica-se que o coeficiente B das regressões lineares, juntamente com o desvio padrão fornecem resultados muito pouco confiáveis, pela pequena distância entre as distribuições, mostrando que o coeficiente linear (*offset*) da reta para cada disfonia estão muito próximas.

Tabela 85 - Análise estatística da relação cepstral HNR (400 à 2000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	131	52,91269	0,403913664
A regr. robusto	137	54,25536	0,396024526
Assimetria	95	14,83926	0,156202737
B regr. clássica	131	52,26883893	0,39899877
B regr. robusto	117	35,25916	0,301360342
Curtose	115	49,38153	0,429404609
Desvio absoluto	120	60,56166	0,5046805
Desvio regr. clássica	112	127,84252	1,141451071
Desvio padrão	122	82,842518	0,679037033
Desvio regr. robusta	112	127,84252	1,141451071
Mediana	93	5	0,053763441
Média	113	19,39172	0,171608142

Os resultados obtidos para esta nova faixa de frequência (que contém as principais frequência formantes da vogal /e/) mostram que as distribuições entre as disfonias estão mais separadas. Apesar dos coeficientes (A e B) das regressões lineares permitirem discriminar um maior número de disfonias, o nível de separabilidade obtido não supera os resultados obtidos com os desvios de regressão linear (tanto clássica como robusta).

A confiabilidade na distinção das patologias destas medidas de dispersão são superiores aos obtidos com o fonema /a/, indicando que o intervalo de variação das estimativas de ruído pela técnica cepstral são maiores que naquela vogal.

Tabela 86 - Análise estatística da relação cepstral HNR (2000 à 5000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	123	29,8	0,242276423
A regr. robusto	117	33,8	0,288888889
Assimetria	76	10,94786893	0,144050907
B regr. clássica	107	14,48874	0,135408785
B regr. robusto	108	9,65916	0,089436667
Curtose	110	66,832832	0,6075712
Desvio absoluto	117	25,78153	0,220354957
Desvio regr. clássica	102	127,8522089	1,253453029
Desvio padrão	117	76,76354	0,656098632
Desvio regr. robusta	104	125,85224	1,210117692
Mediana	114	23,8	0,20877193
Média	112	24,8	0,221428571

Nesta banda espectral, as medidas de desvio das regressões lineares (clássica e robusta) mantém a mesma capacidade de discriminação de patologias obtida na faixa de frequência entre 400 e 2000 Hz, apesar da redução do número de pares distinguíveis. As medidas realizadas neste intervalo de frequência para o fonema /a/ fornecem distribuições de patologias que estão mais distantes entre si, em relação fonema /e/.

Os desvios das regressões lineares (tanto robusta quanto clássica) fornecem os resultados mais confiáveis na identificação das patologias, confirmando os resultados já obtidos para o fonema /a/.

Tabela 87 - Análise estatística da relação cepstral HNR (5000 à 8000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	92	5,82958	0,063365
A regr. robusto	98	7,17225	0,073186224
Assimetria	92	74,831	0,813380435
B regr. clássica	101	54,61938	0,540785941
B regr. robusto	107	46,61938	0,43569514
Curtose	101	51,17365	0,506669802
Desvio absoluto	102	3,82958	0,037544902
Desvio regr. clássica	109	91,44896	0,838981284
Desvio padrão	114	54,44898	0,477622632
Desvio regr. robusta	109	87,44896	0,802284037
Mediana	92	8,82958	0,095973696
Média	89	9,82958	0,110444719

A redução do intervalo de variação das medidas acústicas ao longo do sinal de voz para um intervalo de frequências entre 5000 à 8000 Hz pode ser sentida pela diminuição no número de discriminações e distância entre as distribuições de patologias.

A discriminação patológica é melhor realizada quando se emprega as medidas de desvio das regressões lineares, fato este que aparenta demonstrar a existência de uma tendência linear dos valores obtidos para as disfonias. Assim, quanto mais esparsos os valores em relação a esta reta, maior a probabilidade de detecção de disfonias mais graves.

Tabela 88 - Análise estatística da relação cepstral HNR (60 à 2000 Hz) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	87	3,83925893	0,044129413
A regr. robusto	85	3,83925893	0,045167752
Assimetria	86	5	0,058139535
B regr. clássica	96	7	0,072916667
B regr. robusto	97	3	0,030927835
Curtose	110	57,75197	0,525017909
Desvio absoluto	132	16,94227	0,12835053
Desvio regr. clássica	102	52,38153	0,513544412
Desvio padrão	113	11,78153	0,104261327
Desvio regr. robusta	99	54,38153	0,549308384
Mediana	93	22,83928	0,245583656
Média	102	25,83928	0,253326275

Considerando a faixa de frequência sugerida por Krom (1993) nota-se que a performance foi reduzida em função da utilização de uma vogal (/e/) que exige uma maior tensão da prega vocal do que o fonema /a/.

Durante as análises realizadas neste fonema, pode-se notar que as medidas que mais propiciaram identificar as disfonias foram os desvios das regressões lineares clássica e robusta.

Em alguns casos, foi possível identificar as patologias estudadas através de medidas de tendência central (média e mediana) e os coeficientes A e B das regressões lineares. Entretanto, estas devem ser consideradas como acontecimentos isolados, pois a tendência não se manteve entre os fonemas.

Tabela 89 - Análise estatística do pico cepstral proeminente (CPP) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	144	111,20363	0,772247431
A regr. robusto	147	113,077163	0,769232401
Assimetria	118	25,9580686	0,219983632
B regr. clássica	68	3,02818	0,044532059
B regr. robusto	70	2,02818	0,028974
Curtose	105	27,93888	0,266084571
Desvio absoluto	142	78,0713456	0,549798208
Desvio regr. clássica	113	21,90168	0,193820177
Desvio padrão	121	48,81237	0,403408017
Desvio regr. robusta	113	21,90168	0,193820177
Mediana	155	118,7296512	0,76599775
Média	158	120,7296512	0,764111716

Utilizando avaliações estatísticas sobre os picos cepstrais (representando as componentes harmônicas) percebe-se que o emprego do fonema /i/ permite que um conjunto maior de pares disfônicos sejam identificáveis. As medidas de tendência central oferecem os melhores resultados, confirmando serem confiáveis estas estimativas estatísticas pelo nível e média da discriminação.

As medidas de coeficiente angular A das regressões lineares (clássica e robusta) também confirmaram sua capacidade de identificação de patologias para este fonema. Estas evidências mostram que a medida de pico cepstral (considerando a tendência central ao longo do sinal de voz) apresenta bons resultados no diagnóstico de disfonias.

Tabela 90 - Análise estatística da relação cepstral HNR (60 à 400 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	114	30,5068	0,267603509
A regr. robusto	119	38,91066	0,326980336
Assimetria	98	29,13689	0,297315204
B regr. clássica	66	0	0
B regr. robusto	76	0	0
Curtose	121	32,59714	0,269397851
Desvio absoluto	99	3,02818	0,030587677
Desvio regr. clássica	114	23,56892	0,206744912
Desvio padrão	84	5,63009	0,067024881
Desvio regr. robusta	113	21,56892	0,190875398
Mediana	113	25,53498	0,225973274
Média	112	21,5349756	0,192276568

Comparativamente aos resultados obtidos com o emprego dos fonemas /a/ e /e/, o aumento da tensão nas pregas vocais para produção deste fonema implicou em baixa distância entre as distribuições dos tipos de voz. Pode-se afirmar que a confiabilidade das medidas realizadas para a vogal /i/ neste intervalo de frequência são inferiores aos demais fonemas, apesar do número de discriminações entre estes ser equiparável.

As medidas de coeficiente angular A das regressões lineares foram as que melhor separam os pares disfônicos, apesar das considerações aqui apresentadas.

Tabela 91 - Análise estatística da relação cepstral HNR (400 à 2000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	105	27,4320756	0,261257863
A regr. robusto	114	31,4602556	0,275967154
Assimetria	87	22,80776	0,262158161
B regr. clássica	95	28,31455	0,298047895
B regr. robusto	107	20,31455	0,189855607
Curtose	111	21,19325	0,19093018
Desvio absoluto	114	50,5408	0,443340351
Desvio regr. clássica	134	68,59716	0,511919104
Desvio padrão	127	67,56898	0,532039213
Desvio regr. robusta	134	67,5971556	0,504456385
Mediana	104	21,4039	0,205806731
Média	105	30,4320756	0,289829291

A troca de banda espectral resulta, para o fonema /i/ no aumento da distância entre as distribuições dos tipos de voz estudados, além da elevação no número de pares discrimináveis pelas estimativas estatísticas na relação de energia entre as componentes harmônicas e o ruído. As medidas de desvios mostram-se as mais eficazes (dentre os demais parâmetros determinados nesta faixa de frequência) para identificação das patologias.

Pode-se verificar que não existe uma relação entre o incremento da tensão nas pregas vocais e o número ou nível de discriminação obtido, pois os parâmetros extraídos do fonema /e/ que garantiram uma maior confiabilidade na identificação das patologias.

Tabela 92 - Análise estatística da relação cepstral HNR (2000 à 5000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	107	5	0,046728972
A regr. robusto	114	8,63009	0,075702544
Assimetria	89	0	0
B regr. clássica	66	0	0
B regr. robusto	93	0	0
Curtose	116	45,7000356	0,393965824
Desvio absoluto	93	12,136876	0,130504043
Desvio regr. clássica	106	35,16507	0,331745943
Desvio padrão	91	31,13689	0,342163626
Desvio regr. robusta	106	35,16507	0,331745943
Mediana	90	5,40386	0,060042889
Média	86	3	0,034883721

Considerando os resultados obtidos na tab. 92, verifica-se novamente a redução da performance de identificação das disfonias. As medidas com maior capacidade de discriminação são os desvios das regressões lineares e a medida de curtose.

Os resultados da discriminação para este fonema são inferiores aos obtidos tanto para a vogal /a/ quanto a vogal /e/. Percebe-se que muitas das patologias identificáveis possuem distribuições

extremamente próximas, mostrando que esta faixa de frequência não é confiável para discriminação patológica.

Tabela 93 - Análise estatística da relação cepstral HNR (5000 à 8000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	76	0	0
A regr. robusto	78	0	0
Assimetria	97	73,82791	0,761112474
B regr. clássica	86	15,5068	0,180311628
B regr. robusto	117	27,93884	0,238793504
Curtose	121	65,2317856	0,539105666
Desvio absoluto	138	37,40851	0,271076159
Desvio regr. clássica	116	58,47385	0,504084914
Desvio padrão	113	49,94348	0,441977699
Desvio regr. robusta	117	57,47385	0,491229487
Mediana	83	0	0
Média	87	0,0281756	0,000323857

Analisando-se a energia do ruído de alta frequência, em relação à faixa espectral analisada anteriormente, os resultados apresentaram uma melhora estatisticamente significativa, principalmente na distância entre as distribuições de disfonias, sendo que este evento também ocorreu nos demais fonemas.

Percebe-se que as medidas de desvio das regressões lineares apresentam bons resultados na discriminação (fato que também ocorre para demais fonemas), além do parâmetro de curtose.

Verifica-se que o maior número de discriminações ocorre para o índice desvio padrão, apesar do nível de separabilidade das disfonia não ser muito elevado. Deve-se considerar cuidadosamente esta medida pois pode ser um resultado eventual que com o emprego de um maior número de sinais de voz à analisar resulte na redução do número de pares patológicos distinguíveis.

Tabela 94 - Análise estatística da relação cepstral HNR (60 à 2000 Hz) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	109	24,5068	0,224833028
A regr. robusto	108	26,5068	0,245433333
Assimetria	88	10,54074	0,119781136
B regr. clássica	67	0	0
B regr. robusto	71	0	0
Curtose	125	55,59715	0,4447772
Desvio absoluto	106	0	0
Desvio regr. clássica	113	16,5689156	0,146627572
Desvio padrão	78	0	0
Desvio regr. robusta	115	12,5689156	0,109294918
Mediana	103	20,5068	0,199095146
Média	102	18,5068	0,181439216

Considerando esta banda espectral, nota-se que muitas medidas acústicas não são confiáveis para discriminação patológica. As medidas de dispersão mostram pela distância entre as distribuições

de disfonias que o aumento da tensão sobre as pregas vocais resulta em uma estabilidade maior na função de vocalização na glote, resultando na pouca diferença entre ruído de soprosidade e componentes harmônicas nesta região espectral. Somente a medida de curtose apresentou resultados satisfatórios para a identificação das disfonias.

Pode-se verificar que o emprego da técnica cepstral para avaliação da relação entre componentes harmônicas e ruído turbulento na curva de densidade espectral de potência apresenta bons resultados quando realizada sobre o fonema /a/. Neste caso, é útil considerar as faixas de frequência entre 60 à 400 Hz e 60 à 2000 Hz, devido a concentração de energia e posicionamento das formantes nestas bandas.

A estimativa de pico cepstral proeminente demonstrou ser muito eficiente na discriminação patológica para todos os fonemas apresentados, principalmente no fonema /i/ devido as características de estabilidade já mencionadas.

4.2.7 Medida de Relação Harmônica-Ruído HNR

Para permitir a identificação de patologias, foi considerado o emprego da técnica sugerida por Yumoto *et al* (1982). Através da determinação de um segmento periódico “médio” de um sinal vocálico (teoricamente apresentando apenas componentes harmônicas), pode-se determinar a quantidade de ruído existente na voz disfônica e assim, identificar a patologia correspondente.

Através do teste estatístico não-paramétrico, utilizado sobre todas as combinações (dois à dois) de tipos de voz, avalia-se a capacidade de discriminação dos índices acústicos baseados nesta estimativa.

Tabela 95 - Análise estatística da medida HNR para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	127	61,12423	0,48129315
A regr. robusto	121	61,12423	0,505158926
Assimetria	84	0	0
B regr. clássica	111	5,65916	0,050983423
B regr. robusto	103	5,65916	0,054943301
Curtose	100	64,93926	0,6493926
Desvio absoluto	106	7,34267	0,069270472
Desvio regr. clássica	110	59,69172	0,542652
Desvio padrão	121	19,75536	0,163267438
Desvio regr. robusta	110	59,69172	0,542652
Mediana	128	64,89465	0,506989453
Média	129	61,094648	0,473601922

As medidas estatísticas aplicadas sobre as estimativas de HNR realizadas ao longo do sinal de voz indicam que os parâmetros de tendência central (média e mediana) possibilitam identificar um maior número de disfonias, juntamente com as estimativas do coeficiente angular (A) das regressões

lineares (tanto clássica como robusta). Os índices de desvio das regressões lineares também oferecem uma confiabilidade na identificação das disfonias, juntamente com a medida de curtose.

Comparando os resultados deste método com os resultados obtidos para medida de HNR no domínio cepstral, nota-se que esta última é mais eficiente tanto em quantidade de pares disfônicos distinguíveis quanto em nível de discriminação, além da menor exigência de processamento computacional no seu cálculo.

Tabela 96 - Análise estatística da medida HNR para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	103	58,75197	0,570407476
A regr. robusto	101	51,75197	0,512395743
Assimetria	79	8,5	0,107594937
B regr. clássica	88	5,65916	0,064308636
B regr. robusto	91	3,82958	0,042083297
Curtose	104	56,1	0,539423077
Desvio absoluto	100	43,3081	0,433081
Desvio regr. clássica	117	84,88155	0,725483333
Desvio padrão	87	76,499841	0,879308517
Desvio regr. robusta	116	88,38155	0,761909914
Mediana	117	59,75197	0,510700598
Média	117	59,75197	0,510700598

Com o emprego do fonema /e/, pode-se perceber que as medidas de desvio das regressões lineares melhoram a performance na discriminação dos tipos de voz, principalmente pelo aumento da distância entre as distribuições de disfonias. Outra evidência é a redução da capacidade de distinção dos padrões patológicos pelas medidas de tendência central e coeficiente angular das regressões lineares.

Estes dois eventos mostram que o valores médios de HNR das vozes analisadas (fonema /e/) são muito próximos, e que a variabilidade ao longo do sinal de voz é uma importante característica de distinção entre as patologias.

Tabela 97 - Análise estatística da medida HNR para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	95	42,13689	0,443546211
A regr. robusto	94	44,13689	0,469541383
Assimetria	75	12,506786	0,166757147
B regr. clássica	113	1,0563512	0,009348241
B regr. robusto	117	13,68645	0,116978205
Curtose	107	32,0621656	0,299646407
Desvio absoluto	88	6,03497	0,068579205
Desvio regr. clássica	115	48,8990876	0,425209457
Desvio padrão	95	31,46705	0,331232105
Desvio regr. robusta	115	50,8990956	0,442600831
Mediana	101	45,5971032	0,451456467
Média	106	45,5971056	0,430161374

A utilização da vogal /i/ produz um decaimento na performance discriminatória das avaliações estatísticas realizadas sobre a estimativa da relação harmônico-ruído. Apesar das medidas de desvio das regressões lineares (clássica e robusta) ainda mostrarem-se mais eficientes nesta tarefa, pode-se considerar que por possuir componentes de alta frequência em maior número que os fonemas /a/ e /e/ (a segunda e terceira formantes localizam-se a partir de 2000 Hz), estas influenciam negativamente nos resultados de discriminação pois vozes patológicas são caracterizadas pela presença de grande quantidade de ruído de alta frequência. Com a perda de magnitude das componentes formantes e simultâneo aumento da energia de ruído de alta frequência, a separação entre vozes normais e as disfônicas fica comprometida.

Pode-se verificar que o emprego do fonema /e/, juntamente com medidas de desvio de regressão linear, oferece melhores resultados quando este método é desenvolvido, apesar do fonema /a/ também possuir similar capacidade discriminatória. Cabe citar também que o método HNR por técnica cepstral (uma solução alternativa para este tipo de cálculo) aparentemente fornece melhores resultados com menor esforço computacional.

4.2.8 Medida de Relação Harmônica - HR

Outra forma de avaliar a quantidade de ruído turbulento, provocado pelo aumento de escape de ar pelas pregas vocais, é quantificar a relação de energia entre os picos harmônicos e as componentes não harmônicas entre estas proeminentes magnitudes.

As tabelas abaixo mostram os resultados da capacidade de discriminação dos tipos de voz utilizando avaliações estatísticas sobre as estimações de HR ao longo de cada sinal de voz analisado.

Tabela 98 - Análise estatística da medida HR para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	123	58,6194	0,476580488
A regr. robusto	109	60,6194	0,556141284
Assimetria	69	1	0,014492754
B regr. clássica	86	16,0097	0,186159302
B regr. robusto	74	21,6097	0,292022973
Curtose	98	82,56163	0,842465612
Desvio absoluto	105	16,12239	0,153546571
Desvio regr. clássica	120	49,831	0,415258333
Desvio padrão	121	20,95197	0,173156777
Desvio regr. robusta	118	49,831	0,42229661
Mediana	90	9,00968	0,100107556
Média	87	14,01936	0,161142069

As estimativas estatísticas desenvolvidas ao longo do sinal de voz mostram que os índices acústicos de coeficiente angular (A) e desvio das regressões lineares produzem os melhores resultados na tarefa de identificar as disfonias, conforme mostra a tab. 98.

O parâmetro acústico de curtose fornece a maior confiabilidade das estimativas pois a distância entre os pares patológicos é superior ao resultado obtido para as demais medidas.

Comparando os resultados destes índices acústicos com os obtidos para a medida HNR, percebe-se que a eficiência para o fonema /a/ é maior quando emprega-se aquela metodologia de cálculo para relação entre componentes harmônicas e não harmônicas (à exceção do nível de discriminação para a curtose).

Tabela 99 - Análise estatística da medida HR para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	109	27,0097	0,247795413
A regr. robusto	108	19,0097	0,176015741
Assimetria	79	4	0,050632911
B regr. clássica	68	2,82958	0,041611471
B regr. robusto	76	14,94227	0,196608816
Curtose	90	20,95196	0,232799556
Desvio absoluto	100	56,17364	0,5617364
Desvio regr. clássica	120	115,19303	0,959941917
Desvio padrão	114	84,17367	0,738365526
Desvio regr. robusta	122	114,19303	0,936008443
Mediana	99	20,0097	0,202118182
Média	96	20,0097	0,208434375

Os reflexos da utilização da vogal /e/ sobre os cálculos estatísticos desenvolvidos sobre este tipo de avaliação da condição soprosa dos pacientes são sentidos principalmente nas medidas de desvio de regressões lineares, onde ocorre um aumento de confiabilidade correspondente ao incremento do nível/média das discriminações.

Isto também é aplicado para medidas de dispersão como desvio padrão e desvio absoluto. Os demais parâmetros acústicos sofreram redução ou na quantidade de pares distinguíveis ou na distância entre as distribuições de patologias.

Tabela 100 - Análise estatística da medida HR para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	75	8,5	0,113333333
A regr. robusto	70	5,5	0,078571429
Assimetria	85	13,90386	0,163574824
B regr. clássica	98	6,43204	0,065633061
B regr. robusto	98	2,40386	0,024529184
Curtose	81	2,50679	0,030948025
Desvio absoluto	98	41,4456256	0,422914547
Desvio regr. clássica	114	44,44563	0,389873947
Desvio padrão	94	27,0417656	0,287678357
Desvio regr. robusta	115	45,44563	0,395179391
Mediana	77	10,90386	0,141608571
Média	74	12,90386	0,174376486

O emprego do fonema /i/ apresenta resultados muito inferiores na estimativa da quantidade de ruído turbulento na glote. Pode-se perceber que a distância entre as distribuições de patologias, assim como na quantidade de pares separáveis (em alguns parâmetros), apresentam resultados inferiores em relação as duas outras vogais analisadas.

As medidas com menor capacidade de identificar os tipos de voz estudados são as de tendência central, mais especificamente a média e mediana, além do coeficiente angular das regressões lineares.

O emprego desta técnica de avaliação de soproidade é indicada para o fonema /e/, sendo utilizado o desvio de uma das regressões lineares (clássica ou robusta), estimado ao longo do sinal de voz. Os resultados apresentam um grau de similaridade com os obtidos pelo índice HNR.

4.2.9 Medidas de Relação Harmônica - RA

Baseado na técnica apresentada na seção 3.3.7, determinou-se a eficiência da medida RA na discriminação patológica. Na composição do sinal à ser expandido (para refinamento da resolução espectral da transformada discreta de Fourier) foram considerados trechos com comprimentos iguais a três e seis períodos de frequência fundamental, marcados sincronamente. Os resultados dos testes de hipótese utilizando as medidas baseadas nestas estimativas da relação harmônica são apresentados abaixo, para os fonemas /a/, /e/ e /i/.

Tabela 101 - Análise estatística da medida RA com 3 períodos de fundamental para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	86	5	0,058139535
A regr. robusto	88	4	0,045454545
Assimetria	83	0	0
B regr. clássica	92	18,95197	0,205999674
B regr. robusto	90	9,95195	0,110577222
Curtose	101	19,1	0,189108911
Desvio absoluto	68	0	0
Desvio regr. clássica	72	59,88871	0,831787639
Desvio padrão	72	15,07903	0,209430972
Desvio regr. robusta	72	53,88871	0,748454306
Mediana	62	8	0,129032258
Média	66	7	0,106060606

Pode-se verificar que a quantidade de pares de tipos de voz distinguíveis e os níveis de discriminação são baixos. Algumas inclusive fornecem distribuições de patologias que estão muito próximas, favorecendo a ocorrência de erros de identificação. As medidas que proporcionam uma razoável discriminação das disfonias são os desvios de regressão linear (clássica e robusta), levando-se em conta o valor médio de separabilidade dos tipos de voz.

Com os resultados acima, e considerando outras medidas de avaliação do nível de ruído turbulento nas cavidade glotal, pode-se verificar que o emprego desta medida na discriminação patológica não é justificado, principalmente pelo esforço computacional exigido.

Tabela 102 - Análise estatística da medida RA com 6 períodos de fundamental para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	89	2	0,02247191
A regr. robusto	81	1	0,012345679
Assimetria	86	1	0,011627907
B regr. clássica	90	16,83928	0,187103111
B regr. robusto	82	9,66884	0,117912683
Curtose	101	26,10967893	0,258511673
Desvio absoluto	62	0	0
Desvio regr. clássica	76	54,88871	0,722219868
Desvio padrão	73	14,8	0,202739726
Desvio regr. robusta	71	49,88871	0,702657887
Mediana	66	7	0,106060606
Média	69	8	0,115942029

Mesmo com o emprego de trechos contendo seis períodos sucessivos de frequência fundamental na composição do sinal de análise, os resultados não se alteram significativamente. Verifica-se que as medidas de desvio das regressões lineares têm a melhor capacidade de discriminação de patologias, considerando todas as estimativas realizadas no teste de hipótese.

Tabela 103 - Análise estatística da medida RA com 3 períodos de fundamental para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	97	49,8097	0,513502062
A regr. robusto	90	52,8097	0,586774444
Assimetria	105	13,00968	0,123901714
B regr. clássica	104	8,83926	0,084992885
B regr. robusto	92	10,62958	0,115538913
Curtose	113	58,21269	0,515156549
Desvio absoluto	76	29,142668	0,383456158
Desvio regr. clássica	108	51,12958	0,473422037
Desvio padrão	97	42,97225	0,443012887
Desvio regr. robusta	109	56,12958	0,514950275
Mediana	97	58,3097	0,601130928
Média	92	45,8097	0,497931522

Com o emprego do fonema /e/, verifica-se uma melhora significativa nas avaliações das estruturas de vocalização. A quantidade de pares distinguíveis é semelhantes para todas as medidas acústicas realizadas sobre as estimativas da relação harmônica RA. Apesar do aumento da distância entre as distribuições de patologias, estas não superaram o melhor resultado obtido para o fonema /a/ (desvio das regressões lineares pela técnica clássica).

Tabela 104 - Análise estatística da medida RA com 6 períodos de fundamental para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	95	49,8097	0,524312632
A regr. robusto	89	53,8097	0,604603371
Assimetria	93	8	0,086021505
B regr. clássica	97	0,82958	0,008552371
B regr. robusto	90	2,82958	0,031439778
Curtose	120	66,042272	0,550352267
Desvio absoluto	73	34,14267	0,467707808
Desvio regr. clássica	101	49,12958	0,486431485
Desvio padrão	93	41,97225	0,451314516
Desvio regr. robusta	100	48,12958	0,4812958
Mediana	98	58,3097	0,594996939
Média	94	42,80968	0,455422128

Utilizando seis períodos da fundamental para avaliar a quantidade de ruído nas vogais /e/, percebe-se que a medida curtose apresenta a melhor performance na discriminação das patologias (tanto em número de discriminações quanto em separação das distribuições de patologias). Pode-se perceber pelas tabelas anteriores que este índice acústico não ocorreu isoladamente somente nesta tabela. Juntamente a este parâmetro acústico encontram-se as medidas de desvio das regressões lineares (clássica e robusta).

Tabela 105 - Análise estatística da medida RA com 3 períodos de fundamental para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	83	13,65827	0,16455747
A regr. robusto	88	10,65827	0,121116705
Assimetria	90	0	0
B regr. clássica	110	4,63009	0,042091727
B regr. robusto	90	0	0
Curtose	114	8,52818	0,074808596
Desvio absoluto	79	16,4602156	0,208357159
Desvio regr. clássica	107	36,93208	0,345159626
Desvio padrão	83	34,83594	0,41971012
Desvio regr. robusta	108	34,93208	0,323445185
Mediana	84	0,6582656	0,007836495
Média	75	0,0281756	0,000375675

Pela tab. 105, constata-se a diminuição da distância entre as distribuições das disfonias, reduzindo assim a confiabilidade das discriminações quando realizadas sobre o fonema /i/. Curiosamente observa-se que alguns parâmetros acústicos experimentaram um aumento no número de disfonias distinguíveis (curtose, ambos os desvios de regressão linear e coeficiente B das regressões lineares).

Tabela 106 - Análise estatística da medida RA com 6 períodos de fundamental para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	78	12,65827	0,162285513
A regr. robusto	89	14,65827	0,164699663
Assimetria	111	1	0,009009009
B regr. clássica	110	4,02818	0,036619818
B regr. robusto	99	5,02818	0,050789697
Curtose	115	5,52818	0,04807113
Desvio absoluto	94	17,864082	0,190043426
Desvio regr. clássica	111	29,335942	0,264287766
Desvio padrão	97	39,3641156	0,405815625
Desvio regr. robusta	113	29,33594	0,259610088
Mediana	86	2,6582656	0,030910065
Média	82	0,0281756	0,000343605

Verifica-se pelos resultados para o fonema /i/, quando aplica-se a técnica de estimação de ruído para trechos com comprimento igual à seis períodos da fundamental, um aumento na distância entre as distribuições de tipos de voz, diferente da tendência observada nos fonemas /a/ e /e/, aparentemente confirmando os efeitos negativos sobre a avaliação de ruído turbulento encontrados na estimativa HNR. As medidas mais confiáveis na identificação das disfonias são os índices de dispersão, ou seja, desvio padrão e desvios de regressão linear (clássica e robusta).

Os resultados obtidos pelo teste de hipótese mostraram que esta estimativa de quantidade de ruído turbulento é ineficaz na discriminação das disfonias, mesmo com a variação do comprimento

do trecho à ser expandido e emprego de técnicas híbridas na estimação da frequência fundamental (para marcação síncrona dos períodos de F_0).

Os parâmetros acústicos demonstram que os valores estimados ao longo do sinal de voz são muito similares entre as patologias, dificultando a discriminação patológica.

4.2.10 Medidas de Ruído no Sinal de Voz - NS

Com base na técnica sugerida por Muta *et al* (1988), estima-se a quantidade de ruído utilizando-se propriedades da transformada discreta de Fourier com quantidade de pontos variável conforme o período da fundamental. Empregando-se técnicas estatísticas (descritas na seção 3.2.9) sobre estas estimativas de ruído no sinal de voz, determina-se a performance discriminatória destes índices acústicos, onde os resultados são apresentados nos parágrafos seguintes.

Tabela 107 - Análise estatística da medida NS para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	121	54,95198	0,454148595
A regr. robusto	116	56,95198	0,490965345
Assimetria	115	37,240662	0,323831843
B regr. clássica	103	20,11829	0,195323204
B regr. robusto	104	20,11829	0,193445096
Curtose	108	72,164	0,668185185
Desvio absoluto	73	0	0
Desvio regr. clássica	103	47,44894	0,46066932
Desvio padrão	88	12,44893893	0,141465215
Desvio regr. robusta	103	45,44894	0,441251845
Mediana	129	58,95198	0,456992093
Média	127	56,95198	0,448440787

Observando os resultados obtidos na tab. 107, verifica-se que as medidas de tendência central (média e mediana), juntamente com o coeficiente angular (A) das regressões lineares possibilitam identificar um maior número de patologias com maior confiabilidade.

Entretanto, os melhores resultados são obtidos pela utilização da curtose sobre as estimativas de ruído (NS) ao longo do sinal de voz. Pode-se perceber que a performance discriminatória destes parâmetros acústicos é similar à obtida através da técnica HNR, já apresentada.

Tabela 108 - Análise estatística da medida NS para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	96	36,80968	0,383434167
A regr. robusto	100	33,80968	0,3380968
Assimetria	72	0	0
B regr. clássica	87	10,66884	0,122630345
B regr. robusto	84	11,83926	0,140943571
Curtose	112	79,00328	0,705386429
Desvio absoluto	98	49,87367993	0,508915101
Desvio regr. clássica	112	88,53286	0,790471964
Desvio padrão	98	78,70328	0,803094694
Desvio regr. robusta	112	87,53286	0,781543393
Mediana	119	34,8097	0,292518487
Média	117	37,8097	0,323159829

Com a utilização da vogal /e/, os resultados obtidos com as medidas de dispersão tem o número e nível/média de discriminação incrementados; em contrapartida, as medidas de tendência central tornam-se menos confiáveis. A medida de curtose também segue a tendência de aumento de confiabilidade.

Comparando os resultados obtidos pelas estimativas de ruído turbulento, pode-se perceber que o fonema /e/ é o mais adequado para este tipo de avaliação. Aparentemente a maior tensão da prega vocal (em relação à vogal /a/), juntamente com a configuração da cavidade de ressonância concentrando a energia acústica abaixo de 2000 Hz parecem influenciar positivamente na identificação das patologias.

Tabela 109 - Análise estatística da medida NS para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	125	65,3709156	0,522967325
A regr. robusto	124	51,34274	0,414054355
Assimetria	72	0	0
B regr. clássica	95	0	0
B regr. robusto	83	0	0
Curtose	125	44,32117	0,35456936
Desvio absoluto	114	30,634652	0,268725018
Desvio regr. clássica	118	76,203622	0,645793407
Desvio padrão	121	74,29297	0,613991488
Desvio regr. robusta	116	76,20362	0,656927759
Mediana	115	71,8709156	0,624964483
Média	111	66,8709156	0,602440681

Diferente das demais estimativas de ruído (HNR, HR e HNR por cepstrum), observa-se uma elevação na quantidade de pares disfônicos distinguíveis. Além disto, observa-se a manutenção das distâncias entre as distribuições de patologias. As medidas de tendência central e coeficiente A das regressões lineares mostram-se mais confiáveis com a elevação da tensão sobre as pregas vocais. O

emprego das medidas de dispersão, particularmente dos desvios das regressões lineares, oferecem os melhores resultados na discriminação patológica.

Percebe-se então que o emprego do fonema /i/ para avaliação das estruturas de vocalização para esta técnica de estimação de ruído garante que um bom número de patologias seja identificado de modo confiável.

4.2.11 Medida de Ruído no Sinal de Voz - NNE

A estimativa da energia de ruído aqui apresentada utiliza os intervalos entre as componentes harmônicas para determinar a quantidade de ruído espectral. Considerada com a estimação mais robusta do ruído turbulento (Kasuya *et al*, 1986), aplicou-se o teste estatístico Mann-Whitney para determinar sua capacidade discriminatória, obtendo-se assim a tabela abaixo.

Tabela 110 - Análise estatística da medida NNE

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Fonema /a/			
60 e 400 Hz	113	46,63928	0,412736991
400 e 2000 Hz	137	120,44898	0,879189635
2000 e 5000 Hz	123	116,73767	0,949086748
5000 e 8000 Hz	132	100,24105	0,759401894
1000 e 5000 Hz	136	121,94898	0,896683676
Fonema /e/			
60 e 400 Hz	122	62,25197	0,510262049
400 e 2000 Hz	96	43,61938	0,454368542
2000 e 5000 Hz	126	84,06101	0,667150873
5000 e 8000 Hz	130	77,71831	0,597833154
1000 e 5000 Hz	125	83,06098	0,66448784
Fonema /i/			
60 e 400 Hz	110	68,16969	0,619724455
400 e 2000 Hz	110	45,91066	0,417369636
2000 e 5000 Hz	139	106,0771556	0,763145004
5000 e 8000 Hz	135	78,79977	0,583702
1000 e 5000 Hz	141	124,0771556	0,879979827

A banda entre 1000 e 5000 Hz é sugerida por Kasuya *et al* (1986) como a de melhor capacidade na identificação de algumas patologias. Os resultados demonstram que estas considerações são corretas, apesar do intervalo de frequência entre 5000 e 8000 Hz mostrar melhores resultados para o fonema /e/ e as estimativas realizadas entre 400 e 2000 Hz para /a/ e 2000 e 5000 Hz para /i/ apresentarem capacidade de discriminação similar ao intervalo de 1000 à 5000 Hz.

A utilização da vogal /i/ mostra-se mais eficaz para identificar as patologias, pois simultaneamente distingue um maior número de pares de tipos de voz com maior confiabilidade (distância entre as distribuições de patologias).

Comparando-se estes resultados com outras estimativas de ruído ou relação harmônica-ruído, comprova-se a robustez desta técnica para a avaliação das condições da cavidade de vocalização (laringe e pregas vocais).

4.2.12 Outras medidas

Algumas medidas foram propostas para realizar a identificação de vozes normais e disfônicas, considerando a percepção de soproidade, aspereza e rouquidão. Mesmo não empregando avaliações diretamente sobre a curva de densidade espectral de potência, estas podem servir para avaliação simultânea da relação sinal-ruído e perturbação da frequência fundamental.

O primeiro parâmetro considerado é P/A (*peak to average ratio*), descrito na seção 3.3.13, e os resultados do teste de hipótese Mann-Whitney são apresentados nos parágrafos que seguem.

Tabela 111 - Análise estatística da relação P/A para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	84	10,62236893	0,126456773
A regr. robusto	85	10,62236893	0,124969046
Assimetria	84	18,34267	0,218365119
B regr. clássica	99	25,11826	0,253719798
B regr. robusto	98	11,10861	0,113353163
Curtose	94	31,90861	0,339453298
Desvio absoluto	105	22,9224	0,218308571
Desvio regr. clássica	123	95,22423	0,774180732
Desvio padrão	110	46,38156	0,421650545
Desvio regr. robusta	123	94,22423	0,76605065
Mediana	96	21,62237	0,225233021
Média	94	24,62237	0,261940106

Empregando a medida de relação entre pico e média das amostras da voz, verifica-se que a variabilidade da estimativa ao longo do fonema /a/ relaciona-se com extensão de anormalidades nas estruturas de vocalização. Isto é particularmente percebido pelos resultados de discriminação obtidos pelas medidas de desvio de regressão linear (clássica e robusta), além das medidas de dispersão (como desvio padrão e curtose).

Tabela 112 - Análise estatística da relação P/A para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	109	44,30968	0,406510826
A regr. robusto	109	47,30968	0,434033761
Assimetria	74	5	0,067567568
B regr. clássica	103	10,62958	0,103199806
B regr. robusto	103	14,62958	0,142034757
Curtose	100	43,0014	0,430014
Desvio absoluto	129	76,25197	0,591100543
Desvio regr. clássica	127	77,26166	0,608359528
Desvio padrão	123	78,25197	0,636194878
Desvio regr. robusta	127	77,26166	0,608359528
Mediana	108	60,30968	0,558422963
Média	110	62,30968	0,566451636

Utilizando-se o fonema /e/ para efetuar os cálculos estatísticos, percebe-se que o número de discriminações se eleva para quase todas as medidas acústicas empregadas, mesmo que a máxima distância entre as distribuições de disfonias somente seja conseguida no fonema /a/. Apesar disto, parâmetros como a média e mediana, além dos desvios padrão e absoluto mostram-se tão eficientes quanto as medidas de desvio das regressões lineares.

Aparentemente existe uma relação entre o aumento de tensão nas pregas vocais (com a produção do fonema /e/), que permite uma maior separabilidade das disfonias, motivada talvez pela redução do intervalo de variação das medidas acústicas (aumento da distância das patologias).

Tabela 113 - Análise estatística da relação P/A para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	101	61,81455	0,612025248
A regr. robusto	101	58,81455	0,582322277
Assimetria	118	1,02818	0,00871339
B regr. clássica	109	2,0563556	0,018865648
B regr. robusto	111	2,02818	0,018271892
Curtose	130	31,2214256	0,240164812
Desvio absoluto	114	81,44465	0,714426754
Desvio regr. clássica	127	117,50106	0,925205197
Desvio padrão	123	104,4728656	0,849372891
Desvio regr. robusta	126	116,50106	0,924611587
Mediana	94	65,81455	0,700154787
Média	94	65,81455	0,700154787

Comprova-se pelo emprego do fonema /i/ que o aumento da tensão nas pregas vocais permite que as medidas baseadas na relação P/A apresentem uma elevação na performance na identificação das patologias. Percebe-se, principalmente pelas medidas de dispersão, ou seja, desvios das regressões lineares (clássica e robusta), desvio padrão e desvio absoluto que os resultados são melhores que os fornecidos pelos fonemas /a/ e /e/.

O emprego das medidas de desvio de regressão linear sobre as estimativas da relação entre o pico e a valor médio retificado do sinal de voz é útil para discriminação patológica, pelo número de disfonias identificáveis e sua confiabilidade.

A medida RPK, baseada no coeficiente de Pearson, teve sua capacidade de discriminação dos tipos de voz testada. Os resultados são apresentados nas tabelas abaixo. Os resultados são equivalentes ao cálculo do coeficiente de correlação entre trechos do sinal de voz separados por um intervalo igual ao período da fundamental.

Tabela 114 - Análise estatística da relação RPK para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	132	62,58156	0,474102727
A regr. robusto	135	82,38156	0,610233778
Assimetria	124	62,75196	0,506064194
B regr. clássica	109	3,82958	0,035133761
B regr. robusto	122	9,48874	0,077776557
Curtose	122	53,74228	0,440510492
Desvio absoluto	121	61,92239	0,511755289
Desvio regr. clássica	117	69,55197	0,594461282
Desvio padrão	112	59,75197	0,533499732
Desvio regr. robusta	115	69,55197	0,604799739
Mediana	133	86,55198	0,650766767
Média	132	81,55197	0,617817955

As medidas de média e mediana, juntamente com a estimativa do coeficiente angular (A) das regressões lineares (clássica e robusta) permitem que se identifique com mais confiabilidade (além do maior número de discriminações) as patologias analisadas.

Comparando-se os resultados obtidos neste índice com os apresentados pela medida P/A, nota-se que a quantidade de disfonias distinguíveis é maior, apesar da distância entre as distribuições de patologias não acompanhar esta relação.

Tabela 115 - Análise estatística da medida RPK para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	116	38,63928	0,333097241
A regr. robusto	117	44,63928	0,381532308
Assimetria	117	21,9127	0,187288034
B regr. clássica	97	9,00968	0,092883299
B regr. robusto	95	0,8	0,008421053
Curtose	111	17,12237893	0,154255666
Desvio absoluto	114	41,75197	0,366245351
Desvio regr. clássica	113	54,76165	0,484616372
Desvio padrão	110	45,76165	0,416015
Desvio regr. robusta	112	55,76165	0,497871875
Mediana	112	51,75197	0,462071161
Média	112	49,75197	0,444214018

Já para o fonema /e/, percebe-se uma redução tanto no número quanto no nível das discriminações. Muitos parâmetros acústicos apresentam resultados de separabilidade bastante próximos entre si, como as medidas de tendência central e de dispersão.

Tabela 116 - Análise estatística da medida RPK para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
A regr. clássica	111	43,16507	0,388874505
A regr. robusto	117	51,1650656	0,437308253
Assimetria	128	35,19324	0,274947188
B regr. clássica	114	17,59133	0,154309912
B regr. robusto	118	50,5913356	0,428740132
Curtose	128	38,19324	0,298384688
Desvio absoluto	105	58,2929856	0,555171291
Desvio regr. clássica	107	61,2929856	0,572831641
Desvio padrão	107	58,2929856	0,544794258
Desvio regr. robusta	107	61,2929856	0,572831641
Mediana	114	62,56893	0,548850263
Média	109	56,56893	0,518981009

Entretanto, com o emprego dos procedimentos estatísticos sobre o fonema /i/, percebe-se um aumento no nível de separabilidade das distribuições de disfonias. Algumas medidas acústicas entretanto apresentam um decremento na quantidade de pares de tipos de voz.

A estimativa RPK sobre os sinais de voz permite determinar-se similaridade entre trechos periódicos do sinal de voz, e desta forma, avalia perturbações da frequência fundamental, apesar da influência redutiva do coeficiente de Pearson com o aparecimento de ruídos de alta frequência. Assim, como boa parte das patologias é caracterizada por serem perceptualmente identificáveis como soprosas, é de se esperar uma performance inferior na quantidade de patologias identificáveis, principalmente para uma condição de maior tensão nas pregas vocais providenciada pelo fonema /i/. Entretanto, as distâncias entre as distribuições de patologias é garantida pelas patologias cujas vozes são identificadas como roucas, como carcinomas e paralisias unilaterais. O aumento da estabilidade na produção da voz também auxilia na redução do intervalo de variação da voz normal, facilitando a discriminação entre padrões normais e patológicos.

4.3 Medidas Residuais

Para utilização das técnicas de filtragem inversa adaptativa, apresentadas na seção 3.4, é necessário definir a quantidade de coeficientes que serão estimados nos modelos paramétricos. Para os modelos AR, empregando os algoritmos de gradiente descendente e de valor mínimo médio quadrático recursivo, são utilizados 12 coeficientes ($P=12$). Já para o filtro de Kalman faz-se necessário utilizar uma combinação de coeficientes para a porção autorregressiva (P) e a porção exógena (N) do modelo ARX que dá suporte à esta técnica. As combinações testadas foram:

$$\{(P, N)\} = \{(1,1), (3,3), (6,4), (6,8), (8,4), (8,6), (12,6)\} \quad (117)$$

A abordagem de análise dos resultados da discriminação para as técnicas residuais é um pouco diferente do método até agora adotado. Espera-se determinar, principalmente, qual o algoritmo adaptativo é mais eficiente para identificar as disfonias, além da verificação de quais medidas acústicas e fonemas fornecem os melhores resultados.

4.3.1 Coeficiente de Excesso

A medida de coeficiente de excesso (baseada na curtose da distribuição de magnitudes do sinal residual) foi empregado sobre toda a extensão do sinal obtido pela filtragem inversa. Os valores obtidos por esta estimativa devem indicar que quanto maior a extensão patológica, mais platicúrtica (plana) deve ser a distribuição de magnitudes. Para vozes normais, deve-se obter uma distribuição leptocúrtica, ou seja, mais estreita. Os resultados obtidos pelo teste de hipótese utilizado são apresentados nas tabs. 117, 118 e 119.

Estas tabelas mostram a performance apresentada pelos algoritmos para cada medida acústica. Os filtros de Wiener (empregando algoritmo de gradiente e recursiva minimização do erro médio quadrático) e os filtros de Kalman foram considerados na análise. Para este último conjunto de filtros foram consideradas variações na quantidade de coeficientes à estimar, sendo que a combinação é apresentada da seguinte forma: o primeiro número apresentado compreende o número de coeficientes da porção autorregressiva do modelo ARX, enquanto o segundo valor corresponde à porção exógena.

Tabela 117 - Análise estatística do coeficiente de excesso (EX) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	109	83,26098	0,763862202
LMS	122	132,18335	1,083470082
Kalman 1/1	110	55,26101	0,502372818
Kalman 3/3	98	22,15235	0,226044388
Kalman 6/4	93	38,15238	0,410240645
Kalman 6/8	102	45,231401	0,443445108
Kalman 8/4	100	56,78198	0,5678198
Kalman 8/6	94	35,95237	0,382472021
Kalman 12/6	92	39,44107893	0,42870738

Pelos resultados apresentados quando utiliza-se o fonema /a/, percebe-se claramente que o algoritmo por estimação recursiva do erro médio quadrático mínimo (LMS) fornece a maior quantidade de discriminações, juntamente com a maior confiabilidade dentre as técnicas adaptativas consideradas, pois o nível e média de discriminação garantem tal conclusão.

O algoritmo por gradiente descendente também apresenta bons resultados na identificação de patologias, indicando que os modelos que não fazem uso da porção exógena são capazes de distinguir os pares de tipos de voz analisados.

Considerando o conjunto de coeficientes avaliados para o filtro de Kalman, percebe-se que não há um padrão bem definido que relacione a capacidade de discriminação com a quantidade de coeficientes utilizados. Apenas pode ser citado que o par 8/4 e 1/1 fornecem os melhores resultados, sendo que o primeiro fornece uma confiabilidade maior.

Tabela 118 - Análise estatística do coeficiente de excesso (EX) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	106	58,75197	0,554263868
LMS	110	93,56165	0,850560455
Kalman 1/1	100	43,92239	0,4392239
Kalman 3/3	100	26,00968	0,2600968
Kalman 6/4	106	29,0097	0,273676415
Kalman 6/8	118	59,75197	0,506372627
Kalman 8/4	89	24,92239	0,280026854
Kalman 8/6	106	28,95197	0,273131792
Kalman 12/6	100	12	0,12

O emprego da vogal /e/ permite que os algoritmos de Kalman melhorem a performance discriminatória. Neste caso, verifica-se que com o emprego de 6 coeficientes para a porção autorregressiva, consegue-se um número superior de pares disfônicos distinguíveis que para o fonema anterior (particularmente na configuração 6/8).

O algoritmo adaptativo LMS mantém-se como o mais eficaz para discriminar as disfonias, apesar da redução da capacidade discriminatória frente à vogal /a/. Percebe-se também que os algoritmos que se valem do modelo AR na estimativa do sinal residual tiveram a distância entre as disfonias reduzida, o mesmo ocorrendo para os algoritmos de Kalman que empregam maior quantidade de coeficientes para o modelo ARX.

Tabela 119 - Análise estatística do coeficiente de excesso (EX) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	109	57,67186	0,529099633
LMS	105	74,26942	0,70732781
Kalman 1/1	106	38,0136	0,358618868
Kalman 3/3	104	33,01358	0,317438269
Kalman 6/4	101	39,0136	0,386273267
Kalman 6/8	101	54,04177	0,53506703
Kalman 8/4	109	33,01359	0,302876972
Kalman 8/6	91	22,01359	0,241907582
Kalman 12/6	92	22,01359	0,239278152

Na tab. 119 percebe-se que a capacidade de distinção de pares disfônicos pelo índice acústico EX, quando emprega-se modelos AR, reduz-se ainda mais. Provavelmente o aumento de tensão nas pregas vocais produza um sinal residual cuja distribuição seja muito semelhante entre as disfonias, principalmente as que resultam em vozes soprosas. Mesmo assim, sua confiabilidade ainda é superior aos algoritmos de Kalman considerados.

Utilizando um modelo ARX percebe-se que o par 8/4 fornece um maior número de tipos de voz identificáveis, enquanto que o par 6/8 garante uma maior distância entre as distribuições das disfonias.

Considerando os resultados obtidos para o coeficiente de excesso (EX), pode-se afirmar que o emprego das técnicas adaptativas, desconsiderando a correlação existente no sinal residual (porção exógena), fornecem a melhor performance na discriminação, particularmente o algoritmo adaptativo por LMS.

4.3.2 Pico da Autocorrelação

A mediana da amplitude do segundo pico máximo de autocorrelação do sinal residual (que é estimado ao longo de suas amostras) também foi analisada estatisticamente para verificar sua performance discriminatória. São considerados os fonemas /a/, /e/ e /i/ na determinação deste índice acústico.

A variação desta medida indica os efeitos da patologia sobre as estruturas de vocalização, pois a existência perceptual de aspereza ou soproidade implica na redução da magnitude deste parâmetros, ou seja, diminuição do segundo pico máximo da autocorrelação.

Tabela 120 - Análise estatística do pico da autocorrelação (PA) para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	121	115,63096	0,955627769
LMS	116	144,90026	1,249140172
Kalman 1/1	94	28,62958	0,30457
Kalman 3/3	94	65,98195	0,701935638
Kalman 6/4	84	41,01938	0,488325952
Kalman 6/8	85	36,8097	0,433055294
Kalman 8/4	101	36,83928	0,364745347
Kalman 8/6	85	19,0097	0,223643529
Kalman 12/6	83	15,0097	0,180839759

Os resultados mostram que ambos os algoritmos que empregam modelo AR (com 12 coeficientes) para estimação do sinal residual (elemento de excitação do trato supraglotal teórico) conseguem discriminar um grande número de patologias, além de possibilitarem as maiores distâncias entre as disfonias (que pode ser verificado pelos valores obtidos para o nível e média da discriminação). Comparando com a medida EX, estes resultados são melhores que os obtidos com aquele método de avaliação da laringe.

Apesar do filtro de Kalman com uma configuração 8/4 de coeficientes apresentar um maior número de discriminações, utilizando-se três coeficientes para a porção autorregressiva e exógena consegue-se que as distribuições estejam separadas à uma distância maior que nas demais combinações de coeficientes para esta técnica adaptativa.

Tabela 121 - Análise estatística do pico da autocorrelação (PA) para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	109	89,99168	0,825611743
LMS	110	111,28034	1,011639455
Kalman 1/1	98	53,00131	0,540829694
Kalman 3/3	114	69,99162893	0,613961657
Kalman 6/4	109	56,45864	0,517969174
Kalman 6/8	113	104,08039	0,921065398
Kalman 8/4	118	63,80131	0,540689068
Kalman 8/6	104	78,96207	0,759250673
Kalman 12/6	107	75,080339	0,701685411

Com o emprego do fonema /e/, percebe-se que a quantidade de pares disfônicos identificáveis pelo parâmetro acústico eleva-se para os filtros de Kalman, juntamente com o nível e média das discriminações. A configuração 8/4 possibilita um maior número de discriminações, enquanto que o uso de 6 coeficientes para a porção autorregressiva e 8 para a porção exógena resulta em uma maior confiabilidade na identificação das patologias.

Mesmo com estas constatações, o emprego dos algoritmos por LMS e gradiente descendente são mais capazes de discriminar patologias utilizando esta medida acústica.

Outro fato a ser comentado é que a utilização do fonema /e/ para extração desta medida acústica permite que os resultados obtidos para o filtro de Kalman sejam melhores do que para a vogal /a/. O mesmo é verificado para o parâmetro acústico EX. Aparentemente a maior tensão sobre as pregas vocais, juntamente com as suas características formantes tornam possível a diferenciação entre as disfonias através do sinal residual obtido por este método recursivo.

Tabela 122 - Análise estatística do pico da autocorrelação (PA) para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	118	119,44704	1,012263051
LMS	113	96,56674	0,85457292
Kalman 1/1	63	79,0668	1,255028571
Kalman 3/3	97	46,56	0,48
Kalman 6/4	80	60,0918	0,7511475
Kalman 6/8	98	100,5704	1,026228571
Kalman 8/4	74	8,38229	0,113274189
Kalman 8/6	79	47,58812	0,602381266
Kalman 12/6	80	56,0578056	0,70072257

Comparando os resultados da tab. 122 com os observados nos demais fonemas, o número de discriminações reduz-se significativamente, indicando que as características de energia de alta frequência desta vogal influem negativamente na identificação das patologias através deste índice acústico.

Com uma diminuição não tão acentuada, as distâncias entre as distribuições de patologias também reforçam esta conclusão sobre a utilização do pico da autocorrelação empregada no sinal

residual do fonema /i/. Nota-se entretanto que a relação entre os algoritmos empregados mantém-se praticamente a mesma, ou seja, os filtros adaptativos que ignoram a porção exógena são capazes de identificar melhor as disfonias, enquanto que o emprego do filtro de Kalman sobre esta vogal é útil quando emprega-se a combinação 6/8 de coeficientes.

Pelos resultados obtidos, pode-se verificar que estas medidas acústicas (EX e PA) mostram-se extremamente eficientes para os fonemas /a/ (que garante a melhor confiabilidade na identificação das disfonias) e /e/ (que permite a distinção de um maior número de pares de tipos de voz).

4.3.3 Medida de Perturbação - *Jitter*

A utilização de medidas de perturbação para avaliação das estruturas de produção de fala possibilita que seja realizada distinção entre patologias que produzem voz soprosa daquelas que geram voz rouca. Alterações no padrão oscilatório das pregas vocais devem ser avaliadas para identificação correta da patologia.

O sinal residual possibilita que seja realizada a medição síncrona dos períodos sucessivos da frequência fundamental baseado nas diferenças máximas que ocorrem ao longo do resíduo da filtragem inversa.

Tabela 123 - Análise estatística do *jitter* para fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	112	76,801308	0,685725964
LMS	125	83,25073893	0,666005911
Kalman 1/1	121	109,08039	0,901490826
Kalman 3/3	124	102,45866	0,826279516
Kalman 6/4	124	97,45866	0,785956935
Kalman 6/8	122	91,79163	0,75239041
Kalman 8/4	115	83,45866	0,725727478
Kalman 8/6	120	96,45866	0,803822167
Kalman 12/6	117	92,45864	0,790244786

Utilizando as medidas de perturbação (*jitter*), constata-se que os resultados obtidos para os algoritmos de Kalman são mais confiáveis pois mantém mais afastadas as distribuições das patologias entre si. Apesar do número de discriminação favorecer a técnica recursiva de erro médio quadrático (LMS), as combinações 3/3, 6/4 e 6/8 fornecem resultados muito similares à este algoritmo.

Comparativamente às medidas acústicas extraídas dos resíduos da filtragem adaptativa inversa, pode-se concluir que a estimativa de perturbação da frequência fundamental possibilita a melhor performance na identificação de patologias.

Tabela 124 - Análise estatística do *jitter* para fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	125	46,79164	0,37433312
LMS	114	88,909921	0,779911588
Kalman 1/1	105	75,08036893	0,715051133
Kalman 3/3	118	105,07066	0,890429322
Kalman 6/4	118	106,0803389	0,898985923
Kalman 6/8	124	102,08037	0,82322879
Kalman 8/4	120	122,07066	1,0172555
Kalman 8/6	127	93,07065893	0,732839834
Kalman 12/6	133	44,06096	0,331285414

A tab. 124 indica que a utilização dos filtros de Kalman (particularmente nas configurações de 12/6 e 8/6 para o modelo paramétrico) conseguem distinguir um número maior de pares de tipos de voz. Outro indicativo desta característica dos filtros de Kalman é que a combinação de coeficientes 8/4 permite que as distribuições de patologias estejam bem separadas uma das outras, reduzindo assim a quantidade de erros durante a identificação das anormalidades do trato vocal.

A técnica por gradiente descendente (modelo AR) mostra-se pouco eficiente quando solicitada a separar os pares patológicos. O mesmo pode ser dito para o filtro de Kalman de combinação 12/6, que apesar de discriminar o maior número de patologias, apresenta uma baixa confiabilidade nesta atividade.

Tabela 125 - Análise estatística do *jitter* para fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	114	41,1696256	0,361137067
LMS	114	36,14151	0,317030789
Kalman 1/1	113	71,16968	0,629820177
Kalman 3/3	110	37,1696456	0,337905869
Kalman 6/4	125	48,1696656	0,385357325
Kalman 6/8	141	49,1696856	0,348721174
Kalman 8/4	120	43,14151	0,359512583
Kalman 8/6	137	43,1696856	0,315107194
Kalman 12/6	134	23,4456256	0,174967355

Com a elevação da tensão nas cordas vocais, pode-se verificar uma melhora na quantidade de pares de tipos de voz discrimináveis. Entretanto, constata-se simultaneamente que a distância entre as distribuições de patologias reduz-se drasticamente. Provavelmente a “estabilidade artificial” criada nas estruturas de vocalização possibilita uma medição mais precisa dos períodos sucessivos da frequência fundamental.

Os filtros de Kalman mostram grande eficiência quando aplicados sobre este fonema (além da vogal /e/) na identificação de um número maior de disfonias, pois a desc correlação provida pela porção exógena no sinal residual possibilita uma melhor definição dos picos sucessivos (cujo

distanciamento corresponde ao período da fundamental) e assim, caracterização mais precisa das diferenças entre as patologias.

4.3.4 Medida de Pico Residual

A utilização desta medida das máximas diferenças entre amostras do sinal residual, juntamente com a perturbação da frequência fundamental (*jitter*) visam avaliar o comportamento oscilatório das pregas vocais através do erro de predição linear (ou sinal de excitação do modelo AR/ARX).

Estas avaliações são pertinentes pois modificações ao longo da amplitude do sinal residual podem indicar disfonias específicas provocadas pelo aumento de aspereza (ou rouquidão) quando a voz é avaliada perceptualmente. Disfonias como hipercinesia ou paralisia unilateral podem produzir sinais residuais onde as máximas diferenças de amplitudes podem apresentar um padrão aleatório condizente com a extensão disfônica.

Os resultados da performance discriminatória desta medida acústica em relação aos algoritmos adaptativos são apresentados nos parágrafos a seguir, juntamente com os resultados numéricos da avaliação do poder discriminatório dos algoritmos adaptativos.

Tabela 126 - Análise estatística da medida pico para o fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	104	2,82958	0,0272075
LMS	102	4,00968	0,039310588
Kalman 1/1	101	14,84894	0,147019208
Kalman 3/3	103	7,00968	0,068055146
Kalman 6/4	94	17,83928	0,189779574
Kalman 6/8	85	0	0
Kalman 8/4	93	12,62958	0,135801935
Kalman 8/6	99	8,84893893	0,089383222
Kalman 12/6	90	12,84893893	0,142765988

Através da tab. 126, percebe-se que as distâncias entre as distribuições de probabilidades são mais baixa dentre os parâmetros acústicos obtidos dos resíduos, reduzindo assim a confiabilidade do parâmetro acústico pico. Verifica-se que a estimativa deste índice acústico apresenta maior confiabilidade quando calculado sobre o sinal residual obtido do filtro de Kalman, apesar da quantidade de pares identificáveis ser maior para os algoritmos LMS e gradiente descendente, ou seja, através de um modelo paramétrico AR.

Estes resultados indicam que a mediana das máximas diferenças do sinal residual são pouco eficientes na discriminação das patologias.

Tabela 127 - Análise estatística da medida pico para o fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	122	26,83926	0,219993934
LMS	121	31,83928	0,263134545
Kalman 1/1	124	17,12237093	0,138083637
Kalman 3/3	118	18,00968	0,152624407
Kalman 6/4	135	16,82958	0,124663556
Kalman 6/8	146	49,78155	0,340969521
Kalman 8/4	134	31,95194893	0,23844738
Kalman 8/6	128	25,94227	0,202673984
Kalman 12/6	111	32,95195	0,296864414

Com a utilização do fonema /e/ percebe-se uma melhora significativa nos resultados da discriminação das disfonias, permitindo que as distribuições das patologias tenham suas distâncias aumentadas, juntamente com a elevação da quantidade de pares de tipos de voz.

Os filtros de Kalman demonstram possuir a melhor performance na identificação das patologias, particularmente nas combinações de coeficientes 6/8 e 8/4. Mesmo com estas evidências, deve-se considerar que os valores conjuntos de número, nível e média da discriminação, que permitem avaliação do teste de hipótese Mann-Whitney, indicam a baixa capacidade de identificação das disfonias.

Tabela 128 - Análise estatística da medida pico para o fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	107	8,02818	0,07502972
LMS	100	3	0,03
Kalman 1/1	93	4,506786	0,048460065
Kalman 3/3	93	18,50679	0,198997742
Kalman 6/4	121	40,5068	0,334766942
Kalman 6/8	103	17,01358	0,165180388
Kalman 8/4	106	34,5068	0,325535849
Kalman 8/6	114	43,01359	0,377312193
Kalman 12/6	103	27,5068	0,267056311

Na tab. 128, verifica-se que os filtros de Kalman fornecem melhor performance na avaliação patológica, principalmente quando se considera 6 e 8 coeficientes, juntamente com 8 e 6 coeficientes, para os respectivos filtros de Kalman.

A utilização do fonema /i/ para o diagnóstico de patologias mostra-se também pouco eficiente. Entretanto, constata-se pelos resultados obtidos que o emprego de sinais de voz com concentração energética a partir de 2000 Hz (segunda e terceira formantes localizadas nesta banda espectral) permitem que a diferenciação entre as disfonias seja mais confiável (comparando o nível de discriminação entre /e/ e /i/ em relação à /a/).

Entretanto, a proximidade entre as distribuições das patologias impede uma estimativa confiável, comparando-se aos resultados obtidos para outras medidas acústicas.

4.3.5 Medida de nivelamento espectral do filtro inverso - SFF

Esta medida permite avaliar a quantidade de ruído espectral provocado pela passagem de ar turbulento através das pregas vocais, e assim identificar patologias que provoquem vozes soprosas ou roucas. Juntamente com as medidas de nivelamento espectral do sinal residual e a relação sinal-ruído, espera-se identificar adequadamente as patologias que afetam a laringe e pregas vocais, baseado em modificações na estrutura formante do fonema analisado.

Tabela 129 - Análise estatística da medida SFF para o fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	119	44,75197	0,376066975
LMS	126	45,63928	0,362216508
Kalman 1/1	109	38,951972	0,357357541
Kalman 3/3	109	26,83928	0,246231927
Kalman 6/4	100	20,8097	0,208097
Kalman 6/8	91	28,63928	0,314717363
Kalman 8/4	108	47,551972	0,440296037
Kalman 8/6	103	37,75197	0,366523981
Kalman 12/6	100	46,75197	0,4675197

Empregando os algoritmos para estimação dos filtros de Wiener, particularmente o LMS, obtém-se melhor performance na identificação patológica com o uso da medida de avaliação do mascaramento da estrutura formante do sinal de voz pelo ruído.

Pode-se verificar também que o emprego de um número superior de coeficientes da porção autorregressiva dos filtros de Kalman implica na melhora da capacidade de distinção dos pares disfônicos. Além disto, reduzindo-se o número de coeficientes da porção exógena (quando aplica-se os cálculos sobre o fonema /a/) produz-se uma maior separação nas distribuições de patologias.

Tabela 130 - Análise estatística da medida SFF para o fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	118	45,95197	0,389423475
LMS	117	47,95197	0,409845897
Kalman 1/1	130	56,95197	0,438092077
Kalman 3/3	127	37,122392	0,292302299
Kalman 6/4	127	40,122392	0,315924346
Kalman 6/8	129	51,95197	0,40272845
Kalman 8/4	112	29,12239	0,260021339
Kalman 8/6	120	31,95197	0,266266417
Kalman 12/6	112	33,12239	0,295735625

Na tab. 130, os resultados mostram que com o emprego de medidas acústicas baseadas em sinais residuais há um maior poder de discriminação quando se faz uso do fonema /e/. Este fato é observável em quase todas as índices acústicos desenvolvidos sobre os erros de estimação da filtragem adaptativa.

Para os filtros de Kalman, verifica-se que a utilização de um número de coeficientes abaixo de 6 para a porção exógena permite que um número razoável de pares de patologias sejam identificáveis. O fato da combinação 1/1 para o modelo ARX resultar em um número maior de patologias discrimináveis (com um nível de separabilidade das distribuições de disfonias igualmente superior) deve-se a baixa quantidade de “formantes” identificáveis pelo modelo considerado. Assim, patologias como carcinomas e granulomas resultaram em uma curva de densidade espectral de potência muito mais nivelada (*flatness*) do que vozes normais. O procedimento matemático que identifica esta característica do filtro adaptativo torna-se mais eficiente.

A influência da porção exógena também aumenta a performance da identificação das patologias. Verifica-se pelos pares 6/4 e 6/8, e 8/4 e 8/6, que com o aumento dos coeficientes desta porção nos filtros de Kalman, um número maior de pares de tipos de voz é discriminável, além do aumento da confiabilidade da medida acústica.

Cabe ressaltar também que os filtros de Wiener apresentam um pequeno decréscimo na capacidade de identificar as patologias em relação ao fonema /a/, apesar de não ser significativamente diferente se for considerado a manutenção da distância entre as distribuições das disfonias.

Tabela 131 - Análise estatística da medida SFF para o fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	106	39,5349756	0,372971468
LMS	115	41,5349756	0,361173701
Kalman 1/1	106	31,5068	0,297233962
Kalman 3/3	118	40,5068	0,343277966
Kalman 6/4	123	51,5068	0,418754472
Kalman 6/8	112	46,5068	0,415239286
Kalman 8/4	109	38,5068	0,353273394
Kalman 8/6	108	41,5068	0,384322222
Kalman 12/6	105	31,5068	0,300064762

Os resultados obtidos pela extração do parâmetro acústico SFF do fonema /i/ comprovam mais uma vez a utilidade do filtro de Kalman na avaliação do mascaramento das componentes formantes do sinal de voz por ruído proveniente de turbulência na cavidade glotal.

Particularmente, verifica-se que as combinações de coeficientes 6/4 e 6/8 do modelo ARX mostram-se mais capacitadas na identificação das patologias. Entretanto, a performance da medida acústica no número de disfonias discrimináveis decai com a elevação da tensão sobre as pregas vocais.

Os filtros de Wiener são influenciáveis com a alteração da composição formante da vogal analisada. Com a mudança de localização destas componentes espectrais, a performance dos mesmos decai.

Como uma das medidas que diretamente avaliam a influência do ruído turbulento na voz do paciente, percebe-se que a distância entre as distribuições das disfonias não é tão elevada quando parâmetros acústicos como a medida de perturbação (*jitter*), ou o coeficiente de excesso. Isto indica que disfonias como cistos, nódulos e pólipos (com característica soprosa) são de difícil identificação. Esta característica perceptual demonstra a grande similaridade existente entre as patologias que elevam a turbulência do fluxo de ar na glote.

4.3.6 Medida de nivelamento espectral do sinal residual - SFR

Para avaliar o mascaramento das componentes harmônicas, extraiu-se a medida acústica SFR do sinal residual com finalidade de identificação das patologias com características soprosas e roucas. O emprego de um número maior de índices acústicos para avaliar a quantidade de ruído provocado pelo aumento da turbulência nas pregas vocais deve-se a semelhança acústica de disfonias como fenda fusiforme, nódulos e cistos, por exemplo.

Seguem os resultados de performance discriminatória da medida SFR (*spectral flatness of residue*), para os fonemas /a/, /e/ e /i/.

Tabela 132 - Análise estatística da medida SFR para o fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	98	41,0097	0,418466327
LMS	126	54,8887	0,435624603
Kalman 1/1	88	3	0,034090909
Kalman 3/3	83	11,01936	0,132763373
Kalman 6/4	62	0	0
Kalman 6/8	72	1	0,013888889
Kalman 8/4	68	0,112692	0,001657235
Kalman 8/6	68	1	0,014705882
Kalman 12/6	92	22,01359	0,239278152

Constata-se pelo emprego do fonema /a/ para estimação da medida SFR que tanto o número discriminações quanto o nível/média de separabilidade das disfonias é muito baixo para os filtros de Kalman. O sinal residual determinado por estes algoritmos apresenta características espectrais muito semelhantes

Os parâmetros acústicos baseados nos filtros de Wiener apresentam resultados compatíveis com os obtidos para a medida SFF (anteriormente apresentada). Para o índice SFR, fornecem melhor performance na discriminação das patologias.

Tabela 133 - Análise estatística da medida SFR para o fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	112	27,460981	0,24518733
LMS	96	38,26096	0,398551667
Kalman 1/1	97	45,6097	0,470203093
Kalman 3/3	116	39,1524	0,33752069
Kalman 6/4	111	4,01935893	0,036210441
Kalman 6/8	94	19,0097	0,202230851
Kalman 8/4	84	5,8	0,069047619
Kalman 8/6	78	0	0
Kalman 12/6	96	2	0,020833333

Novamente verifica-se que com o aumento da tensão das pregas vocais (para produção do fonema /e/) o número de tipos de voz identificáveis é maior, para os filtros de Kalman, onde a concentração de energia localiza-se abaixo de 2000 Hz.

O emprego de um número maior de coeficientes para o modelo ARX desta técnica adaptativa resulta na baixa performance discriminatória, indicando que as modificações espectrais das componentes harmônicas do sinal de voz são muito similares.

Outro resultado importante é que os algoritmos de gradiente descendente e LMS tem a capacidade de separabilidade das disfonias reduzida, pela diminuição das distâncias entre as distribuições dos tipos de voz.

Tabela 134 - Análise estatística da medida SFR para o fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	131	43,634676	0,33308913
LMS	126	79,1415	0,628107143
Kalman 1/1	108	29,81456	0,276060741
Kalman 3/3	130	22,53172	0,173320923
Kalman 6/4	98	21,23081	0,216640918
Kalman 6/8	125	38,634662	0,309077296
Kalman 8/4	91	3	0,032967033
Kalman 8/6	97	11	0,113402062
Kalman 12/6	93	2	0,021505376

Nota-se na tab. 134 a melhora nos resultados provocada pela extração da medida SFR a partir do fonema /i/, principalmente nos filtros Kalman. A combinação 6/8 de coeficientes para o modelo ARX, a mais eficiente para este tipo de filtro, pois permite que as distribuições de patologias não estejam tão próximas uma das outras.

Entretanto, o emprego de filtros de Wiener fornece uma maior confiabilidade na discriminação patológica, pelos valores obtidos para o número, nível e média de discriminação.

A utilização deste parâmetro acústico (SFR) para análise da condição, disfônica ou não, das estruturas de vocalização apresenta baixa performance, principalmente no emprego do fonema /a/. A

distribuição de energia ao longo da curva de densidade espectral de potência influencia diretamente nos resultados discriminatórios.

4.3.7 Medida de Relação Sinal-Ruído

Esta última medida acústica realizada sobre os sinais residuais é baseada no fato de que este sinal corresponde ao erro de estimação de um modelo paramétrico para o fonema pronunciado. Assim, sua quantificação indica o nível de ruído existente no sinal de voz.

O objetivo é compor juntamente com as medidas SFR e SFF, um conjunto de parâmetros que avaliem diretamente o escape de ar pela cavidade glotal.

Tabela 135 - Análise estatística da medida SNR para o fonema /a/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	109	14,83926	0,13614
LMS	110	25,83928	0,234902545
Kalman 1/1	101	38,83928	0,384547327
Kalman 3/3	99	17,83926	0,180194545
Kalman 6/4	79	11,83926	0,149864051
Kalman 6/8	78	12,83926	0,164605897
Kalman 8/4	87	26,63926	0,306198391
Kalman 8/6	91	15,00968	0,164941538
Kalman 12/6	87	21,63926	0,248727126

Pelos resultados apresentados na tab. 135, verifica-se que os algoritmos que implementam os filtros de Wiener conseguem discriminar um número maior patologias em relação aos filtros de Kalman. Mas algumas configurações de coeficientes (1/1, 8/4 e 12/6) garante confiabilidade na identificação das disfonias, mesmo com um número menor de pares de tipos de voz separáveis.

Tabela 136 - Análise estatística da medida SNR para o fonema /e/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	122	45,781552	0,375258623
LMS	120	48,95197	0,407933083
Kalman 1/1	124	62,95198	0,507677258
Kalman 3/3	132	32,12237	0,243351288
Kalman 6/4	119	30,12237	0,25312916
Kalman 6/8	135	59,95198	0,444088741
Kalman 8/4	110	28,12239	0,255658091
Kalman 8/6	109	21,95195	0,201394037
Kalman 12/6	109	19,12237	0,175434587

A utilização da vogal /e/ indica que os filtros de Kalman são mais eficientes para discriminação das patologias, particularmente nas combinações 1/1 e 6/8. A utilização de um número baixo de coeficientes resultando em uma separabilidade das distribuições é proveniente da maior magnitude das amostras residuais em relação às outras configurações do filtro de Kalman. Pelos

resultados obtidos, conclui-se que a magnitude do sinal residual para patologias graves, como carcinomas, seja muito maior que as vozes normais (para este tipo de modelo ARX), facilitando assim a discriminação das demais disfonias.

Verifica-se que os filtros de Kalman com número de coeficientes abaixo de 6 para a porção autorregressiva resulte em maior performance discriminatória.

Tabela 137 - Análise estatística da medida SNR para o fonema /i/

	Número de Discrim.	Nível de Discrim.	Média das Discrim.
Gradiente	102	31,53498	0,309166471
LMS	106	42,53498	0,401273396
Kalman 1/1	116	29,5068	0,254368966
Kalman 3/3	108	38,5068	0,356544444
Kalman 6/4	109	46,5068	0,42666789
Kalman 6/8	112	38,5068	0,343810714
Kalman 8/4	105	30,5068	0,290540952
Kalman 8/6	105	38,5068	0,366731429
Kalman 12/6	98	24,5068	0,250069388

Pela tab. 137, verifica-se novamente a superioridade (não muito acentuada) das medidas baseadas nos filtros de Kalman em relação aos algoritmos gradiente descendente e LMS. Empregando 6 coeficientes para a porção autorregressiva do modelo ARX, obtém-se melhor eficácia na identificação dos tipos de voz analisados.

A utilização deste parâmetro acústico apresenta resultados semelhantes aos obtidos pelo índice SFF, indicando que estas duas metodologias de avaliação do escape de ar glotal (com elevada turbulência) devem ser consideradas em modelos de discriminação multivariável.

As medidas acústicas baseadas em sinal residual, principalmente EX, PA e Jitter produziram os melhores resultados, devido a maior robustez (pelos valores de número, nível e média da discriminação) em relação às medidas como SFF e SNR. Dentre os parâmetros acústicos considerados, os de menor performance foram pico e SFR. Modificações no método de avaliação das perturbações em amplitude do sinal residual (pico) podem resultar em aumento da sua capacidade discriminatória.

Verificou-se pelos resultados que as medidas residuais oferecem uma maior performance discriminatória quando extraídas do fonema /e/, possivelmente devido à característica de concentração de energia a partir de 2000 Hz (segunda e terceira formantes) e menor tensão nas pregas vocais, em relação à vogal /i/.

Considerando os algoritmos adaptativos, verificou-se que as medidas EX e PA são mais eficientes na identificação das patologias quando são extraídas a partir do sinal residual obtido pelos filtros Wiener (principalmente LMS), enquanto que os parâmetros acústicos para avaliação direta da turbulência aérea na glote (SFF, SFR e SNR) tem maior performance pelo emprego de filtros de Kalman. Cabe salientar que não houve a preponderância de uma combinação específica na

identificação das patologias, indicando a necessidade de utilização de várias configurações do modelo ARX para avaliação de anomalias nas pregas vocais.

Entretanto, se for considerado a aplicação de um único algoritmo de filtragem para extração das medidas acústicas, percebe-se que o mais indicado é o filtro de Wiener, estimado por LMS.

A análise da performance computacional demonstrou que o emprego de filtros de Kalman demanda muitas operações matriciais, tornando-o extremamente lento. Já o algoritmo por gradientes descendente, empregado na estimação do filtro de Wiener, mostrou ser 5 vezes mais rápido que o filtro Kalman na configuração 6/6 (tamanho matricial equivalente).

Uma última consideração sobre o filtro de Kalman deve ser feita: verificou-se que o emprego de um número de coeficientes alto para a estimação do modelo ARX resulta em baixa performance na discriminação patológica, fato evidenciado pelos resultados obtidos pela combinação de coeficientes 12/6. Possivelmente a porção exógena, durante a decorrelação do sinal residual, retire informação sobre as disfonias, padronizando os sinais de voz. Assim, pode-se esperar que uma análise dos coeficientes obtidos resulte na estimação direta das disfonias, ao invés de analisar-se diretamente o sinal residual.

4.4 Resultados da Identificação Patológica empregando

Redes Neurais Artificiais

Uma vez determinado o poder de discriminação, individual, dos parâmetros acústicos, pode-se então construir configurações de redes neurais artificiais que possam ser capazes de identificar mais eficientemente os padrões patológicos que serão apresentados.

Cada padrão apresentado é representado pelo conjunto de valores proveniente dos índices acústicos e do vetor de saída, para uma determinada voz (patológica ou não). A descrição do vetor de saída é apresentada na tab. 5, na seção 3.6.1.

Os parâmetros de controle de aprendizado da rede neural *feedforward* com treinamento por *backpropagation* foram controlados manualmente, visando buscar a convergência assintótica para o mínimo global. Verificou-se que o emprego do termo de momento (α) acarretava em dois acontecimentos distintos:

1. Quando o valor é superior à 0,3 (aproximadamente), os pesos de algumas conexões tornam-se instáveis, crescendo exponencialmente, incapazes de serem processados pelo computador.
2. Quando o valor é inferior à 0,3 (aproximadamente), a convergência torna-se extremamente lenta, possivelmente pela existência de uma grande quantidade de regiões planas no espaço de pesos ω_i .

Com estas duas constatações, ignorou-se este parâmetro (ou seja, $\alpha=0$) em todas as configurações analisadas.

A quantidade de camadas na rede neural também foi analisada pois define geometricamente a maneira com que as “patologias” serão definidas espacialmente. Primeiramente avaliou-se a capacidade de discriminação da rede neural utilizando apenas uma camada intermediária. Esta arquitetura mostrou-se insuficiente para classificar a maioria das configurações de padrões de entrada, fornecendo uma baixa taxa de discriminação (apenas 20% de acerto para padrões não-treinados para uma construção empregando 9 células na camada de entrada, 50 células na camada intermediária e 22 células na camada de saída; a entrada contém os índices acústicos obtidos a partir do sinal residual obtido por LMS, juntamente com o tipo do fonema pronunciado e o sexo da pessoa). Entretanto, quando empregou-se um grande número de informações para as duas principais características encontradas nas patologias estudadas (periodicidade do sinal de voz e quantidade de ruído turbulento), foi possível empregar somente uma camada intermediária.

Outro fator importante que deve ser considerado simultaneamente é a quantidade de neurônios empregados. Verificou-se que elevando consideravelmente o número de células nas camadas intermediárias, a convergência igualmente aumenta. Entretanto, os erros na etapa de teste destas construções neurais é incrementado, uma vez que os agrupamentos torna tantos que são suficientes para envolver cada padrão treinado, reduzindo assim a capacidade de generalização e interpolação da rede neural para vozes não treinadas. Procurou-se manter uma pequena quantidade de células nas camadas intermediárias.

Para avaliar a capacidade das redes neurais em identificar as patologias, utilizou-se uma função degrau para cada neurônio da camada de saída (que informa probabilisticamente discriminação de uma disfonia a partir das medidas acústicas apresentadas), ou seja,

$$f(x) = \begin{cases} 1, & x \geq 0,5 \\ 0, & x < 0,5 \end{cases} \quad (118)$$

Assim, quando a probabilidade de uma patologia é igual ou superior à 0,5, constata-se que a rede neural identificou características acústicas no sinal de voz que “confirmam” a existência desta disfonia.

Para realizar o teste das redes neurais, empregou-se um conjunto de medidas acústicas de vozes que não foram empregadas no treinamento destas estruturas. A tab. 138 apresenta os padrões (normais e patológicos), bem como a quantidade vozes por patologia.

Tabela 138 - Relação de patologias e quantidade de pessoas para teste das redes neurais

Patologia	N. de casos
Cisto	2
Edema de Reinke	1
Fenda Fusiforme	10
Nódulo	2
Normal	23
Pólipo	2
Total	40

Nos testes preliminares empregando um classificador neural, foram analisadas as capacidades de identificação das disfonias empregando apenas grupos correlatos de parâmetros acústicos, ou seja, grupos em que os índices acústicos apresentam o mesmo objetivo na identificação de características acústicas (temporais ou espectrais). Na última seção são apresentados estruturas de redes neurais em que utiliza-se uma gama maior de parâmetros acústicos que possibilitam uma melhor identificação das disfonias. A descrição destes grupos e os resultados produzidos por estes, são apresentados nas seções a seguir.

4.4.1 Configurações empregando perturbação na amplitude e período da fundamental.

No primeiro grupo de teste são consideradas apenas as alterações da frequência fundamental, cujas vozes podem ser avaliadas perceptualmente como roucas ou ásperas. Patologias como carcinomas, hipercinesia e laringite crônica devem ser bem identificadas pelas redes neurais construídas a partir destas medidas acústicas.

Considerando que as medidas de perturbação (da ordem zero até a ordem quatro) tanto para período quanto amplitude da fundamental, apresentaram os melhores resultados na discriminação patológica, estes foram considerados em algumas configurações de entrada. Também foram utilizados os parâmetros PVI, média móvel com 5 e 11 coeficientes e DPF. A seguir, são apresentadas as configurações de entrada testadas, considerando os parâmetros acústicos envolvendo medidas de perturbação do período e amplitude da frequência fundamental.

Pode-se verificar nas configurações apresentadas que a rede neural é treinada para reconhecer simultaneamente o fonema apresentado e o sexo da pessoa cuja voz é analisada. A apresentação do fonema tem o intuito de fazer com que esta estrutura neural saiba reconhecer as diferenças de tensão nas pregas vocais importantes pois algumas vozes para /e/ e /i/ apresentam muita similaridade quando pronunciadas por pacientes disfônicos. Além disto, muitas medidas acústicas são melhor empregadas quando se considera um determinado fonema. A rede neural pode então operar como um especialista analisando características específicas das patologias para cada um dos fonemas apresentados.

O parâmetro sexo apresentado à rede neural é motivado pelo fato de existir diferenças orgânicas e funcionais na laringe e pregas vocais entre homens e mulheres. Patologias como fenda fusiforme são identificadas de modo diferenciado conforme o sexo do paciente pois anatomicamente as pregas vocais das mulheres apresentam uma abertura maior, na região posterior da glote, em relação ao homem, além da maior tensão nas pregas vocais (maior frequência fundamental). Assim, a rede neural deve compreender estas diferenças e capturar as características patológicas para ambos os sexos.

A seguir, são apresentadas as configurações de entrada utilizadas com o objetivo de discriminar as disfonias estudadas.

Configuração 01, contendo 22 entradas:

- Sexo (0 para mulheres e 1 para homens);
- Fonema (0 para /a/, 0,5 para /e/ e 1 para /i/);
- PVI e médias móveis com 5 e 11 coeficientes, para período da fundamental
- Desvio Absoluto e Padrão das perturbações de ordem zero, 1, 2, 3 e 4, do período da fundamental;
- DPF e média móvel com 11 coeficientes para amplitude da fundamental;
- Curtose das perturbações de ordem zero, 1, 2, 3 e 4, da amplitude da fundamental.

Configuração 02, contendo 22 entradas:

- Sexo e Fonema;
- PVI e médias móveis com 5 e 11 coeficientes, para período da fundamental
- Desvio clássico e robusto da regressão linear, das perturbações de ordem zero, 1, 2, 3 e 4, do período da fundamental;
- DPF e média móvel com 11 coeficientes para amplitude da fundamental;
- Curtose das perturbações de ordem zero, 1, 2, 3 e 4, da amplitude da fundamental.

Configuração 03, contendo 15 entradas:

- Sexo e Fonema;
- PVI e médias móveis com 5 e 11 coeficientes, para período da fundamental;
- Desvio clássico e robusto da regressão linear, das perturbações de ordem zero, 1, 2, 3 e 4, do período da fundamental;

Configuração 04, contendo 10 entradas:

- Sexo e Fonema;
- Série temporal, DPF e média móvel com 11 coeficientes, para amplitude da fundamental;
- Curtose das perturbações de ordem zero, 1, 2, 3 e 4, da amplitude da fundamental.

Configuração 05, contendo 10 entradas:

- Sexo e Fonema;
- JR, JF, PVI e médias móveis com 5 e 11 coeficientes, para período da fundamental;
- DPF, série temporal e média móvel com 11 coeficientes para amplitude da fundamental.

Configuração 06, contendo 17 entradas:

- Sexo e Fonema;
- Desvio Absoluto e Padrão das perturbações de ordem zero, 1, 2, 3 e 4, do período da fundamental;
- Curtose das perturbações de ordem zero, 1, 2, 3 e 4, da amplitude da fundamental.

A escolha destas configurações visa analisar a capacidade discriminatória da rede, além de validar as considerações estatísticas que indicaram as melhores medidas acústicas para identificação das patologias. Outra característica que procurou-se avaliar foi os efeitos do emprego de estatística robusta em relação à clássica. Pode-se verificar pelos resultados apresentados na seção 4.1.1 que muitas vezes encontrava-se resultados equivalentes para medidas como desvio clássico e robusto da regressão linear, por exemplo. As estimativas robustas são caracterizadas por uma dispersão de valores menores que as obtidas por técnicas clássicas (geralmente baseadas em curvas de probabilidade gaussiana para as distribuições de valores). Assim, a proximidade destas pode influenciar a capacidade de generalização da rede neural.

Nas configurações 01 e 02, utilizou-se uma grande quantidade de medidas acústicas, mas precisamente às de melhor capacidade discriminatória quando se considera análise das alterações na periodicidade de movimento das pregas vocais.

Já as configurações 03 e 04 correspondem, respectivamente, à medidas de perturbação no período e amplitude da fundamental. O intuito é verificar qual o comportamento da rede neural quando apenas apresenta-se uma destas avaliações de periodicidade do sinal de voz.

Para examinar o efeito separado de medidas acústicas convencionais existentes na literatura em relação às estimações estatísticas das perturbações de ordem superior, para período e amplitude de F_0 , foram utilizadas tanto as configurações 05 e 06, respectivamente. Cabe salientar que a diferença na quantidade de entradas entre ambas pode propiciar uma redução no poder classificatório da rede neural artificial.

Empregando os parâmetros extraídos do conjunto de vozes para teste das estruturas neurais (tab. 139), obtém-se a seguinte tabela, contendo o percentual de acerto desta estrutura para todos os padrões treinados (considerando que cada pessoa pronunciou os três fonemas analisados e cada uma destas vozes é aplicada na rede neural) e separadamente para /a/, /e/ e /i/.

Tabela 139 - Classificação patológica por redes neurais para medidas de perturbação do período e amplitude da fundamental

Config.	Número de células por camada	Erro médio por padrão	% de acerto total	% de acerto para /a/	% de acerto para /e/	% de acerto para /i/
01	22/22/22/22	0,08342	60,00%	55,00%	60,00%	65,00%
01	22/25/25/22	0,04109	61,67%	62,50%	60,00%	62,50%
02	22/22/22/22	0,06190	58,33%	45,00%	52,50%	77,50%
02	22/25/25/22	0,04135	55,83%	47,50%	47,50%	72,50%
03	15/22/22/22	0,25551	74,17%	55,00%	77,50%	90,00%
03	15/25/25/22	0,26867	34,17%	42,50%	25,00%	35,00%
04	10/22/22/22	0,32048	39,17%	47,50%	50,00%	25,00%
04	10/25/25/22	0,12615	43,33%	37,50%	40,00%	52,50%
05	10/22/22/22	0,14476	52,50%	50,00%	60,00%	47,50%
05	10/25/25/22	0,06923	60,83%	50,00%	65,00%	67,50%
06	17/22/22/22	0,23585	69,17%	62,50%	65,00%	80,00%
06	17/25/25/22	0,25544	75,83%	65,00%	75,00%	87,50%
06	17/25/25/22	0,23883	60,83%	55,00%	57,50%	70,00%

A coluna “Número de células por camada” representa a quantidade de neurônios empregados na camada de entrada (onde não ocorre processamento e sim apenas armazenamento temporário dos padrões de treinamento/teste), as quantidades de neurônios para as duas camadas intermediárias, e a quantidade de neurônios da camada de saída (onde obtém-se o resultado propriamente dito da rede neural). A coluna “Erro médio por padrão” indica o erro médio apresentado por cada padrão em um ciclo completo (apresentação de todos os padrões). O critério de parada da rede utilizado é que este erro seja inferior à 0,01, ou que o número de ciclos totais seja igual à 10000.

O objetivo de efetuar-se o treinamento em no máximo 10000 iterações é verificar a potencialidade das medidas acústicas, além de se analisar a suficiência da quantidade de parâmetros utilizados na camada de entrada. Configurações com maior velocidade de convergência podem indicar que as medidas acústicas empregadas conseguem identificar adequadamente o conjunto de patologias estudadas.

Existe outro motivo para utilizar este critério para encerramento do processo de aprendizado: quanto mais se reduz o erro de aprendizado, considerando a quantidade de neurônios nas camadas intermediárias, os intervalos de valores dos parâmetros acústicos para as disfonias tornam-se cada vez menores, reduzindo-se assim a capacidade de generalização da rede neural.

Os resultados apresentados na tab. 139 demonstram que a capacidade geral das redes neurais treinadas não é elevada. Entretanto, este resultado é suficiente se for considerado que o especialista humano quando analisa a voz no diagnóstico das patologias também analisa a abertura das pregas vocais e desta forma, a turbulência do fluxo de ar na glote. As medidas de perturbação não são suficientes para prover este tipo de informação. Demonstra-se pelos resultados obtidos que a rede neural é capaz de identificar muitas patologias com uma baixa quantidade de padrões treinados,

indicando que os critérios para treinamento da rede neural são corretos para a tarefa de discriminação patológica a partir de medidas acústicas.

4.4.2 Configurações empregando estimativas de ruído turbulento

Os parâmetros acústicos para avaliação da quantidade de ruído proveniente do fechamento incompleto das pregas vocais são importantes indicadores de disfonias. Nesta seção são consideradas configurações de redes neurais contendo apenas estimativas de ruído na camada de entrada, com o intuito de verificar a sua potencialidade em identificar patologias como fenda fusiforme, cistos, nódulos e pólipos, em que as pregas vocais apresentam movimento periódico normal mas devido à crescimentos orgânicos ou funcionamento inadequado, permitem que uma quantidade (variável conforme a extensão e o tipo da disfonia) de ar turbulento excite a cavidade de ressonância.

Como a quantidade de índices acústicos para estimação do ruído turbulento é elevada, foram considerados apenas aqueles que apresentaram melhor capacidade estatística (avaliada pelo teste Mann-Withney) na discriminação dos pares de tipos de voz.

A seguir, apresenta-se as configurações de entrada utilizando parâmetros de avaliação de ruído para discriminação das disfonias.

Configuração 07, contendo 14 entradas:

- Sexo e Fonema;
- Média e desvio clássico da regressão linear do primeiro e segundo harmônicos (H1 e H2);
- Média e desvio clássico da regressão linear da energia espectral nas bandas de 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz.

Configuração 08, contendo 14 entradas:

- Sexo e Fonema;
- Mediana e desvio robusto da regressão linear do primeiro e segundo harmônicos (H1 e H2);
- Mediana e desvio robusto da regressão linear da energia espectral nas bandas de 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz.

Configuração 09, contendo 14 entradas:

- Sexo e Fonema;
- Média e desvio clássico da regressão linear da estimativa CPP;
- Média e desvio clássico da regressão linear da estimativa HNR por técnica cepstral nas bandas de 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz e 50 à 2000 Hz.

Configuração 10, contendo 14 entradas:

- Sexo e Fonema;
- Mediana e desvio robusto da regressão linear da estimativa CPP;
- Mediana e desvio robusto da regressão linear da estimativa HNR por técnica cepstral nas bandas de 60 à 400 Hz, 400 à 2000 Hz, 2000 à 5000 Hz, 5000 à 8000 Hz e 50 à 2000 Hz.

Configuração 11, contendo 10 entradas:

- Sexo e Fonema;
- Desvio clássico da regressão linear da estimativa P/A
- Média e desvio clássico da regressão linear da estimativa SNL nas bandas 100 à 2600 Hz e 100 à 5100 Hz;
- Estimativa HFPR acima de 6000 Hz;
- Estimativa NNE nas bandas 1000 à 5000 Hz e 2000 à 5000 Hz.

Configuração 12, contendo 10 entradas:

- Sexo e Fonema;
- Desvio robusto da regressão linear da estimativa P/A
- Mediana e desvio robusto da regressão linear da estimativa SNL nas bandas 100 à 2600 Hz e 100 à 5100 Hz;
- Estimativa HFPR acima de 6000 Hz;
- Estimativa NNE nas bandas 1000 à 5000 Hz e 2000 à 5000 Hz.

Configuração 13, contendo 12 entradas:

- Sexo e Fonema;
- Desvio clássico da regressão linear do primeiro e segundo harmônicos (H1 e H2);
- Desvio clássico da regressão linear da energia espectral nas bandas de 60 à 400 Hz e 400 à 2000 Hz;
- Média da estimativa CPP;
- Desvio clássico da regressão linear da estimativa HNR por técnica cepstral nas bandas de 400 à 2000 Hz e 2000 à 5000 Hz;
- Estimativa NNE nas bandas 1000 à 5000 Hz e 2000 à 5000 Hz;
- Desvio clássico da regressão linear da estimativa P/A;

Configuração 14, contendo 12 entradas:

- Sexo e Fonema;
- Desvio robusto da regressão linear do primeiro e segundo harmônicos (H1 e H2);
- Desvio robusto da regressão linear da energia espectral nas bandas de 60 à 400 Hz e 400 à 2000 Hz;
- Mediana da estimativa CPP;
- Desvio robusto da regressão linear da estimativa HNR por técnica cepstral nas bandas de 400 à 2000 Hz e 2000 à 5000 Hz;
- Estimativa NNE nas bandas 1000 à 5000 Hz e 2000 à 5000 Hz;
- Desvio robusto da regressão linear da estimativa P/A;

As configurações 07 e 08 são consideradas apenas medidas que avaliam a curva de densidade espectral de potência dos sinais de voz (normal e patológico). O objetivo é verificar se a rede neural é capaz de identificar patologias observando modificações nas componentes harmônicas (primeira e segunda) e avaliar a influência do nivelamento espectral que patologias como carcinoma e paralisia unilateral provocam. A avaliação espectral também é realizada pelas redes neurais que empregam as configurações 09 e 10, utilizando medidas acústicas obtidas por técnicas cepstrais.

Considerando outras medidas extraídas no domínio da frequência, as configurações 11 e 12 foram desenvolvidas para avaliar simultaneamente características de aspereza e sopro que existem em todas as patologias da laringe e pregas vocais.

As configurações 13 e 14 empregam os melhores parâmetros acústicos (do ponto de vista das avaliações estatísticas realizadas sobre as medidas de avaliação de ruído), buscando-se um conjunto ótimo de índices acústicos que propiciem à rede neural identificar características que diferenciem as patologias analisadas.

Tabela 140 - Classificação patológica por redes neurais para medidas de ruído

Config.	Número de células por camada	Erro médio por padrão	% de acerto total	% de acerto para /a/	% de acerto para /e/	% de acerto para /i/
07	14/22/22/22	0,01893	51,67%	37,50%	50,00%	67,50%
07	14/25/25/22	0,09994	35,00%	27,50%	25,00%	52,50%
08	14/22/22/22	0,05103	44,17%	47,50%	42,50%	42,50%
08	14/25/25/22	0,00998	37,50%	35,00%	30,00%	47,50%
09	14/22/22/22	0,00998	43,33%	35,00%	52,50%	42,50%
09	14/25/25/22	0,00999	50,83%	47,50%	37,50%	67,50%
10	14/22/22/22	0,00998	39,17%	17,50%	60,00%	40,00%
10	14/25/25/22	0,01003	54,17%	55,00%	35,00%	72,50%
11	10/22/22/22	0,03963	40,83%	40,00%	35,00%	47,50%
11	10/25/25/22	0,03817	49,17%	57,50%	35,00%	55,00%
12	10/22/22/22	0,10578	46,67%	17,50%	55,00%	67,50%
12	10/25/25/22	0,06643	45,83%	17,50%	62,50%	57,50%
13	12/22/22/22	0,06405	36,67%	42,50%	27,50%	40,00%
13	12/25/25/22	0,04767	55,00%	52,50%	62,50%	50,00%
14	12/22/22/22	0,06419	40,00%	30,00%	40,00%	50,00%
14	12/25/25/22	0,07709	54,17%	47,50%	55,00%	60,00%

Percebe-se pelos resultados obtidos que as medidas acústicas de ruído turbulento apresentam capacidade inferior, corroborando as evidências estatísticas da seção 4.2. Verifica-se em algumas configurações que quando o critério de parada baseado no erro médio por padrão não é alcançado, o percentual de acerto é alto. Isto demonstra que a baixa quantidade de padrões de treinamento empregado (mesmo com aplicação de ruído na camada de entrada) deve ser contrabalançada com uma exigência de convergência mais “relaxada”, para que a generalização possa ocorrer.

4.4.3 Configurações empregando resíduos da filtragem inversa

Nesta seção analisa-se a capacidade da rede neural quando são utilizados parâmetros acústicos extraídos da filtragem inversa (realizada através de técnicas adaptativas). A quantidade de informação provida por estes algoritmos é, teoricamente, suficiente para que as características acústicas encontradas sejam identificadas, ou seja, perturbação na periodicidade do sinal de voz e presença de ruído em determinadas faixas de frequência que resultam na descaracterização espectral dos fonemas analisados.

Entretanto, as análises estatísticas que precedem a escolha das entradas dos classificadores neurais acusaram que algumas medidas acústicas como Pico (para avaliação da perturbação em amplitude do sinal residual, e indiretamente do sinal de voz) e SFR apresentam pouca capacidade discriminatória.

As configurações apresentadas nesta seção empregam os índices acústicos obtidos a partir um único algoritmo adaptativo. Apesar de ser verificado que a utilização combinada de diversos índices

para as técnicas de estimação residual acarretaria em melhor capacidade discriminatória, a utilização de um único algoritmo reduz o esforço computacional.

Configuração 15, contendo 9 entradas:

- Sexo e Fonema;
- Pico, P/A, Jitter, EX, SFR, SFF e SNR estimados da filtragem inversa através do algoritmo LMS.

Configuração 16, contendo 9 entradas:

- Sexo e Fonema;
- Pico, P/A, Jitter, EX, SFR, SFF e SNR estimados da filtragem inversa através do algoritmo gradiente descendente.

Configuração 17, contendo 9 entradas:

- Sexo e Fonema;
- Pico, P/A, Jitter, EX, SFR, SFF e SNR estimados da filtragem inversa através do filtro de Kalman com configuração 6/8 para coeficientes do modelo ARX.

Configuração 18, contendo 9 entradas:

- Sexo e Fonema;
- Pico, P/A, Jitter, EX, SFR, SFF e SNR estimados da filtragem inversa através do filtro de Kalman com configuração 6/4 para coeficientes do modelo ARX.

Configuração 19, contendo 9 entradas:

- Sexo e Fonema;
- Pico, P/A, Jitter, EX, SFR, SFF e SNR estimados da filtragem inversa através do filtro de Kalman com configuração 8/4 para coeficientes do modelo ARX.

As configurações apresentadas variam apenas no algoritmo adaptativo empregado. Apesar da quantidade inferior de medidas acústicas utilizadas para treinamento da rede neural, os resultados demonstram que o emprego de parâmetros acústicos que extraiam informações como o nível de periodicidade das pregas vocais e a quantidade de ruído proveniente de escape de ar permitem um maior número de identificações corretas de disfonias, comparando-se com as configurações apresentadas nas seções 4.4.1 e 4.4.2.

Tabela 141 - Classificação patológica por redes neurais para medidas residuais

Config.	Número de células por camada	Erro médio por padrão	% de acerto total	% de acerto para /a/	% de acerto para /e/	% de acerto para /i/
15	9/22/22/22	0,02264	63,00%	72,00%	60,00%	54,00%
15	9/25/25/22	0,01527	46,00%	42,00%	57,00%	39,00%
16	9/22/22/22	0,02136	51,00%	42,00%	42,00%	69,00%
16	9/25/25/22	0,01373	62,00%	45,00%	78,00%	63,00%
17	9/22/22/22	0,04251	66,00%	75,00%	57,00%	66,00%
17	9/25/25/22	0,01786	60,00%	54,00%	72,00%	54,00%
18	9/22/22/22	0,03247	48,00%	54,00%	42,00%	48,00%
18	9/25/25/22	0,02419	48,00%	60,00%	54,00%	30,00%
19	9/22/22/22	0,05030	60,00%	81,00%	51,00%	48,00%
19	9/25/25/22	0,02212	70,00%	63,00%	81,00%	66,00%

A taxa de acerto das redes neurais empregando medidas baseadas no sinal residual demonstram a necessidade de que um maior número de informações seja apresentado na entrada, mesmo com algumas redundâncias, para que possa classificar adequadamente as patologias.

4.4.4 Configuração mista

Buscando-se melhorar a capacidade discriminatória das redes neurais, duas configurações são apresentadas, agregando os parâmetros acústicos com maior potencial estatístico para identificação das patologias, e ainda considerando as configurações já apresentadas.

Claramente percebe-se que a elevação da quantidade de células na camada entrada é importante para aumento da taxa de acerto da rede neural. Outra evidência é que as avaliações de periodicidade do sinal de voz e nível de ruído turbulento são necessárias para que esta estrutura de classificação possa capturar as características que diferenciam as patologias.

Outro ponto importante é o manuseio cuidadoso dos parâmetros de aprendizado para que a generalização da rede neural seja mantida, seguido da redução do erro médio por padrão. Isto é necessário pois a quantidade de padrões por disfonia utilizada para treinamento é baixo.

Baseado nestas evidências, são apresentados os seguintes parâmetros de entrada para as redes neurais (que posteriormente serão definidas quanto à quantidade de neurônios nas camadas intermediárias):

Configuração 20, contendo 25 entradas:

- Sexo e Fonema;
- Desvio clássico da regressão linear, das perturbações de ordem zero, 1, 2, 3 e 4, do período da fundamental;
- Curtose das perturbações de ordem zero, 1, 2, 3 e 4, da amplitude da fundamental;

- Média e desvio clássico da regressão linear do primeiro e segundo harmônicos (H1 e H2);
- Mediana da estimativa CPP;
- Desvio clássico da regressão linear da estimativa P/A;
- Desvio clássico da regressão linear da estimativa HNR por técnica cepstral nas bandas de 60 à 400 Hz e 400 à 2000 Hz;
- Estimativa NNE nas bandas 1000 à 5000 Hz e 2000 à 5000 Hz.
- Desvio clássico da regressão linear da energia espectral nas bandas de 400 à 2000 Hz e 2000 à 5000 Hz.
- Estimativa PA e EX da filtragem inversa por algoritmo LMS;
- Estimativa Br.

Configuração 21, contendo 25 entradas:

- Sexo e Fonema;
- Desvio robusta da regressão linear, das perturbações de ordem zero, 1, 2, 3 e 4, do período da fundamental;
- Curtose das perturbações de ordem zero, 1, 2, 3 e 4, da amplitude da fundamental;
- Desvio robusta da regressão linear do primeiro e segundo harmônicos (H1 e H2);
- Mediana da estimativa CPP;
- Desvio robusta da regressão linear da estimativa P/A;
- Desvio robusta da regressão linear da estimativa HNR por técnica cepstral nas bandas de 60 à 400 Hz e 400 à 2000 Hz;
- Estimativa NNE nas bandas 1000 à 5000 Hz e 2000 à 5000 Hz.
- Desvio robusta da regressão linear da energia espectral nas bandas de 400 à 2000 Hz e 2000 à 5000 Hz;
- Estimativa PA e EX da filtragem inversa por algoritmo LMS;
- Estimativa Br.

A utilização destas medidas acústicas na composição do vetor de entrada do classificador neural deve-se aos resultados estatísticos apresentados em seções anteriores. Também é baseada na precisão obtida nas redes neurais em que se empregou separadamente grupos de medidas para perturbação de amplitude e período da fundamental, quantidade de ruído turbulento pela adução incompleta das pregas vocais, além das medidas extraídas do sinal residual (PA e EX).

A diferença entre estes dois conjuntos de valores está no uso de estimadores estatísticos clássicos (configuração 20) e robustos (configuração 21). O objetivo é verificar qual o comportamento da rede neural quando informações mais imunes à ruídos de estimação são empregadas para classificação de dados, mais particularmente, discriminação de patologias. A quantidade de acerto quando se aplica o conjunto de vozes de teste é apresentado na tabela a seguir.

Tabela 142 - Classificação patológica por redes neurais para medidas acústicas

Config.	Número de células por camada	Erro do último ciclo	% de acerto total	% de acerto para /a/	% de acerto para /e/	% de acerto para /i/
20a	25/35/22	0,03181	67,65%	61,76%	70,59%	70,59%
20b	25/12/12/22	0,16526	69,61%	52,94%	64,71%	91,18%
20c	25/22/22/22	0,00999	58,82%	58,82%	64,71%	52,94%
21a	25/35/22	0,00995	68,63%	58,82%	67,65%	79,41%
21b	25/12/12/22	0,04693	60,78%	55,88%	73,53%	52,94%
21c	25/22/22/22	0,00996	69,61%	79,41%	67,65%	61,76%

Pelos resultados obtidos, apesar da baixa quantidade de padrões treinados, o emprego de redes neurais, juntamente com a utilização de estimativas robustas e modificações no procedimento de cálculo de algumas medidas estatísticas permitem que patologias que afetam a laringe e as pregas vocais possam ser identificadas, probabilisticamente, através de análise da voz da pessoa examinada.

Pelas tabs. 139, 140, 141 e 142 pode-se verificar que as medidas de perturbação permitem uma boa avaliação das estruturas glotais, principalmente quando um número suficiente destas compõem o vetor de entrada. As medidas de relação sinal-ruído não apresentam a mesma eficiência na discriminação patológica. Provavelmente, o emprego de um maior número de sinais de voz para treinamento das redes neurais permita que um maior número de disfonias possam ser identificadas.

As estimativas robustas dos parâmetros acústicos permitem que a convergência das redes neurais seja mais rápida, pois a dispersão de seus valores é menor. Assim, as distribuições das patologias no espaço multivariável são leptocúrticas, permitindo uma distinção melhor destas disfonias pelo classificador (com base das vozes de treinamento).

O emprego de redes neurais artificiais como classificador de medidas acústicas para identificação de patologias mostra-se promissor. Entretanto, a variabilidade no percentual de sucesso destas estruturas depende da quantidade de vozes para treinamento. Isto porque os resultados apresentados nas tabs. 139, 140, 141 e 142 demonstram a existência de uma grande quantidade de mínimos locais, influenciada pela baixa quantidade de padrões de treinamento. O emprego de mais vozes por patologias pode representar um ponto chave na utilização desta metodologia para identificação de alterações orgânicas ou funcionais na laringe.

5. Conclusões

O diagnóstico automático de patologias do trato vocal empregando métodos baseados unicamente na voz do paciente ainda demanda estudos mais extensos pois as similaridades entre algumas disfonias, juntamente com a complexidade dinâmica do sistema produtor de fala ainda não produzem métodos 100% confiáveis.

Entretanto, pode-se verificar através do emprego de procedimentos estatísticos que a mediana e a regressão linear robusta na estimação de medidas acústicas baseadas no sinal de voz possibilitam que algumas disfonias sejam corretamente identificadas, como alguns casos de carcinoma, paralisias unilaterais, laringite, entre outras.

Considerando também o emprego de classificadores não-lineares como as redes neurais artificiais, pôde-se definir um método automático de identificação de patologias, que apesar de requisitar grande esforço computacional, tem a vantagem de ser simples, exigindo apenas que seja realizada a gravação e digitalização da voz através de placas comuns de aquisição de dados. Este instrumento é mais uma ferramenta que auxilia fonoaudiólogos e otorrinolaringologista no diagnóstico, ainda permitindo o acompanhamento da evolução terapêutica.

Este trabalho procurou evidenciar o uso de diversas ferramentas matemáticas que permitissem abordar adequadamente e solucionar o problema “identificação de patologias através de análise acústica”. Para tanto, empregou-se técnicas de processamento digital de sinais, algoritmos adaptativos, métodos estatísticos e classificadores neurais.

A bibliografia consultada apresenta um grande número de medidas acústicas baseadas em características temporais e espectrais do sinal de voz. A maioria destes artigos busca relacionar estes índices acústicos com características perceptuais como rouquidão, sopro e aspereza, que são fortes indicativos da existência de anormalidades na laringe e pregas vocais; assim justifica-se a quantidade de medidas acústicas analisadas, bem como variações estatísticas destas para verificar quais são potencialmente úteis para identificar diretamente as disfonias. Estas análises foram realizadas considerando individualmente cada uma das medidas acústicas disponíveis, conforme método descrito na seção 3.5. Apesar que o ideal seria empregar técnicas como os discriminantes de Fischer, análise bayesiana ou classificação hierárquica onde pode-se considerar a avaliação multivariada de todas as disfonias estudadas, a quantidade elevada de medidas acústicas dificultaria a análise da influência de cada parâmetro acústico na identificação das patologias. O método de Mann-Whitney, juntamente com uma pequena modificação para mensuração da distância entre as

distribuições de patologias em um espaço unidimensional permite que avalie-se, exaustivamente, todas as combinações dois-à-dois dos tipos de voz analisados, possibilitando a escolha adequada dos parâmetros acústicos que são mais eficientes na identificação das patologias.

Muitas das medidas acústicas apresentadas neste trabalho correspondem à modificações realizadas (geralmente algorítmicas) nos procedimentos descritos em artigos, pois muitos destes algoritmos envolvem a seleção manual do trecho “mais estável” deste sinal. Empregando-se algumas técnicas estatísticas, avaliou-se inclusive o comportamento de cada índice acústico ao longo do sinal de voz. Os resultados nas classificações (estatística e por redes neurais) ilustram inclusive que alterações das medidas acústicas ao longo do sinal de fala (5 segundos de vocalização sustentada), representados por estimativas de dispersão, apresentam uma relação mais direta com diversas patologias.

Considerando as medidas de perturbação da frequência fundamental (seja em período ou amplitude, ciclo-a-ciclo), verificou-se que as medidas de desvio para derivadas discretas (ou perturbação) de ordem superior são extremamente úteis para diagnóstico automático de patologias. Outra evidência não encontrada em artigos consultados é que a curtose das medidas de perturbação em amplitude auxilia eficazmente na discriminação de disfonias. Com o agravamento das disfonias, a distribuição das amplitudes da frequência fundamental torna-se mais platicúrtica, ou seja, as oscilações destas amplitudes torna-se elevada.

Para avaliar o nível de turbulência do fluxo de ar através da glote, foram analisados muitos procedimentos matemáticos encontrados na literatura. Algumas medidas como HNR, Ra e Hr mostraram-se pouco eficientes para a tarefa de discriminação patológica pois conceitualmente baseiam-se em características visuais da curva de densidade espectral de potência. Isto pode ser comprovado pela avaliação estatística (por Mann-Whitney) apresentada no capítulo 4, onde um pequeno número de pares patológicos é separável através destes parâmetros.

Outras medidas acústicas, que são descritas matematicamente nos respectivos artigos, como HNR por técnica cepstral, NNE e SNL são extremamente úteis na avaliação da quantidade de ruído e desta forma, identificar patologias como cistos, nódulos e pólipos, todas definidas perceptualmente como soprosas.

Avaliações da quantidade de energia do sinal de voz também mostra-se útil na discriminação patológica pois permite avaliar o nivelamento na curva de densidade espectral de potência. Sabe-se que com o aumento da quantidade de ruído na glote, as estruturas formantes são descaracterizadas, juntamente com a elevação das componentes não harmônicas e as de alta frequência. Isto explica a versatilidade da avaliação da energia em determinadas bandas espectrais, principalmente entre 400 e 5000 Hz, onde pode-se verificar mais claramente estas alterações espectrais conforme a disфонia e sua extensão.

O emprego da técnica de filtragem inversa também propicia a identificação de um grande número de patologias, com reduzida quantidade de medidas acústicas extraídas do sinal residual.

Foram propostos o emprego de filtros adaptativos (Wiener e Kalman) para permitir que um grande trecho do sinal de voz seja analisado, evitando-se novamente a demarcação de um trecho “mais estável” para estimativa de medidas acústicas. Apesar do volume extraordinário de cálculos exigidos por esta metodologia, os resultados indicam que são potencialmente úteis na identificação de anormalidades na laringe e pregas vocais, pois fornecem parâmetros acústicos que extraem o nível de periodicidade do sinal de voz juntamente com estimativas da quantidade de ruído turbulento. Considerando o filtro de Kalman, constata-se que este ainda merece futuros estudos, principalmente em relação aos coeficientes da porção exógena. Acredita-se que durante o processo de decorrelação do sinal residual, informações sobre o comportamento do trato vocal podem estar sendo armazenadas nestes coeficientes, o que pode resultar em um conjunto de novos parâmetros acústicos estimados diretamente do filtro e não depender de análise posterior do sinal residual.

Muitas variações sobre os índices acústicos estudados podem ser realizadas em trabalho futuros como alteração dos comprimentos da janela de dados e da sobreposição inter-janela, variação da quantidade de coeficientes dos filtros adaptativos, estimativa de ruído em diversas bandas espectrais (válido para a maioria das medidas acústicas de ruído analisadas) e emprego de estimativas de períodos da frequência fundamental mais robustos.

A partir do levantamento estatístico das medidas acústicas com melhor capacidade de discriminação das disfonias, empregou-se redes neurais artificiais para identificar as patologias. A escolha desta técnica deve-se às características de generalização, abstração das características acústicas encontradas nas patologias da laringe, além da capacidade de representar implicitamente modelos não-lineares. A similaridade entre as disfonias, juntamente com a complexidade e redundância do trato vocal na produção da fala indicam a presença de não-linearidades.

A utilização de redes neurais artificiais capacita a identificação de patologias a partir de medidas extraídas do sinal de voz. O fato de se apresentar estimativas para a rede neural permite que estas tenham um conjunto de entradas bastante reduzido, ao invés de apresentar-se o “sinal de voz” diretamente. O motivo desta abordagem é tentar ensinar a rede a identificar as características que o especialista de voz normalmente busca quando examina um paciente.

O modelo empregado (*feedforward* com treinamento por *backpropagation*) foi considerado por ser o mais difundido. Entretanto, sabe-se de antemão que alguns requisitos seriam necessário para que a rede neural apresente como resultado a probabilidade de ocorrência de uma determinada disфонia a partir de um conjunto de medidas acústicas, é necessário que obtenha-se um mínimo global com um conjunto suficiente de padrões para treinamento. O primeiro requisito pode ser alcançado com o modelo neural utilizado, apesar da necessidade de controlar-se adequadamente a taxa de aprendizado, ou pelo emprego de RBF (*radial basis function*) ou outras arquiteturas que considerem a sobreposição de padrões das disfonias.

Já o requisito da quantidade de padrões para treinamento exige que mais pacientes sejam examinados por especialistas e suas vozes, gravadas. Apesar da quantidade de padrões utilizada neste

trabalho estar aquém do ideal, a metodologia empregada para identificação das patologias mantém-se válida, motivada principalmente pela robustez das medidas acústicas. O aumento da quantidade de padrões possibilitará que a quantidade de acertos por parte das estruturas neurais possa ser incrementada, permitindo uma confiabilidade maior dos instrumentos que possam ser construídos a partir das técnicas apresentadas.

Este trabalho demonstra que o pré-diagnóstico automático das disfonias pode ser realizado apesar da necessidade de que mais estudos sejam realizados para compreensão melhor dos aspectos dinâmicos da produção da fala e as modificações que as patologias provocam neste sistema orgânico. Novos índices acústicos podem ser desenvolvidos para capturar alterações temporais e espectrais, enquanto que os requisitos de memória e tempo de processamento dos parâmetros acústicos apresentados são passíveis de redução. O potencial dos instrumentos construídos a partir das técnicas apresentadas resultam na sua utilização por especialistas da voz como meio de prevenção de disfonias, pois com a melhora da eficiência dos modelos matemáticos utilizados poder-se-á identificar estas patologias nos estágios iniciais. A redução do tempo e do custo dos exames para otorrinolaringologistas e fonoaudiólogos possibilita que um maior número de pacientes possam ser examinado.

Bibliografia

- AHMED, N. (1987). Adaptive Filtering, In: Elliot, D. F. (Ed.), *Handbook of Digital Signal Processing - Engineering Applications*, Academic Press Inc., San Diego.
- ASKENFELT, A. G. E HAMMARBERG, B. (1986). Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. *Journal of Speech and Hearing Research*, 29, 50-64.
- ASKENFELT, A. GAUFFIN, J. SUNDBERG, J. E KITZING, P. (1980). A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency. *Journal of Speech and Hearing Research*, 23, 258-273.
- AWAN, S. N. E FRENKEL, M. L. (1992). Improvements in estimating the harmonics-to-noise ratio of the voice. *Journal of Voice*, 8, 255-262.
- BAKEN, R. J. (1996). *Clinical Measurement of Speech and Voice*, Singular Publishing Group Inc., San Diego.
- BENDAT, J. S. E PIERSON, A. G. (1980). *Engineering Application of Autocorrelation and Spectral Analysis*, John Wiley & Sons, Inc., New York.
- BIELAMOWICZ, S., KREIMAN, J., GERRATT, B. R., DAUER, M. S. E BERKE, G. S. (1996). Comparison of voice analysis systems for perturbation measurement. *Journal of Speech and Hearing Research*, 39, 126-134.
- BLESS, D. M. (1991). Measurement of vocal function. *Otolaryngologic Clinics of North America*, 24, 1023-1033.
- BOONE, D. R. E MCFARLANE, S. C. (1994). *A Voz e a Terapia Vocal*, Artes Médicas, Porto Alegre.
- BOX, G. E. P. E JENKINS, G. M. (1976) *Time Series Analysis - Forecasting and Control*, Revised Edition, Prentice-Hall Inc., Englewood Cliffs.
- BRIGHAM, E. O. (1988) *The Fast Fourier Transform and Its Applications*, Prentice-Hall Inc., Englewood Cliffs.
- CADZOW, J. A. (1994). *Discrete-Time Systems - An Introduction with Interdisciplinary Applications*, Prentice-Hall Inc., Englewood Cliffs.
- CAVALLO, S. A., BAKEN, R. J. E SHAIMAN, S. (1984). Frequency perturbation characteristics of pulse register phonation. *Journal of Communication Disorders*, 17, 231-243.
- CHILDERS, D. G. E BAE, K. S. (1992). Detection of laryngeal function using speech and electroglottographic data. *IEEE Transactions on Biomedical Engineering*, 39, 19-25.
- CHILDERS, D. G., HICKS, D. M., MOORE, G. P., ESKENAZI, L. E LALWANI, A. L. (1990). Electroglottography and vocal fold physiology. *Journal of Speech and Hearing Research*, 33, 245-254.
- CRYSTAL, T. H. E JACKSON, C. L. (1970). Extracting and processing vocal pitch for laryngeal disorder detection. *Journal of the Acoustical Society of America*, 48, 1, n. 1, parte 1, 118.
- DAVIS, M. H. A. E VINTER, R. B. (1985). *Stochastic Modelling and Control*, Chapman and Hall Ltd, London.
- DAVIS, S. B. (1970). Preliminary results using inverse filtering of speech for automatic evaluation of laryngeal pathology. *Journal of the Acoustical Society of America*, 58, suppl. 1, S111-S112.

- DAVIS, S. B. (1979). Acoustic characteristics of normal and pathological voices. In Lass, N. J. (Ed.) *Speech and Language: Advances in Basic Research and Practice*, vol. 1, New York Academic Press, 271-335.
- DEAL, R. E. E EMANUEL, F. W. (1978). Some waveform and spectral features of vowel roughness. *Journal of Speech and Hearing Research*, 21, 250-264.
- DELLER JR., J. R., PROAKIS, J. G. E HANSEN, J. H. L. (1993). *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., New York.
- DUBNOWKI, J. J. R., SCHAFER, R. W. E RABINER, L. R. (1976). Real-time digital hardware pitch detector, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24, 2-8.
- DUIFHUIS, H., WILLEMS, L. F. E SLUYTER, R. J. (1982). Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *Journal of the Acoustical Society of America*, 71, v. 6, 1568-1579.
- EMANUEL, F. W. E WHITEHEAD, R. L. (1979). Harmonic levels and vowel roughness. *Journal of Speech and Hearing Research*, 22, 829-840.
- EMANUEL, F. W., LIVELY, M. A. E MCCOY, J. F. (1973). Spectral noise levels and roughness ratings for vowels produced by males and females. *Folia Phoniatrica*, 25, 110-120.
- ESKENAZI, L., CHILDERS, D. G. E HICKS, D. M. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research*, 33, 298-306.
- EUSTACE, C. S., STEMPLE, J. C. E LEE, L. (1996). Objective measures of voice production in patients complaining of laryngeal fatigue. *Journal of Voice*, 10, 146-154.
- FEX, B., FEX, S., SHIROMOTO, O. E HIRANO, M. (1994). Acoustic analysis of functional dysphonia: before and after voice therapy (accent method). *Journal of Voice*, 8, 163-167.
- FISHER, W. M., MONSEN, R. B. E ENGBRETSON, A. M. (1975). Variations in the normal male glottal wave. *Journal of the Acoustical Society of America*, 58, suppl. 1, S41.
- FROKJAER-JENSEN, B. E PRYTZ, S. (1976). Registration of voice quality, *B & K Technical Review*, 3-17.
- FUKAZAWA, T., EL-ASSUOOTY, A. E HONJO, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *Journal of the Acoustical Society of America*, 83, 1189-1193.
- GECKINLI, N. E YAVUZ, D. (1976). Algorithm for pitch extraction using zero-crossing interval sequence. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 25, 559-564.
- GELFER, M. P. (1995). Fundamental frequency, intensity and vowel selection: effects on measures of phonatory stability. *Journal of Speech and Hearing Research*, 38, 1189-1198.
- GUBRYNOWICZ, R., MIKIEL, W. E ZARNECKI, P. (1980). Na acoustic method for the evaluation of the state of the larynx source in case involving pathological changes in the vocal folds. *Archives of Acoustics*, 5, 3-30.
- HALL, K. D. (1995). Variations across time in acoustic and electroglottographic measures of phonatory function in women with and without vocal nodules. *Journal of Speech and Hearing Research*, 38, 783-793.
- HAMLET, S. L. E GEOFFREY, V. C. (1976). Effect of a dental prosthesis on speaker - specific characteristics of voice. *Journal of Speech and Hearing Research*, 19, 639-650.
- HARTMANN, E. E CRAMON, D. V. (1984). Acoustic measurement of voice quality in central dysphonia. *Journal of Communication Disorders*, 17, 425-440.
- HAYKIN, S. (1994). *Neural Networks*, MacMillan Publishing Co., Englewoods Cliffs.
- HEBB, D. O. (1949). *The Organization of Behavior*, John Wiley and Sons, Inc., New York.

- HECKER, M. H. L. E KREUL, E. J. (1970). Description of the speech of patients with câncer of the vocal folds. Part I: Measures of fundamental frequency. *Journal of the Acoustical Society of America*, 49, n. 4, parte 2, 1275-1282.
- HEMERLY, E. M. (1996). *Controle por Computador de Sistemas Dinâmicos*, Editora Edgard Blücher Ltda, São Paulo.
- HERTZ, J., KROGH, A., PALMER, R. G. (1986). *Introduction to the Theory of Neural Computation*, Addison-Wesley Publishing Co., Redwood City.
- HERZEL, H., BERRY, D., TITZE, I. R. E SALEH, M. (1994). Analysis of vocal disorders with methods from nonlinear dynamics. *Journal of Speech and Hearing Research*, 37, 1008-1019.
- HESS, W. (1995). Determination of glottal excitation cycle in running speech. *Phonetica*, 52, 196-204.
- HIKI, S. E IMAIZUMI, S. (1970). Acoustical analysis for voice disorders: An attempt towards clinical applications. *Journal of the the Acoustical Society of America*, 58, suppl. 1, S111.
- HILLENBRAND, J., CLEVELAND, R. A. E ERICKSON, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37, 769-778.
- HIRAOKA, N., KITAZOE, Y., UETA, H, TANAKA, S. E TANABE, M. (1984). Harmonic-intensity analysis of normal and hoarse voices. *Journal of the Acoustical Society of America*, 76, 1648-1651.
- HOEL, P. G. (1971) *Introduction to Mathematical Statistics*, John Wiley & Sons, Inc., New York.
- HOLMES, J. N. (1975). Low-frequency phase distortion of speech recordings. *Journal of the Acoustical Society of America*, 1975, 747-749.
- HONIG, M. L. E MESSERSCHMITT, D. G. (1984). *Adaptive Filters: Structures, Algorithms and Applications*, Kluwer Academic, Boston.
- HONJO, I. E ISSHIKI, N. (1980). Laryngoscopic and voice characteristics of aged persons, *Archives of Otolaryngology*, 106, 149-150.
- HORII, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. *Journal of Speech and Hearing Research*, 22, 5-19.
- HORII, Y. (1982). Jitter and shimmer differences among sustained vowel phonations. *Journal of Speech and Hearing Research*, 25, 12-14.
- IMAZUMI, S. (1979). Acoustic measures of roughness in pathological voice. *Journal of Speech and Hearing Research*, 23, 202-209.
- ISHIZAKA, K. (1975). Computer simulation of pahologic vocal cord vibration. *Journal of the Acoustical Society of America*, 58, suppl. 1, S111.
- ISSHIKI, N. YANAGIHARA, N. E MORIMOTO, M. (1966). Approach to the objective diagnosis of hoarseness. *Folia Phoniatica*, 18, 393-400.
- ISSHIKI, N., KITAGIMA, K., KOJIMA, H. E HARITA, Y. (1978). Turbulent noise in dysphonia. *Folia Phoniatica*, 30, 214-224.
- ISSHIKI, N., OKAMURA, H., TANABE, M. E MORIMOTO, M. (1969). Differential diagnosis of hoarseness. *Folia Phoniatica*, 21, 9-19.
- IWATA, S. (1972). Periodicities of pitch perturbations in normal and pathologic larynges. *Laryngoscope*, 82, 87-96.
- JACKSON, L. B. (1989) *Digital Filters and Signal Processing*, 2 ed., Kluwer Academic Publishers, Boston.
- JOHN, P. W. M. (1990). *Statistical Methods in Engineering and Quality Assurance*, John Wiley & Sons Inc., London.

- JOHNSON, R. A. E WICHERN, D. W. (1992). *Applied Multivariate Statistical Analysis*, Prentice-Hall Inc., Upper Saddle River.
- KASUYA, H., OGAWA, S., MASHIMA, K E EBIHARA, S. (1986). Normalized noise energy as a acoustic measure to evaluate pathologic voice. *Journal of the Acoustical Society of America*, 80, 1329-1334.
- KAY, S. M. (1988). *Modern Spectral Estimation - Theory and Application*, Prentice-Hall Inc., Englewood Cliffs.
- KLINGHOLZ, F, E MARTIN, F. (1985). Quantitative spectral evaluation of shimmer and jitter. *Journal of Speech and Hearing Research*, 28, 169-174.
- KOIKE, Y. (1969). Vowel amplitude modulations in patients with laryngeal diseases. *Journal of the Acoustical Society of America*, 45, 839-844.
- KOIKE, Y. (1973). Applications of some acoustic measures for the evaluation of laryngeal dysfunction, *Studia Phonologica*, 7, 17-23.
- KOIKE, Y. E MARKEL, J. (1975). Application of inverse filtering for detecting laryngeal pathology. *Annals of Otolaryngology*, 84, 117-124.
- KOIKE, Y., TAKAHASHI, H. E CALCATERRA, T. C. (1977). Acoustic measures for detecting laryngeal pathology, *Journal of Phonetics*, 14, 517-524.
- KOJIMA, H., GOULD, W. J., LAMBIASE, A. E ISSHIKI, N. (1980) Computer analysis of hoarseness, *Acta Otolaryngologica*, 89, 547-554.
- KOUFMAN, J. A. (1991). Approach to the patient with a voice disorder. *Otolaryngologic Clinics of North America*, 24, 989-998.
- KOUFMAN, J. A. E ISAACSON, G. (1991). The spectrum of vocal dysfunction. *Otolaryngologic Clinics of North America*, 24, 985-988.
- KROM, G. DE (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*, 36, 254-266.
- KROM, G. DE (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research*, 38, 794-811.
- LATHI, B. P. (1968). *An Introduction to Random Signals and Communication Theory*, International Textbook Company, Scranton.
- LEVINSON, S. E. E SCHMIDT, C. E. (1983). Adaptive computation of articulatory parameters from the speech signal. *Journal of the Acoustical Society of America*, 74, 1145-1154.
- LIEBERMAN, P. (1961). Perturbations in vocal pitch, *Journal of the Acoustical Society of America*, 33, 597-603.
- LIEBERMAN, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges, *Journal of the Acoustical Society of America*, 35, 344-353.
- LIVELY, M. A. E EMANUEL, F. W. (1970). Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. *Journal of Speech and Hearing Research*, 13, 498-502.
- MAKSYM, J. N. (1973). Real-time pitch extraction by adaptive prediction of the speech waveform. *IEEE Transactions on Audio and Electroacoustics*, 21, 149-160.
- MARPLE, S. L., JR. (1987) *Digital Spectral Analysis with Applications*, Prentice-Hall Inc., Englewood Cliffs.
- MARTIN, D., FITCH, J. E WOLFE, V. (1995). Pathologic voice type and the acoustic prediction of severity. *Journal of Speech and Hearing Research*, 38, 765-771.

- MATSUSHITA, H., HIRANO, M. E KAWASAKI, H. (1970). Vibration of the vocal cords in some pathological conditions. *Journal of the Acoustical Society of America*, 58, suppl. 1, S112.
- MCCANDLESS, S. S. (1974). An algorithm for automatic formant extraction using linear prediction spectra. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 22, 135-141.
- MILENKOVIC, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research*, 30, 529-538.
- MOORER, J. A. (1974). The optimum comb method of pitch period analysis of continuous digitized speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 22, 330-338.
- MURRY, T. (1975). Some acoustic features of hoarseness. *Journal of the Acoustical Society of America*, 58, suppl. 1, S111.
- MURRY, T. E DOHERTY, E. T. (1980). Selected acoustic characteristics of pathologic and normal speakers. *Journal of Speech and Hearing Research*, 23, 361-369.
- MUTA, H., BAER, T., WAGATSUMA, K., MURAOKA, T. E FUKUDA, H. (1988). A pitch-synchronous analysis of hoarseness in running speech. *Journal of the Acoustical Society of America*, 84, 1292-1301.
- NARANJO, N. V., LARA, E. M., RODRÍGUEZ, I. M. E GARCÍA, G. C. (1994). High-frequency components of normal and dysphonic voice. *Journal of Voice*, 8, n. 2, 157-162.
- NITTROUER, S., MCGOWAN, R., MILENKOVIC, P. H. E BEEHLER, D. (1990). Acoustic measurements of men's and women's voices: a study of context effects and covariation. *Journal of Speech and Hearing Research*, 33, 761-775.
- NOLL, A. M. (1964). Short-time spectrum and "cepstrum" techniques for vocal-pitch detection. *Journal of the Acoustical Society of America*, 36, 296-302.
- NOLL, A. M. (1966). Cepstrum pitch determination. *Journal of the Acoustical Society of America*, 41, 293-309.
- OPPENHEIM, A. V. E SCHAFER, R. W. (1976) *Digital Signal Processing*, Prentice-Hall Inc., Englewood Cliffs.
- OPPENHEIM, A. V. E SCHAFER, R. W. (1989), *Discrete-Time Signal Processing*, Prentice-Hall Inc., Englewood Cliffs.
- PARSONS, T. W. (1986). *Voice and Speech Processing*, McGraw-Hill, New York.
- PINTO, N. B. E TITZE, I. R. (1990). Unification of perturbation measures in speech signals. *Journal of the Acoustical Society of America*, 87, 1278-1289.
- POTTER, R., KOPP, G. E GREEN, H. (1947). *Visible Speech*, D. Van Nostrand Company, Inc., New York.
- PROSEK, R. A., MONTGOMERY, A. A., WALDEN, B. E. E HAWKINS, D. B. (1987). An evaluation of residue features as correlates of voice disorders. *Journal of Communication Disorders*, 20, 105-107.
- RABINER, L. R. E SCHAFER, R. W. (1978). *Digital Processing of Speech Signals*, Prentice-Hall Inc., Englewood Cliffs.
- READ, C., BUDER, E. H. E KENT, R. D. (1990). Speech analysis systems: a survey. *Journal of Speech and Hearing Research*, 33, 363-374.
- READ, C., BUDER, E. H. E KENT, R. D. (1992). Speech analysis systems: an evaluation. *Journal of Speech and Hearing Research*, 35, 314-332.
- REMACLE, M. E TRIGAUX, I. (1991). Characteristics of nodules through the high-resolution frequency analyzer. *Folia Phoniatrica*, 43, 53-59.

- RODRIGUES, S., BEHLAU, M. E PONTES, P. (1994). Proporção harmônica-ruído: valores para indivíduos adultos brasileiros. *Acta Awho*, 13, n. 3, 112-116.
- RONTAL, E., RONTAL, M. E ROLNICK, M. I. (1975). Objective evaluation of vocal pathology using voice spectrography. *Annals of Otolaryngology*, 84, 662-671.
- ROSS, M. J., SHAFFER, H. L., COHEN, A., FREUDBERG, R. E MANLÉY, H. J. (1974). Average magnitude difference function pitch extractor. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 5, 353-362.
- ROSSITER, D. E HOWARD, D. M. (1994). Albert: A system for interactive analysis and display of voice source and acoustic parameters. *Proceedings of the Institute of Acoustics : Speech and Hearing Conference*, November.
- ROTHENBERG, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *Journal of the Acoustical Society of America*, 53, n. 6, 1632-1645.
- ROTHENBERG, M. E MAHSHIE, J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research*, 31, 338-351.
- RUMMELHART, D. E. E MCCLELLAND, J. L., eds. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol 1., MIT Press, Cambridge.
- SANDERSON, R. J. E MARAN, A. G. D. (1992). The quantitative analysis of dysphonia. *Clinical Otolaryngology and Allied Sciences*, 17, 440-443.
- SANSONE, F. E. E EMANUEL, F. W. (1970). Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. *Journal of Speech and Hearing Research*, 13, 472-488.
- SANTINA, M. S., STUBBERUD, A. R. E HOSTETTER, G. H. (1994). *Digital Control System Design*, Saunders College Publishing, Orlando.
- SCHOENTGEN, J. E GUCHTENEERE, R. DE (1995). Time series analysis of jitter. *Journal of Phonetics*, 23, 189-201.
- SCHROEDER, M. R. (1968). Period histogram and product spectrum: New methods for fundamental frequency measurement, *Journal of the Acoustical Society of America*, 43, 829-834.
- SHIPP, THOMAS E MCGLONE, R. E. (1970). Comparison of sustained vocal fry and terminal breath group phonation. *Journal of the Acoustical Society of America*, 48, 1, n. parte 1, 119.
- SHOJI, K., REGENBOGEN, E., YU, J. D, E BLAUGRUND, S. M. (1992-1). High-frequency power ratio of breathy voice. *Laryngoscope*, 102, 267-271.
- SHOJI, K., REGENBOGEN, E., YU, J. D., BLAUGRUND, S. M. (1992-2). H-Index: A new measure of glottal efficiency for the pathologic voice. *Laryngoscope*, 102, 1113-1117.
- SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Kogakusha Ltd., Tokyo.
- SMITH, M. E., BERKE, G. S., GERRATT, B. R. E KREIMAN, J. (1992). Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration. *Journal Speech and Hearing Research*, 35, 545-554.
- SÖDERSTRÖM, T. (1994). *Discrete-Time Stochastic Systems - Estimation and Control*, Prentice-Hall International Limited, Cambridge.
- SONDHI, M. M. (1968). New methods of pitch extraction. *IEEE Transactions on Audio and Electroacoustics*, 16, 262-266.
- SORDERSTEN, M E LINDESTAD, P. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research*, 33, 601-611.
- SORENSEN, D. E HORII, Y. (1983). Frequency and amplitude perturbation in the voice of female speakers. *Journal of Communication Disorders*, 16, 57-61.

- SORENSEN, D. E HORII, Y. (1984). Directional perturbation factors for jitter and for shimmer. *Journal of Communication Disorders*, 17, 143-151.
- SVEC, J. G, SCHUTTE, H. K E MILLER, D. G. (1996). A Subharmonic vibratory pattern in normal vocal folds. *Journal of Speech and Hearing Research*, 39, 135-143.
- TAKAHASHI, H. E KOIKE, Y. (1975). Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Oto-Laryngologica*, suppl. 338, 1-24.
- TITZE, I. R. E LIANG, H. (1993). Comparasion of F0 extraction methods for high-precision voice perturbation measurements. *Journal Speech and Hearing Research*, 1120-1133.
- TITZE, I. R., HORII, Y. E SCHERER, R. C. (1987). Some technical considerations in voice perturbation measurements. *Journal of Speech and Hearing Research*, 30, 252-260.
- TRULLINGER, R. W., EMANUEL, F. W., SKENES, L. L. E MALPASS, J. C. (1988). Spectral noise level measurements used to track voice improvement in one patient. *Journal of Communication Disorders*, 21, 447-457.
- UN, C. K. E YANG, S. (1977). A pitch extraction algorithm based on LPC inverse filtering and AMDF. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 25, 565-572.
- WENDAHL, R. W. (1963). Laryngeal analog synthesis of harsh voice quality. *Folia Phoniatica*, 15, 241-250.
- WENDAHL, R. W. (1966). Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. *Folia Phoniatica*, 18, 98-108.
- WIDROW B. E HOFF, M. E. (1960). Adaptive Switching Circuits, In: *1960 IRE WESCON Convention Record*, part 4, 96-104, IRE, New York.
- WINHOLTZ, W. S. E RAMIG, L. O. (1992). Vocal tremor analysis with the vocal demodulator. *Journal of Speech and Hearing Research*, 35, 562-573.
- WOLFE, V. E STEINFATT, T. M. (1987). Prediction of vocal severity within and across voice types. *Journal of Speech and Hearing Research*, 30, 230-240.
- YANAGIHARA, N. (1967). Significance of harmonic changes and noise components in hoarseness. *Journal of Speech and Hearing Research*, 10, 534-543.
- YUMOTO, E., GOULD, W. J. E BAER, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, 71, 1544-1549.

Glossário

<i>Pitch</i>	Termo comumente associado à frequência fundamental.
<i>Jitter</i>	Denomina a perturbação, ciclo-a-ciclo, da frequência fundamental.
<i>Shimmer</i>	Perturbação, ciclo-a-ciclo, das amplitudes periódicas do sinal de voz.
<i>Short-term</i>	Referente à utilização de janelas sobre um sinal no desenvolvimento matemático.
<i>Long-term</i>	Referente à utilização de toda extensão do sinal no desenvolvimento matemático.
<i>Aliasing</i>	Sobreposição espectral quando a taxa de amostragem não obedece os requisitos de Nyquist.
Espectrograma	Gráfico tridimensional onde relaciona-se tempo x frequência x energia.
<i>Lag</i>	Intervalo temporal.
<i>Cepstrum</i>	Equivalente do espectro densidade de potências no domínio cepstral.
<i>Quefreny</i>	Equivalente da frequência no domínio cepstral.
<i>Liftering</i>	Operação de filtragem no domínio cepstral.
<i>Rharmônico</i>	Termo referente às componentes harmônicas no domínio cepstral.
Leptocúrtico	Termo indicando que uma distribuição de amostras é menos achatada que a distribuição normal.
Platicúrtico	Termo indicando que uma distribuição de amostras é mais achatada que a distribuição normal.
Curtose	Medida estatística que emprega o momento de quarta ordem para avaliar grau de achatamento da função distribuição.
<i>Peak-picking</i>	Técnica para estimação da frequência fundamental, ciclo-a-ciclo, pela busca dos pontos de máximo no sinal de voz.
<i>Zero-crossing</i>	Técnica para estimação da frequência fundamental, ciclo-a-ciclo, pela busca dos pontos onde o sinal de voz cruza o eixo das abscissas (zero).
<i>Clusters</i>	Agrupamentos no espaço n-dimensional que apresentam alguma característica matemática que os assemelha.
<i>Feedforward</i>	Tipo de rede neural onde as informações são propagadas no sentido entrada-saída.
<i>Backpropagation</i>	Técnica de treinamento de redes neurais onde a informação de erro é propagada no sentido saída-entrada.
<i>Spectral flattness</i>	Indicador do grau de achatamento (tendência ao ruído branco) de uma curva de densidade espectral de potência.

Apêndice

Pathologic Discrimination through Acoustic Analysis of Voice

Marcelo de Oliveira Rosa, José Carlos Pereira
School of Engineering of São Carlos
University of São Paulo - Brazil
Av. Dr. Carlos Botelho, 1465, C.P. 359
13560-250 - São Carlos - São Paulo - Brazil
{marceloI, pereira}@lim.sel.eesc.sc.usp.br

Simone Adad Araújo, Marcos Grellet
Faculty of Medicine of Ribeirão Preto
University of São Paulo - Brazil
Av. dos Bandeirantes, 3900
14049-900 - Ribeirão Preto - São Paulo - Brazil

Abstract

The acoustic analysis of voice from dysphonic patients has been studied, mainly for classification of voice quality in terms of roughness, breathiness, and hoarseness. By a data reduction of speech signal, researchers have produced several acoustic measures in order to relate them with a perceptual diagnosis. In this work, a proposal was made to modify acoustic parameters in order to support a higher sample quantity, instead of using only a data window by digitalized speech. Also, robust statistics is used to obtain reliable measures that estimate the abnormal condition of a voice signal as well as to discriminate several pathologies of the larynx and vocal cords, without any perceptual consideration. Using only seven measures from adaptive inverse filtering (Kalman and Wiener filters) and a neural network for pattern classification on 228 voice signals of speakers (20 groups of distinct types of pathologies and one of normal speakers) assessed through videolaryngoscopy, results of the pathologic discrimination showed 22% of certainty. Redrawing the patient sample in seven distinct groups (five with defined pathologies, one with all others pathologies, and one control), then the error in the pathologic discrimination is 18%.

This work is supported by Capes

1. Introduction

Traditionally, the techniques of digital signal processing have been used for synthesis and speech recognition, besides noise reduction and automatic echo canceling. These techniques that are used for voice analysis with purpose of pathologic characteristics identification that affect the vocal tract, has been developed with objective to assist voice specialist as laryngologists and speech pathologists on dysphonias diagnosis like nodules, polyps and carcinomas.

Considering that the principal characteristics evaluated during the medical diagnosis are the

periodicity of vocal folds vibration and the level of noise deriving from the turbulent air flow in the glottis, it attempts to determine measurements extracted from voice signal are determined in attempting to relate them with perceptual aspects as breathiness, roughness and hoarseness levels.

Through the fundamental frequency perturbation, Lieberman (1961, 1963) found some organic alterations in the vocal folds. Koike (1973) verified that many slow oscillations on fundamental periods are found in normal speakers and also affect the jitter estimations (perturbation of the F_0). Using moving average to solve this problem, he defined the RAP acoustic measurement (Relative average perturbation), in which different length windows (3 and 5) are used for jitter and shimmer assessment, respectively.

The indirect relationship between the fundamental frequency and the jitter mean estimation (average value of the successive differences of pitch periods) led Horii (1979) to define the JF index (jitter factor) as the ratio between mean jitter by mean fundamental frequency.

For assessment of pathologies whose voices are identified like roughness, Deal and Emanuel (1978) showed that this perceptual characteristic presents a direct correlation with the PVI (period variability index) and AVI (amplitude variability index) parameters whose mathematical procedure is based on variance coefficient of the period and amplitude estimation of glottal wave.

Recently, Pinto and Titze (1995) estimated the perturbation measures of glottal signal regarding discrete derivatives of superior order, with the aim to standardize these evaluations of vocal folds movement.

To evaluate the turbulent noise, Shoji et al (1991) measured the mean energy on power spectral density curve of the voice signal starting from 6 kHz, and found the correlation between this acoustical parameter (HFPR or High Frequency Power Ratio) and breathiness voices.

Considering a voice signal with additive noise, Yumoto, Gould and Baer (1982) defined the HNR measurement, using 50 consecutive pitch periods of a sustain vowel to determine a mean cycle of voice signal

without noise. Awan and Frenkel (1994) presented a procedure to improve the resolution of this parameter.

The breathiness characteristic maybe evaluated through the Br acoustical measurement (breathiness ratio) of Fukazawa et al (1988) that determine the energy of high frequency components estimated from second derivate on digital signal.

Using an adaptive comb filtering method in frequency domain, Kasuya et al (1986) define a robust measurement (NNE or Normalized Noise Energy) for assessment of the noised components of signals deriving from pathologies as carcinoma, nodules, polyps and unilateral paralysis.

Krom (1993), using cepstral technique for evaluation of the ratio between harmonic and noised components, defined a new method to HNR estimation that uses a comb-filtering in cepstrum domain. This method evaluate, simultaneously, the noise level and jitter amount of the voice signal.

Another methodology to assess the vocalization structures, suggested by Davis (1979), uses the inverse filtering for determination of the residue signal. Based on speech production model (Deller, 1993), the AR coefficients are estimated through Levinson algorithm for most stable segment of sustained vowel and thus the parametric estimation error is obtained, i.e., it is determined the residue to be analysed.

The signal proprieties are evaluated from acoustical measurements of periodicity (APQ and PPQ), noise level (SFF and SFR, respectively, Spectral Flatness of Filter and Spectral Flatness of Residue) besides EX (excess coefficient) and P/A (peak of autocorrelation), in order to determine, acoustically, the presence of not or this dysphonias like carcinoma, cyts, unilateral paralysis and polyps.

The propose of this work is to define an estimation technique of acoustic parameters (it estimated robustly through statistic methods) applying inverse filtering over all sustained vowel signal through robust adaptive filtering (Wiener and Kalman filters). Later, these measures are used in pathological discrimination and not only in perceptual assessment of normal and pathologic voices.

By the ability of generalization and abstraction, artificial neural networks was used in order to classify the acoustical patterns. Using only the voice signal, it defines an automatic method of pathologies identification.

2. Method

Using adaptive algorithms for residue signal estimation, real conditions like non-stationary signal can be supported. In the next sections are describe the descent

gradient and recursive least square techniques for Wiener filters, and the Kalman filters.

2.1 Descent Gradient

The first technique considers that the prediction error (or residue signal) $e(n)$ may be obtained by difference between an actual sample $s(n)$ and a linear combination of P past samples, that is to say:

$$\begin{aligned} e(n) &= s(n) - \hat{a}_n^T \mathbf{s}_{n-1} \\ \mathbf{s}_n &= [s(n), s(n-1), \dots, s(n-P+1)] \\ \hat{\mathbf{a}}_n &= [a_1, a_2, \dots, a_P] \end{aligned} \quad (1)$$

It can be shown that the adaptative expression of coefficients $\mathbf{a}_n$ of estimate Wiener filter, is given by Honig and Messerschmitt (1984), and Ahmed (1987):

$$\begin{aligned} \hat{\mathbf{a}}_{n+1} &= \hat{\mathbf{a}}_n + \frac{1-\beta}{\sigma_s^2(n)} \cdot e(n) \cdot \mathbf{s}_{n-1} \\ \sigma_s^2(n) &= \beta \cdot \sigma_s^2(n-1) + (1-\beta) \cdot s^2(n) \end{aligned} \quad (2)$$

where β is the convergence rate of the coefficients, defined by

$$\beta = 1 - \frac{1}{N} \quad (3)$$

These definitions for convergence rate and variance estimate of the voice signal, $\sigma_s^2(n)$, guarantee the algorithm stability.

2.2 Recursive Least Square

The adaptive estimate of vocal tract AR model P -coefficients by root mean square minimization technique is derived from eq 1., and are:

$$\begin{aligned} \hat{\mathbf{a}}_n &= [\mathbf{S}_{n-1}^T \mathbf{S}_{n-1}]^{-1} \cdot [\mathbf{S}_{n-1}^T \mathbf{s}_n] \\ \mathbf{S}_n &= \text{col}[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \\ \mathbf{s}_n &= [s(n-1), s(n-2), \dots, s(n-P+1)] \end{aligned} \quad (4)$$

The autocovariance matrix $\mathbf{P}_n$ which derives from $[\mathbf{S}_{n-1}^T \mathbf{S}_{n-1}]$ has been calculated by a recursive equation. Through the matrix inversion lemma, the actualization of this matrix is given by Santana et al (1994):

$$\mathbf{P}_{n+1} = \mathbf{P}_n - \frac{\mathbf{P}_n \mathbf{s}_{n+1} \mathbf{s}_{n+1}^T \mathbf{P}_n}{1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1}} \quad (5)$$

Therefore, the equation that governs the coefficient adaptations is:

$$\hat{\mathbf{a}}_{n+1} = \hat{\mathbf{a}}_n + \kappa_{n+1} [s(n+1) - \mathbf{s}_{n+1}^T \hat{\mathbf{a}}_n] \quad (6)$$

where

$$\begin{aligned}\kappa_{n+1} &= \mathbf{P}_n \mathbf{s}_{n+1} \delta_{n+1}^{-1} \\ \delta_{n+1} &= 1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1}\end{aligned}\quad (7)$$

Through the eq. 6, the excitation signal of speech production model for each interaction can be obtained, and is given by:

$$e_{n+1} = s(n+1) - \mathbf{s}_{n+1} \hat{\mathbf{a}}_n \quad (8)$$

2.3 Kalman Filter

In this approach, a priori, it is necessary to determine the state equations of vocal tract, for posterior application of the Kalman filter algorithm. The space state estimate is built starting from an ARX model, using the least square method. Thus, the system coefficients can be determined, by eq. 9.

$$A(z^{-1})s_i = C(z^{-1})e_j \quad i = 0, \dots, P-1; j = 0, \dots, N-1 \quad (9)$$

Once the past samples are known, the coefficient matrices A and C can be determined, and the estimate error equation is given by:

$$\begin{aligned}\hat{e}(n) &= s(n) - \mathbf{x}_n \hat{\mathbf{a}}_n \\ \mathbf{x}_n &= [-s(n-1), \dots, -s(n-P-1), e(n-1), \dots, e(n-N-1)] \\ \hat{\mathbf{a}}_n &= [a_1, \dots, a_{P-1}, c_1, \dots, c_{N-1}]^T\end{aligned}\quad (10)$$

By using the minimum square criterion in eq. 10, recursively, the ARX model parameters can be obtained (Davis and Vinter, 1985). In each iteration, the matrices A and C are used in the filtering process, by Kalman filter, and these matrices are transformed on a state-space representation (canonical form), producing then, the following system (Santina et al, 1994):

$$\begin{cases} \mathbf{z}_{n+1} = \mathbf{M}_n \mathbf{z}_n + \mathbf{N}_n w(n) \\ s(n) = \mathbf{H}_n \mathbf{z}_n + w(n) \end{cases} \quad (11)$$

where w_n is an unit variance white noise.

From this time-variant system, one determines the best linear estimate for the voice signal. First, a posteriori space-state estimate is determined by the Kalman filter:

$$\begin{aligned}\hat{\mathbf{z}}_{n+1|n} &= \mathbf{M}_n \hat{\mathbf{z}}_{n|n-1} + \kappa'_n [s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1}] \\ \hat{\mathbf{z}}_{0|-1} &= 0 \\ \kappa'_n &= \frac{\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n}{1 + \mathbf{H}_n \mathbf{P}'_n \mathbf{H}_n^T}\end{aligned}\quad (12)$$

The covariance matrix, $\mathbf{P}'_n$, is obtained, by a recursive way, from the Riccati equation, (13):

$$\begin{aligned}\mathbf{P}'_{n+1} &= \mathbf{M}_n \mathbf{P}'_n \mathbf{M}_n^T + \mathbf{N}_n \mathbf{N}_n^T - \kappa'_n [\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n] \\ \mathbf{P}'_0 &= \rho \mathbf{I}\end{aligned}\quad (13)$$

From these iterations, on each new sample, the estimate error is evaluated, which corresponds to the best excitation signal of vocal tract. Thus:

$$e'(n) = s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1} \quad (14)$$

In fact, it obtains the residue sequence that must correspond to best estimation of excitation signal of speech production model. It is used twelve coefficients for AR models of the Wiener filters, based on studies of Davis (1979), although this researcher considered the use of Levinson algorithm. For the Kalman filters, the existence of an exogen portion in ARX model motivates the utilization of 4 coefficient pairs, i.e. {(6,4), (6,8), (8,4), (8,6)}. This set of pairs has been chosen through the Mann-Withney statistical test, using acoustic measurements extracted from residue signals according to Rosa et al (1997). The four selected pairs have presented the best discrimination of pathologies based on Mann-Withney statistical test that assess the pathology discrimination capability using acoustic measures extracted from residue signals (showed in next section), according to Rosa et al (1997).

2.4 Acoustic Measurements

Using the adaptive algorithms for residue signal estimation, a set of acoustic index that represents the temporal and spectral characteristics of this signal can be determined. Based on Davis (1979), seven parameters are define with this objective.

- SNR (signal to noise ratio): Evaluated from the sum of sample squares in all signal residue.
- Jitter (perturbation of the fundamental frequency): median value of difference among successive periods of pitch.
- Peak: Corresponds to median of the maximal absolute difference of residue signal in the pitch period.

In next measurements, it is considered a frame with 4096 samples of the residue signal.

- EX (coefficient excess): Correspond to median value of the magnitude distribution kurtosis of all frames.
- SFR (spectral flatness of residue): Defined as the ratio between geometric and arithmetic means of the frame spectrum. For several frames it is used the median of this value as SFR measurement.

- SFF (spectral flatness of filter): Calculated by difference between the ratio of geometric mean by arithmetic mean of signal voice at correspond time of frame, and the SFR measure.
- P/A (peak of autocorrelation): Defined as the median value of normalized autocorrelation second peak of frame.

The acoustic parameters Jitter and Peak are similar to APQ and PPQ of Davis (1979). To avoid that rapid alterations in perturbation index can affect the pathologic discrimination capability, it uses median instead of mean as central tendency measurement. Together with P/A, Jitter and Peak are useful to assess the roughness level.

The remaining measurements allow to assess the glottal structure considering pathologies that affect the constriction glottal during the sustained speech, and thus, they determine the noise level arising from the incorrect adduction of the vocal folds, which produces breathiness levels.

To avoid the negative effects on acoustic measurements computation, produced by initial samples of residue due to convergence velocity of the adaptive algorithms, it is used two distinct approaches. The first one is to start the calculation 20 ms apart from the first sample of the estimated residue (this value corresponds to a pitch period of extremely deep voice), that guarantee the convergence of Wiener and Kalman coefficients.

The second approach is used to improve the convergence velocity for hoarseness voices (periodicity loss, decrease of energy from the second formant and increase of the high frequency noise). The initial coefficients of inverse filters are estimated using autocorrelation technique (Levinson algorithm) for 1024 first samples of voice signal. This method is not applied to the Kalman filter because it uses the ARX model estimate instead of AR model.

2.5 Artificial Neural Networks

In order to perform an identification of laryngeal and vocal folds pathologies from the presented acoustic measurements, a feedforward artificial neural networks with backpropagation training was used. The procedure for use this structures is:

- Introduction of the i -nth pattern to be trained on input layer neurons.
- Activation and deactivation of neurons.
- Propagation of the activation function results for neurons of next layer.
- Determination of the error between desired and obtained output of neural network.
- Actualization of the synaptic weights according this error.

These steps correspond to manner of this mathematical structure acquires knowledge basing on the input/output patterns that are "taught". Some necessary mathematical elements have to be defined in order to use in pathological discrimination.

The activation function for each neuron used is a sigmoid, i.e.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (15)$$

In the non-linear classification, it is considered the use of one and two hidden layers, testing how many neurons are necessary to keep the error below a pre-established value during testing phase. The acoustic measurements provided to input layer need be scaled because there is a great magnitude difference among these parameters. Thus, the maximum absolute value of each acoustic parameters used in training phase is increased by 2% and normalized to ± 1 , forcing all others parameters estimations to be in the range $[-1, +1]$. This avoids saturation in activation function.

Each neuron of the output layer produces the probability of occurrence of the specific dysphonia for a set of acoustic measurements in the input layer. Therefore it is associated with a distinct pathology. The number of neurons in this layer correspond to the number of analyzed dysphonias.

This technique of association comes from the capability of neural network to address the probability, a posteriori, of which acoustical measurements in the input layer correspond to a specific pathology. In fact, the weights are actualized in order to minimize a relative entropy of trained patterns, or measure of Kullback-Leibler (Haykin, 1994), i.e.

$$H_{P||Q} = \sum_{\alpha} P_{\alpha} \sum_i \left[Q_{i|\alpha} \cdot \ln \left(\frac{Q_{i|\alpha}}{P_{i|\alpha}} \right) + (1 - Q_{i|\alpha}) \cdot \ln \left(\frac{1 - Q_{i|\alpha}}{1 - P_{i|\alpha}} \right) \right] \quad (16)$$

where P_{α} is the probability, a priori, of pathology occurrence for the presented pattern

$P_{i|\alpha}$ is the probability, a posteriori, of the i -nth cell in output layer for the α training pattern;

$Q_{i|\alpha}$ is the real probability of the i -ésima cell in output layer for the α training pattern α ;

The training of this neural network is performed in batch mode, where all patterns to be learned by this structure are presented in a random order, in each epoch. The training rate is controlled manually. It was rejected the moment term because this causes instability of synaptic weights among layers or low convergence velocity.

The stop criteria is based on the mean error by input/output pattern presented in current epoch. Due to similarity among the used patterns, the desired mean error not must be low, because this can affect the generalization performance of the knowledge about the acoustical characteristics of the pathologies. Dysphonic patterns are superimposed and neural network can compass each patterns individually, in geometric terms, instead of agglomerate them in one cluster only.

3. Test Procedures

For analysis of the presented method, 73 voices were recorded from 48 patients and 25 normal speakers. For pathologic case, there are patients that reported more than one dysphonia. The total number of pathologies is 22 and they are discriminated in tab. 1.

All speakers were diagnosed in Otorrinolaryngology sector and Head and Neck Surgery sector in Clinical Hospital of the São Paulo University Faculty of Medicine, at Ribeirão Preto, Brazil, being that the men's age varied from 6 to 80 years, with mean value of 41.5 years, whereas the women's age had an interval from 21 to 71 years, with mean value of 40.9 years. Videolaryngoscopy and videostroboscopy were used in the assessment of vocalization structure both for normal speakers (non-smokers) and dysphonic patients.

For voice record, an unidirectional microphone (LeSon SM58), with dynamic gain, was used. This signal was stored directly on a digital computer through a SoundBlaster data acquisition and reproduction board (model SB16). The sampling rate was 22050 Hz (Hori, 1979), while each sample was quantized in 16 bits. Using digital equipment directly, only one A/D conversion is necessary, reducing to the utmost production of spurious noise that does not correspond to the patient's voice.

The Portuguese phonemes /a/, /e/ and /i/ were recorded for nearly 5 seconds (controlled by operator) in a comfortable level of pitch and magnitude to the speaker. If the patient did not support the phonation period, it was asked for him to stop uttering. The microphone was kept 5 centimeters from the mouth. All recording was preceded by a training procedure of the speaker on the methodology used, in order to avoid psychological artifacts. Some recordings were rejected and again acquired by noting difference between the vowel phonation and normal pattern of speaker conversation. All samples of voice signal were normalized in magnitude between -1 and +1.

4. Results

The adaptive algorithms was applied in all voice signals presented in tab. 1. From the residue signal, it was extracted the seven acoustic parameters that was used in the neural network classification. The output pattern for each dysphonia is presented in tab. 2. The presence of two configuration types for the output layer is motivated by evaluation of neural network capability to identify pathologies with very similar acoustical characteristics like cysts, Reinke's edema, laryngeal cleft, nodules and polyps.

Various configurations of neural networks for pathological discrimination were tested from the acoustic parameters obtained of the residue signal. During the training and test procedures, it was verified that the use of only one hidden layer resulted in 80% error level, that not allow its application in automatic determination of dysphonias. Thus, it became necessary the use of two hidden layers.

Another important factor is the quantity of neurons in these layers. The use of a large number of cells results in fast neural network convergence. Nevertheless, this cause a low interpolation of non-trained patterns because the neural network memorizes only the trained pathologies, without to abstract the acoustical characteristics of these voices. With purpose to avoid this memorization and some local minimum, it was added to each input of neural network a gaussian noise with zero mean and variance of 10^{-4} (the magnitude of noise is decremented with the increasing of the epochs).

To perform the neural network tests, it was used an additional set of voices (also diagnosed by speech specialists). Over this set, it was extracted acoustical measurements to evaluate the capability of the pathology discrimination by the neural network. These voices are presented in table 3.

The configurations that present best results of the dysphonia identification are describe in the next section. In the most of configuration, it was used acoustical measurements obtained by only one adaptive algorithm.

Configuration 1, containing 9 inputs:

- Sex and Phoneme;
- Peak, P/A, Jitter, EX, SFR, SFF and SNR estimated from inverse filtering using LMS algorithm.

Configuration 2, containing 9 inputs:

- Sex and Phoneme;
- Peak, P/A, Jitter, EX, SFR, SFF and SNR estimated from inverse filtering using descent gradient algorithm.

Configuration 3, containing 9 inputs:

- Sex and Phoneme;
- Peak, P/A, Jitter, EX, SFR, SFF and SNR estimated from inverse filtering using Kalman filter with 6/8 coefficients for ARX model.

Configuration 4, containing 9 inputs:

- Sex and Phoneme;
- Peak, P/A, Jitter, EX, SFR, SFF and SNR estimated from inverse filtering using Kalman filter with 6/4 coefficients for ARX model.

Configuration 5, containing 9 inputs:

- Sex and Phoneme;
- Peak, P/A, Jitter, EX, SFR, SFF and SNR estimated from inverse filtering using Kalman filter with 8/4 coefficients for ARX model.

Configuration 6, containing 9 inputs:

- Sex and Phoneme;
- Jitter and SNR estimated from inverse filtering using Kalman filter with 6/8 coefficients for ARX model
- Peak estimated from inverse filtering using Kalman filter with 6/4 coefficients for ARX model
- EX, P/A and SFF estimated from inverse filtering using LMS algorithm
- SFR estimated from inverse filtering using descent gradient algorithm

The information of the sex is presented to neural network motivated by the existence of acoustic differences between men and women, mainly in the fundamental frequency. Beyond this, the voice's women is pretty roughness than voice's men.

In order to the neural network learns the difference between phonemes analyzed, this information (in codified form) is presented to this neural structure, because /e/ and /i/ are very similar perceptually for some patients.

After the training and test of the neural network, it was obtained the results corresponding to accuracy percentage, being showed in tab. 4.

In column "Configuration", the first number represent the acoustical measurements used in input layer and the second one corresponds to the output vector type. This is motivated to verify the capability of neural network to distinguish quite similar acoustical patterns.

It is verified that a lot of configurations presents poor results for pathological discrimination from acoustical parameters. However, configurations, like 5-1, make possible to identify a good number of pathologies using only voice signal.

It can be noted that in some cases (configuration 1-1), the generalization capability is affected by the increase of neurons in hidden layers. Thus, is necessary to considerate the trade-off between the number of positively identified test pattern and the size of neural network (both the number of hidden layers than number of neurons).

5. Conclusion

The use of residue signals obtained from adaptive filters (Wiener and Kalman) do not require the stationary assumption in the voice signal. The robustness of Kalman filter allows that the coefficients of speech production model being adequately estimated, besides to remove the existent correlation in residue signal.

The acoustical measurements computed from this signal make possible to evaluate the main acoustical characteristics in pathologies that affect the larynx, i.e., periodicity on vocal folds movements and amount of turbulent air flow through glottis. Breathiness, roughness, hoarseness and normal voices are identified by magnitude variation of these characteristics. Such perceptual aspects in vocal tract make valid the use of multivariate classifiers in the identification of pathologies.

The results presented in tab. 4 show that the use of artificial neural networks is efficient in the dysphonias discrimination even using short quantity of the acoustical measurements (seven). The robustness of the median (central tendency of distribution) of these acoustic parameters is very important to assess the larynx and vocal folds dysfunctions.

The generalization of the artificial neural networks is helpful to identify pathologies because of the acoustical parameters variability for each anomaly (organic or functional) in the vocal folds. The abstraction capability of the main characteristics of acoustical patterns presented by the neural network justifies the use of this non-linear classifier. It has been shown that is possible to identify pathologies with similar acoustical properties as cysts, nodules and polyps.

The use of many others acoustical measurements has been considered to allow that the neural classifier supplies more information about the pathologies, increasing their discrimination capability. The use of relationships between harmonic and noise components can increase the efficiency of neural network because many dysphonias present patterns well defined of the alterations in these spectral components.

If a more accurate pathology discrimination is required, independently of computer time consumed, it can be used radial base functions to guarantee that the

convergence occurs to the global minimum to obtain the real probabilities of pathology existence from acoustical measurements. The increase of patterns for training phase also improves the accuracy of the pathology identification. The use of non-supervision neural networks and models that considers the superposing of acoustical patterns have been considered because many pathologies are not totally separables, in statistic terms.

The methods presented that make possible the automatic identification of dysphonias directly of voice signal, became a helpful tool for speech specialist in diagnosis. However, it is necessary more studies to improve the classificatory results.

6. References

- Ahmed, N. "Adaptive Filtering", In: Elliot, D. F. (Eds.) (1987), *Handbook of Digital Signal Processing - Engineering Applications*, Academic Press Inc., San Diego.
- Awan, S. N. and Frenkel, M. L. (1992). Improvements in estimating the harmonics-to-noise ratio of the voice. *Journal of Voice*, 8, 255-262.
- Davis, S. B. (1979). Acoustic characteristics of normal and pathological voices. In Lass, N. J. (Ed.) *Speech and Language: Advances in Basic Research and Practice*, vol. 1, New York Academic Press, 271-335.
- Deal, R. E. and Emanuel, F. W. (1978). Some waveform and spectral features of vowel roughness. *Journal of Speech and Hearing Research*, 21, 250-264.
- Deller Jr., J. R., Proakis, J. G. and Hansen, J. H. L. (1993), *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., New York.
- Fukazawa, T., El-Assuooty, A. and Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *Journal of the Acoustical Society of America*, 83, 1189-1193.
- Haykin, S. (1994). *Neural Networks*, MacMillan Publishing Co., Englewoods Cliffs.
- Honig, M. L. and Messerschmitt, D. G. (1994), *Adaptive Filters: Structures, Algorithms, and Applications*, Kluwer Academic, Boston.
- Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. *Journal of Speech and Hearing Research*, 22, 5-19.
- Kasuya, H., Ogawa, S., Mashima, K. and Ebihara, S. (1986). Normalized noise energy as a acoustic measure to evaluate pathologic voice. *Journal of the Acoustical Society of America*, 80, 1329-1334.
- Koike, Y. (1973). Applications of some acoustic measures for the evaluation of laryngeal dysfunction, *Studia Phonologica*, 7, 17-23.
- Krom, G. de (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*, 36, 254-266.
- Lieberman, P. (1961). Perturbation in vocal pitch. *Journal of the Acoustical Society of America*, 33, 597-603.
- Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America*, 35, 344-353.
- Pinto, N. B. and Titze, I. R. (1990). Unification of perturbation measures in speech signals. *Journal of the Acoustical Society of America*, 87, 1278-1289.
- Rosa, M. O., Pereira, J. C., Araújo, S. A. and Grellet, M. Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis, *IEEE Transactions on Biomedical Engineering* (submitted).
- Santina, M. S., A. R. Stubberud and G. H. Hostetter (1994), *Digital Control System Design*, Saunders College Publishing, Orlando, 1994.
- Shoji, K., Regembogen, E., Yu, J. D., Blaugrund, S. M. (1992). H-Index: A new measure of glottal efficiency for the pathologic voice. *Laryngoscope*, 102, 1113-1117.
- Yumoto, E., Gould, W. J. and Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, 71, 1544-1549.

Pathology	N. of cases
Carcinoma	1
Cyst	4
Partial Cordectomy	4
Total Cordectomy	1
Dysarthria	1
Functional Dysphonia	3
Edema	2
Reinke's edema	6
Hyperemia	1
Thickening of Vocal cords	2
Laryngeal Cleft	8
Granuloma	1
Hyperkinetic dysphonia	1
Hypokinetic dysphonia	1
Chronical Laryngitis	2
Contact Ulceration	4
Leukoplakia	2
Nodule	5
Normal	25
Unilateral Paralysis	1
Polyp	6
Others	2
Total	83

Pathology	Output vector in configuration 1	Output vector in configuration 2
Carcinoma	{1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Cysts	{0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}	{0,0,0,0,0,0,0}
Partial Corpectomy	{0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Total Corpectomy	{0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Dysarthria	{0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Functional Dysphonia	{0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Edema	{0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Reinke's Edema	{0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0}	{0,0,0,0,0,1,0}
Hyperemia	{0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Thickening of Vocal Folds	{0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Laryngeal Cleft	{0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0}	{0,0,0,0,1,0,0}
Granuloma	{0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Hyperkinetic dysphonia	{0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Hypokinetic dysphonia	{0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0}	{0,1,0,0,0,0,0}
Chronical Laryngitis	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0}	{0,1,0,0,0,0,0}
Contact Ulceration	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0}	{0,1,0,0,0,0,0}
Leukoplakia	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0}	{0,1,0,0,0,0,0}
Nodule	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0}	{0,0,0,1,0,0,0}
Normal	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0}	{1,0,0,0,0,0,0}
Others	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}	{0,1,0,0,0,0,0}
Unilateral Paralysis	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}	{0,1,0,0,0,0,0}
Polyp	{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}	{0,0,1,0,0,0,0}

Pathology	N. of cases
Cyst	2
Reinke's edema	1
Laryngeal cleft	10
Nodule	2
Normal	23
Polyp	2
Total	40

Table 3 - Table of pathologies found in neural network tests

Configuration	N. of cells by layer	% of total accuracy	% of accuracy for /a/	% of accuracy for /e/	% of accuracy for /i/
1-1	9/22/22/22	63.00%	72.00%	60.00%	54.00%
1-1	9/25/25/22	46.00%	42.00%	57.00%	39.00%
1-2	9/15/15/7	72.00%	66.00%	81.00%	69.00%
1-2	9/20/20/7	56.00%	60.00%	66.00%	42.00%
2-1	9/22/22/22	51.00%	42.00%	42.00%	69.00%
2-1	9/25/25/22	62.00%	45.00%	78.00%	63.00%
2-2	9/15/15/7	51.00%	57.00%	48.00%	48.00%
2-2	9/20/20/7	66.00%	63.00%	66.00%	69.00%
3-1	9/22/22/22	66.00%	75.00%	57.00%	66.00%
3-1	9/25/25/22	60.00%	54.00%	72.00%	54.00%
3-2	9/15/15/7	73.00%	78.00%	78.00%	63.00%
3-2	9/20/20/7	70.00%	72.00%	78.00%	60.00%
4-1	9/22/22/22	48.00%	54.00%	42.00%	48.00%
4-1	9/25/25/22	48.00%	60.00%	54.00%	30.00%
4-2	9/15/15/7	62.00%	63.00%	63.00%	60.00%
4-2	9/20/20/7	61.00%	60.00%	57.00%	66.00%
5-1	9/22/22/22	60.00%	81.00%	51.00%	48.00%
5-1	9/25/25/22	70.00%	63.00%	81.00%	66.00%
5-2	9/15/15/7	45.00%	45.00%	42.00%	48.00%
5-2	9/20/20/7	67.00%	78.00%	66.00%	57.00%
6-1	9/22/22/22	54.90%	47.06%	70.59%	47.06%
6-1	9/25/25/22	61.76%	47.06%	76.47%	61.76%
6-2	9/15/15/7	78.00%	75.00%	90.00%	69.00%
6-2	9/20/20/7	65.00%	63.00%	66.00%	66.00%

Table 4 - Accuracy of pathologic discrimination by neural networks using acoustic measurements

Pathological Discrimination of Larynx through Neural Network

Marcelo de Oliveira Rosa¹
marceloI@lim.sel.eesc.sc.usp.br

José Carlos Pereira¹
pereira@lim.sel.eesc.sc.usp.br

Marcos Grellet²

¹Department of Electrical Engineering
University of São Paulo at São Carlos
Av. Dr. Carlos Botelho, 1465, CP 359
13560-250, São Carlos, São Paulo, Brazil

²Faculty of Medicine
University of São Paulo at Ribeirão Preto
Av. Bandeirantes, 3900
14049-900, Ribeirão Preto, São Paulo, Brazil

ABSTRACT

The pathological diagnosis of the larynx is done by speech pathologists through the acoustical perception of the voice (harshness, breathiness and hoarseness levels), indirect laryngoscopy, videolaryngoscopy, videostroboscopy and images of the larynx by high speed films. These techniques require from the specialist an accurate sense of perception (vision and hearing) in order to identify abnormal movements of the vocal folds. That qualitative evaluation provides no standard procedures in the identification of the dysphonia. The visualization of the muscles and cartilage of the larynx is “masked” by reflex movements in the vocal tract or by sub-glottal localization of the organic growth, causing an inaccurate pathological identification.

The use of non-invasive techniques, by other hand, presents no interference on the measurements and is a useful tool in the pathological assessment. The voice signal analysis combined with the speech production model [1] [2] produce important information about the vocal structures, allowing to distinguish between normal and abnormal function as well as the origin of pathology (organic or functional).

Generally, the acoustical analysis of the pathologies is based on the characteristics of the vocal cord vibration (perturbation of the period and magnitude of the fundamental frequency), and on the level of turbulence of the airflow through the glottis, relating them with perceptual levels of roughness, breathiness and hoarseness.

The perturbation of the period of the fundamental frequency (jitter) and the magnitude perturbation (shimmer) are important parameters in the assessment of the dysphonies [3] [4], because any disturb generated at the glottal level will be propagated through the resonance cavity and will reflect in the voice signal.

The glottal airflow related to the pathologies like polyps, cyst, nodules and fusiforme fissures gives rise to a noisy signal characterized by harmonic components with low energy, alterations on the formant structure and increase of the high frequency noise [4] [5]. Acoustical measurements were proposed in order to quantify these characteristics and to relate them with perceptual levels.

Due to the complex and non-linear nature of the speech production system, this work uses neural network feedforward with backpropagation training to perform the discrimination of 21 dysphonies. The classical acoustic indexes are used as input (perturbation measurements of the vocal cord and spectral evaluation of noise), while the output represents the several pathologies. In fact, each output of this neural network is the probability [6] of that the acoustical measurements of the patient correspond to a specific pathology.

As the parametric model of vocal tract has been consider time-invariant, the voice signal to be analyzed comes from a sustained speech. The acoustical parameters used in the training and testing of the neural network are extracted from this signal.

The characteristics of generalization and robustness of the neural network were tested and it was verified its efficiency in the pathological discrimination. The proposed system allows automatic identification of dysphonies and can be used as powerful tool in the pre-diagnosis of larynx pathologies.

keywords: neural networks application, pathologic diagnosis automatic, digital signal processing, digital speech processing

REFERENCES

- [1] Deller Jr., J. R., J. G. Proakis and J. H. L. Hansen. *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., New York, 1993.
- [2] Rabiner, L. R. and R. W. Schaffer. *Digital Processing of Speech Signals*, Prentice Hall, Englewood Cliffs, 1978.
- [3] Baken, R. J. *Clinical Measurement of Speech and Voice*, Singular Publishing Group, Inc., San Diego, 1996.
- [4] Yanagihara, N. "Significance of harmonic changes and noise components in hoarseness". *Journal of Speech and Hearing Research*, 1967, 10, 534-543.
- [5] Krom, G. de "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments". *Journal of Speech and Hearing Research*, 1995, 38, 794-811.
- [6] Haykin, S. *Neural Networks*, MacMillan Publishing Co., Englewoods Cliffs, 1994.

Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis

¹Rosa, Marcelo de Oliveira; ¹Pereira, José Carlos; ²Grellet, Marcos

¹School of Engineering of São Carlos, University of São Paulo

²Faculty of Medicine of Ribeirão Preto, University of São Paulo

Abstract

The use of non-invasive techniques in the evaluation of the larynx and vocal tract helps the speech pathologist in the accurate diagnosis of pathologies. Also, the diagnostic at the early stage of the pathology can provide substantial benefits in terms of policy and public health. Furthermore, an instrumentation which allows quantitative evaluation of the disease can be useful in order to build up a data base of each patient for later comparison. This work presents three adaptive methods for acoustical measurement estimation that adequately represent the spectral and temporal characteristics of normal speakers and dysphonic patients. These parameters are defined through robust statistical procedures. The capability of pathologic discrimination is evaluated through of non-parametric statistical test, indicating that even dysphonias with perceptual characteristics very close (like nodules, cysts, polyps and laryngeal cleft) are identifiable.

This work is supported by Capes

1. Introduction

Pathology diagnosis of vocal tract is a field that still demands further investigation, since there is no standard between speech professionals in their activities. This is because the current tools available (indirect laryngoscopy, videolaryngoscopy and videostroboscopy, over there own professional ear) require accurate subjective identification of problems in laryngeal and vocal folds, often resulting a qualitative assessment of these structures. Beyond this, some patients may present a reflexive act in the supraglottal cavity that prevents an appropriate examination of glottal structures, causing this procedure even more inaccurate.

Due to their quantitative non-invasive nature, voice acoustic analysis techniques may represent important instrumentation for clinicians in pre-diagnosis of larynx pathologies. Beyond this, organics defective located on the subglottal part of the vocal cords can only be seen through surgical incision.

Studies related to acoustic analysis are based, mainly, on the periodicity of vocal folds vibration and in the amount of air that escape through the glottis during the speech. The periodicity perturbations are made up of measurements of jitter (variation between successive periods of the fundamental frequency) and shimmer (variation between successive magnitude). Researchers have defined acoustical index, in an effort to relate them with pathologies and/or acoustic perception ([1], [2], [3], [4], [5], [6], [7], [8], [9], [10]). Koike [11] and Iwata[12], using correlation coefficients of magnitude measurements of fundamental frequency, defined a voice pattern for normal speakers and patients with unilateral paralysis and carcinoma. Pinto and Titze [12] searched to unificate the measurements of jitter and shimmer through the analysis of high-order difference, using the median and standard deviation. These statistical measurements have proved satisfactory in pathological classification of voice signals. On the other hand, Wendhal [14], [15] synthesized the jitter and shimmer, in order to represent a harshness voice for comparison with dysphonic voice.

The turbulence in glottal flow resulting from malfunction of vocal cords can be quantified by the noise in spectral components of speech. Pathologies like, for example, polyps, cysts, Reinke's edema, generally reduce the energy of the harmonic components of the spectrum and increase it at non-harmonic ones. That distortion on the spectrum is related to the pathology extent, and several spectral parameters have been proposed in order to assess this unbalance. The relationship among harmonic components and the remaining components ([16], [17], [18], [19], [20], [21], [22]), the difference between low frequency harmonics [23], the amount of spectral energy starting from specific frequency [24] and the analysis of least values of spectral intensity in frequency intervals ([25], [26], [27], [28]), are some of the spectral parameters proposed in the literature.

In another approach, some authors have analyzed the residues of inverse filtering of the speech signal ([29], [30], [31], [32]), which correspond to an estimate of the excitation signal in the mathematical model of the vocal tract. Some parameters of signal residues are extracted for assessment of normal speakers and dysphonic patients.

The present work follows the latter approach, incorporating robust adaptive techniques, which allow a better estimate of the residue signal over all samples (approximately 5 seconds of sustained speech), even in non-stationarity conditions, that occurs for involuntary motion of the supraglottal apparatus.

The combined use of Wiener's and Kalman's filters allows an optimum discrimination of pathologies with similar acoustic characteristics. The arrangement of these algorithms in the determination of acoustic measurements over the 3 Portuguese testing phonemes (/a/, /e/ and /i/), results in discrimination of 21 different pathologies.

2. Samples Acquisition method

The analyzed voices originated from 73 speakers, 48 being dysphonic and 25 persons presenting no pathology. In the first case, there are patients that reported more than one pathology. The samples of normal voices constitute the control group. The total number of evaluated pathologies is 21 and they are discriminated in tab. 1.

Pathology	N. of cases
Carcinoma	1
Cyst	4
Partial Cordectomy	4
Total Cordectomy	1
Dysarthria	1
Functional Dysphonia	3
Edema	2
Reinke's edema	6
Hyperemia	1
Thickening of Vocal cords	2
Laryngeal Cleft	8
Granuloma	1
Hyperkinetic dysphonia	1
Hypokinetic dysphonia	1
Chronical Laryngitis	2
Contact Ulceration	4
Leukoplakia	2
Nodule	5
Normal	25
Unilateral Paralysis	1
Polyp	6
Others	2
Total	83

Table 1 - Table of pathologies found in the analyzed sample

All patients are diagnosed by a medical specialist of Otorrinolaryngology sector and Head and Neck Surgery sector in Clinical Hospital of the Faculty of Medicine of the University of São Paulo at Ribeirão Preto, being that the men's age varied from 6 to 80 years, with mean value in 41.5 years, whereas the women's age had in interval from 21 to 71 years, with mean value in 40.9 years. The diagnosis involved the use of videolaryngoscopy, and, in some cases, was also made use of videostroboscopy for final consideration about the dysphonia. The control group (non smokers) was submitted to the same clinical procedure for confirmation of that has deprived of organic or functional dysphonia.

For recording of the patient voices, an unidirectional microphone (LeSon - SM58), with dynamic gain control was used. The data were stored directly on a digital computer through a SoundBlaster data acquisition and reproduction board model SB16. The sampling rate was 22050 Hz, according with the consideration of temporal resolution of Horii [5], while the samples quantization level was 16 bits, with automatic gain control, in order to avoid saturation of the A/D converter. The use of this equipment, instead standard magnetic recording media, is due to phase distortion in low frequency [34]. The use of professional equipment was not considered because of its high cost and difficult use in a medical clinic. Using directly the digital equipment, only one A/D conversion is necessary, reducing to the utmost production of spurious noise that does not correspond to the patient's voice.

Each speaker is asked to pronounce the Portuguese phonemes /a/, /e/ and /i/ for nearly 5 seconds (controlled by operator) in comfortably levels of pitch and magnitude corresponding to the day to day speech. If the patient did not support the phonation period, it was asked for he stops uttering. That is necessary because increasing air flow in the glottis, the maintenance of phonation is caused by higher tension of vocal folds, producing an inherent increase of the fundamental frequency, and consequently, an artificial stability in speech production [35]. The microphone was maintained 5 centimeters from the mouth. All recording was preceded by a training procedure of the speaker on the methodology used, in order to avoid psychological artifacts. Some recordings were rejected and again acquired by noting difference between vowel phonation and normal pattern of speaker conversation.

3. Residue Signal Estimate and Acoustic Measurements

The inverse filtering found in literature is developed over the stable samples of digitized speech signal, defined to be those occurring 400 ms after phonation onset, on the average, in a 140 ms time frame, according to Prosek [33]. The coefficients of AR filter (14-order) are estimated using the Levinson algorithm [36] due to its efficiency, without the use of high frequency pre-emphasis because, according to Davis [30], there is no improvement in the performance of pathological discrimination. The residue signal was obtained by applying the inverse filter over speech signal, and it corresponds to the excitation signal of the vocal tract model [37].

The evaluation of this method shows that the critical points are in stationary demand of signal by mathematical definition of autoregressive filter estimate [36], requiring, therefore, a time-invariant vocal tract during a few milliseconds. In this way, the tongue movement, for example, even if during sustained speech, may reflect in residue signals like any dysphonia occurring at the glottal level.

Another important aspect is the definition of the starting point for estimation, that is to say, determination of a more stable frame of voice signal. The characteristics of initial air flow combined with the laryngeal constrict differ in a population (normal and pathologic), demanding that the beginning of the estimate to be done according to the speaker. To overcome these difficulties, adaptive algorithms to AR model estimate and residue signal determination were used, so that small modifications in the vocal tract profile will reflect only on the coefficients. The following algorithms were constructed:

- a) estimate by Wiener filter using the descent gradient technique
- b) estimate by Wiener filter using recursive least square
- c) estimate by Kalman filter.

Once the residue signal was defined, it is necessary to determine measurements that represent this signal appropriately to make the pathologic discrimination or even the assessment of the patient glottal structure. In other words, the acoustical parameters corresponding to measurements that handle aspects as glottal signal variation (jitter and shimmer) and the amount of air flow turbulence in the vocal folds are determined.

3.1. Descent Gradient

The first technique considers that the prediction error (or residue signal) $e(n)$ may be obtained by difference between an actual sample $s(n)$ and a linear combination of P past samples, that is to say:

$$e(n) = s(n) - \hat{\mathbf{a}}_n^T \mathbf{s}_{n-1}$$

$$\mathbf{s}_n = [s(n), s(n-1), \dots, s(n-P+1)] \quad (1)$$

$$\hat{\mathbf{a}}_n = [a_1, a_2, \dots, a_P]$$

It can be shown that the adaptation expression of coefficients $\mathbf{a}_n$ of estimate Wiener filter, is given by ([38], [39]):

$$\hat{\mathbf{a}}_{n+1} = \hat{\mathbf{a}}_n + \frac{1-\beta}{\sigma_s^2(n)} \cdot e(n) \cdot \mathbf{s}_{n-1} \quad (2)$$

$$\sigma_s^2(n) = \beta \cdot \sigma_s^2(n-1) + (1-\beta) \cdot s^2(n)$$

where β is the convergence rate of the coefficients and can be defined taking into account the number of analyzed samples, N , that is:

$$\beta = 1 - \frac{1}{N} \quad (3)$$

These definitions for convergence rate and variance estimate of the voice signal, $\sigma_s^2(n)$, guarantee the algorithm stability.

3.2 RLS

The adaptive estimate of vocal tract AR model P -coefficients by root mean square minimization technique was determined using the same procedure as above. The estimate of these coefficients is given by eq. 4.

$$\hat{\mathbf{a}}_n = [\mathbf{S}_{n-1}^T \mathbf{S}_{n-1}]^{-1} \cdot [\mathbf{S}_{n-1}^T \mathbf{S}_n]$$

$$\mathbf{S}_n = \text{col}[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \quad (4)$$

$$\mathbf{s}_n = [s(n-1), s(n-2), \dots, s(n-P+1)]$$

The autocovariance matrix $\mathbf{P}_n$ which derives from $[\mathbf{S}_{n-1}^T \cdot \mathbf{S}_{n-1}]$ has been calculated by a recursive equation. Through the matrix inversion lemma, the actualization of this matrix is given by ([40]):

$$\mathbf{P}_{n+1} = \mathbf{P}_n - \frac{\mathbf{P}_n \mathbf{s}_{n+1} \mathbf{s}_{n+1}^T \mathbf{P}_n}{1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1}} \quad (5)$$

Therefore, the equation that governs the coefficient adaptations is:

$$\hat{\mathbf{a}}_{n+1} = \hat{\mathbf{a}}_n + \kappa_{n+1} [s(n+1) - \mathbf{s}_{n+1} \hat{\mathbf{a}}_n] \quad (6)$$

where

$$\begin{aligned} \kappa_{n+1} &= \mathbf{P}_n \mathbf{s}_{n+1} \delta_{n+1}^{-1} \\ \delta_{n+1} &= 1 + \mathbf{s}_{n+1}^T \mathbf{P}_n \mathbf{s}_{n+1} \end{aligned} \quad (7)$$

Through the eq. 6, the excitation signal of speech production model for each interaction can be obtained, and is given by:

$$e_{n+1} = s(n+1) - \mathbf{s}_{n+1} \hat{\mathbf{a}}_n \quad (8)$$

3.3 Kalman Filter

The last algorithm used is through the estimate of a Kalman filter for the voice signal and the evaluation of its prediction error for each sample. In this approach, a priori, is necessary to determine the state equations of vocal tract, for posterior application of the Kalman filter algorithm. The space state estimate was built starting from an ARX model, using the least square method. Thus, the system coefficients can be determined, by eq. 9.

$$A(z^{-1})s_i = C(z^{-1})e_j \quad i = 0, 1, \dots, P-1; j = 0, 1, \dots, N-1 \quad (9)$$

Once the past samples are known, the coefficient matrices A and C can be determined, and the estimate error equation is given by:

$$\begin{aligned} \hat{e}(n) &= s(n) - \mathbf{x}_n \hat{\mathbf{a}}_n \\ \mathbf{x}_n &= [-s(n-1), \dots, -s(n-P+1), e(n-1), \dots, e(n-N+1)] \\ \hat{\mathbf{a}}_n &= [a_1, \dots, a_{P-1}, c_1, \dots, c_{N-1}]^T \end{aligned} \quad (10)$$

By using the minimum square criterion in eq. 10 (through a recursive process), the ARX model parameters can be obtained ([40], [41]). In each iteration, the matrices A and C are used in the filtering process, using the Kalman filter. Firstly, a transformation of the difference equation (eq. 9) is performed, with the parameters already evaluated, for a state-space representation (canonical form), producing then, the following system [40]:

$$\begin{cases} \mathbf{z}_{n+1} = \mathbf{M}_n \mathbf{z}_n + \mathbf{N}_n w(n) \\ s(n) = \mathbf{H}_n \mathbf{z}_n + w(n) \end{cases} \quad (11)$$

where w_n is an unit variance white noise.

From this time-variant system and based on its stochastic nature, represented by eq. 11, it follows that a best linear estimate for the voice signal can be found. First, a posteriori space-state estimate is determined by the Kalman filter, through the eqs. (12):

$$\begin{aligned} \hat{\mathbf{z}}_{n+1|n} &= \mathbf{M}_n \hat{\mathbf{z}}_{n|n-1} + \kappa'_n [s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1}] \\ \hat{\mathbf{z}}_{0|-1} &= 0 \end{aligned} \quad (12)$$

$$\kappa'_n = \frac{\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n}{1 + \mathbf{H}_n \mathbf{P}'_n \mathbf{H}_n^T}$$

The covariance matrix, $\mathbf{P}'_n$, is obtained, by a recursive way, through the Riccati equation, (13):

$$\begin{aligned} \mathbf{P}'_{n+1} &= \mathbf{M}_n \mathbf{P}'_n \mathbf{M}_n^T + \mathbf{N}_n \mathbf{N}_n^T - \kappa'_n [\mathbf{M}_n \mathbf{P}'_n \mathbf{H}_n^T + \mathbf{N}_n] \\ \mathbf{P}'_0 &= \rho \mathbf{I} \end{aligned} \quad (13)$$

From these iterations, on each new sample, the estimate error is evaluated, which corresponds to the best excitation signal of vocal tract. Thus:

$$e'(n) = s(n) - \mathbf{H}_n \hat{\mathbf{z}}_{n|n-1} \quad (14)$$

Once have been established these algorithms, it one can be determine how many coefficients must be used in the model estimation. For descent gradient and recursive least mean square, 12 coefficients were considered for the vocal tract model. The Kalman filter demands the use of 7 coefficients combination for autoregressive and exogen parts of

ARX model, the following variations being used: $\{(1, 1), (3, 3), (6, 4), (6, 8), (8, 4), (8, 6), (12, 6)\}$, in order to verify its efficiency.

3.4 Acoustics Measurements

Acoustical parameters have been calculated over all voiced portion of the data, resulting, on the average, 120K samples normalized in magnitude between -1 and +1.

Using the algorithms described in the previous section on residue signal estimate, a set of parameters, representing the spectral and temporal characteristics of this signal, can be determined. Based on Davis [30], seven parameters were defined for this purpose.

- SNR (signal-noise ratio): Defined as the sum of square of the residue signal.
- EX (excess coefficient): Defined as a measurement of the kurtosis of magnitude distribution. Speakers without dysphonia have leptokurtic (sharp) distribution, while the distribution is platikurtic (wide) for breathy voice patterns.
- SFR (spectral flatness of residue): is a measure of the masking of harmonic magnitudes by noise. Defined as the ratio between geometric and arithmetic means of the signal residue spectrum.
- SFF (spectral flatness of AR filter): is a measure of the masking of formant components and bandwidths by noise, and can be obtained by difference between the ratio of geometric mean by arithmetic mean of signal voice, and the SFR measure.
- Jitter (perturbation of the fundamental frequency): median value of first order perturbation, i.e., difference between successive periods of pitch. The estimate of these periods is done in a synchronous way upon the residue signal, by searching the major absolute differences among magnitudes in this signal. An autocorrelation is performed in order to determine the region which contains the maximal absolute magnitude difference for each period, and to maintain the synchronism.
- Peak: corresponds to the median of maximal absolute differences of residue signal magnitudes.
- P/A: (peak of autocorrelation): median value of second normalized autocorrelation peak, for evaluation of residue signal periodicity.

Except the SNR, all other parameters are evaluated over a 4096 samples window. The Jitter and Peak measurements are similar to APQ and PPQ parameters, as defined by Davis [30] and Koike and Markel [32]. Instead of mean value, the median was used in the evaluation of the central tendency measurement, since that the median is more robust and insensitive to fast disturbances in perturbation measurements [13]. Jitter and Peak, together with P/A, are useful to assess the voice roughness level.

The remaining measurements allow evaluation of the glottal structure considering the pathologies that affect the glottal constriction during sustained speech, and thus, determine the noise quantity arising from incorrect adduction of vocal cords, which produces inadequate levels of breathiness.

As adaptive methods require an initial quantity of samples in order to converge, two approaches were used to inhibit the effect over the acoustical measurement estimates. The first one is to start the calculation 20 ms apart from the first sample of the estimated residue. This value corresponds to a pitch period of a extremely deep voice (50 Hz), being enough for guarantee the convergence of filter coefficients.

For hoarse voices (periodicity loss, decrease of energy from the 2 formant and increase of the high frequency noise), yet another approach must be used, since that the convergence time may be unsatisfactory and spurious values may deteriorate residue measurements. Thus, the initial values of inverse filter coefficients are determined through the autocorrelation method (Levinson Algorithm) for a set of 1024 first samples of voice signal, rerun through Wiener algorithm, later. That approach was not applied to the Kalman filter because it uses the ARX model estimate instead of AR model.

In order to determine the separability existence among the voice types and the influence of the adaptive methods in this discrimination, the Mann-Withney [42] test was applied over parameters. The use of this test, instead t-test, is due to the restrictions of the samples set by pathology, the lack of knowledge of distribution that governs the acoustical measurements and the impossibility of to define a numerical relationship among pathologies.

For a specific acoustical parameter, the level of separability among groups was verified by an adaptation of the statistical method. The difference between the critical values of non-parametric test and the value to be tested was considered to be the discrimination indicator.

4. Results

Since the number of data samples is small, the statistical test was performed with a low significance level $[\alpha=0.05]$. In section 4.1, the voice types distinction is made based on acoustical parameters regardless the algorithms used. Next, in

the section 4.2, the computational methods are evaluated, effectively. Finally, section 4.3, presents the discriminated pathologies, regarding the acoustical measurements and the adaptive techniques.

4.1 Acoustical Measurements

In order to evaluate the efficiency of the acoustical parameters, all algorithms have been applied upon the 231 combinations of pathological types. Table 2 shows the measurements with the best and worst performance in the discrimination of pathologies, when the /a/ phoneme is used.

The number of discrimination is based on the quantity of rejection of Mann-Withney test. For example, if a normal-carcinoma combination is rejected by the test, a unit is increased in the score of the respective acoustical parameter. The average of this score for all algorithms is shown together to the acoustical parameter.

The level of separability is calculated on the combinations discriminated above and consists in the distance between the critical and established values of the statistical test. The average of separability is the ratio between the level of separability and the number of discrimination.

	Best discrimination	Worst discrimination
N. of discrimination	Jitter (120)	SFR (84.111)
	SFF (107.222)	SNR (93.444)
Level of discrimination	Jitter (92.580)	Peak (8.985)
	EX (56.491)	SFR (14.894)
Average of separability	Jitter (0.771)	Peak (0.056)
	PA (0.5447)	SFR (0.058)

Table 2 - Discrimination of voice types for phoneme /a/

From table 2 it can be seen that the Jitter parameter presents the best result in assessment of voice condition, whereas SFR and Peak are not indicated for this discrimination. Probably the existence of pathologies with strong roughness characteristics (high perturbation of the fundamental frequency), like hyperkinetic dysphonia, chronical laryngitis and leukoplakia, is influencing the results.

Once SFF is a masking measurement of formant structure by noise caused by excessive quantity of glottal airflow, its ability of discrimination is increased for extreme hoarse dysphonia, like cordectomy (partial and total), carcinoma and unilateral paralysis. The same relation between hoarseness pathologies and EX measurement confirms this consideration.

The SFR measurement has low ability of discrimination, beside Peak parameter. Apparently it may be thought that the antagonism between a perturbation measurement provides better ability of pathologies discrimination and another measurements, of same nature, as Peak or SFR do not to possess the same ability. The difference between Jitter and SFR is not very large. It can be seen from table 2 that the Jitter (best parameter) discriminates only 42% more than SFR (worst parameter).

In relation to SFR, it is possible that the spectral flatness of residue signal in some voice types be quite similar, allowing that only extreme dysphonias may be identified by this parameter. The same consideration is valid for Peak measurement, where the separability level is low due to similarity among the residue signals.

Applying the statistical method on the acoustic parameters in phoneme /e/, table 3 is obtained.

	Best discrimination	Worst discrimination
N. of discrimination	Peak (126.556)	SFR (98.222)
	SFF (121.333)	EX (103.889)
Level of discrimination	Jitter (87.024)	SFR (20.146)
	P/A (78.072)	Peak (27.919)
Average of separability	Jitter (0.729)	SFR (0.198)
	P/A (0.710)	Peak (0.220)

Table 3 - Discrimination of voice types for phoneme /e/

Again, the Jitter measurement presents good results (in spite of ranking as third in the number of discrimination), maintaining the same discriminatory performance as phoneme /a/. The same can be concluded for the P/A parameter; confirming that perturbation measurements, of that kind, are important in the pathological assessment.

Although the Peak measurement for the phonemes /a/ and /e/ presents good results regarding the number of discrimination, the other statistical measurements do not allow its use as a reliable parameter.

The acoustical measurement SFR, that maintain low power of discrimination, seems to demonstrate that the residue signals of various voice patterns present quite similar spectral characteristics, making their statistical separation difficult. Through comparison of tables 2 and 3, it can be seen that the phoneme /e/ presents better results than phoneme /a/ in separability test of the voice types.

Finally, phoneme /i/ is analyzed. Table 4 shows, once more, that the acoustical parameter Jitter is a good measurement to distinguish dysphonias. The parameter SFF also appears with reasonable separability levels for all phonemes.

	Best discrimination	Worst discrimination
N. of discrimination	Jitter (123.111)	P/A (89.111)
	SFF (111.333)	EX (102)
Level of discrimination	P/A (68.259)	Pico (21.843)
	Jitter (43.639)	SFR (27.888)
Average of separability	P/A (0.402)	Pico (0.202)
	EX (0.755)	SFR (0.234)

Table 4 - Discrimination of voice types for phoneme /i/

It is interesting to note that, despite its low number of discrimination, the other statistical parameters of P/A allow its use as a good pathological discriminator. The hyperfunctionality of the glottal structures during the uttering of the phoneme /i/ produces a higher tension on the vocal cords, causing quite similar perturbation of the acoustical characteristics, even having discrimination among the considered pairs. That is corroborated by the presence of the P/A index in the worst discrimination column, since the signal periodicity is calculated through the autocorrelation of the residue. It's verified that the potentially this phoneme in discrimination of the voice types is very similar to that obtained by use of phoneme /a/.

4.2 Adaptive algorithms

The effects produced by the use of different algorithms in residue signal estimate is considered here. Table 5 shows the algorithms with better results in discrimination of voice types, regarding the acoustical measurements and phonemes. The two values following Kalman filters correspond to the number of coefficients in autoregressive and exogen parts, respectively. For each combination phoneme x acoustical measurements, two algorithms are presented with better results in pathologic separability, when number and level of discrimination are jointly considered.

	/a/	/e/	/i/
EX	RLS	RLS	RLS
	Gradient	Kalman 6/8	Gradient
P/A	RLS	RLS	Gradient
	Gradient	Gradient	Kalman 6/8
Jitter	RLS	Kalman 8/4	Kalman 1/1
	Kalman 8/6	Kalman 6/4	Kalman 6/8
Peak	Kalman 8/4	Kalman 6/8	Kalman 6/8
	Kalman 6/4	Kalman 6/4	Kalman 8/6
SFF	Kalman 8/4	RLS	Kalman 6/4
	RLS	Kalman 6/8	RLS
SFR	RLS	Gradient	RLS
	Gradient	Kalman 1/1	Gradient
SNR	Kalman 1/1	Kalman 6/8	RLS
	Kalman 8/4	RLS	Kalman 6/4

Table 5 - Algorithms to discrimination for phoneme and acoustic measure

It can be noticed that the Wiener filters (by RLS and descent gradient) presents better discrimination when the acoustical measurements are determined on the residue signal of the phoneme /a/, evidenced by parameters EX, P/A and SFR. Related to the phoneme /e/, Kalman 6/8 filter has the best performance for the ARX model estimate, while that for the phoneme /i/, the algorithms that demonstrates better separability among voice types is RLS Wiener filter.

One important consideration is related to the perturbation measurements, Jitter and Peak, where the technique of residue signal estimate by Kalman filter presents better performance than Wiener filter: the determination of these acoustical measurements is done through the synchronous search of the pitch periods, based on periodical peaks of residue signal. The more prominent of the peaks in the signal, the better is the pitch identification, and consequently, more accurate the perturbation estimate. That indicates the existence of correlation in the residue signal itself, probably due to the characteristics of each dysphonia. Therefore, the techniques that consider this correlation (Kalman filter, in this case) tend to reduce this information in the residue. The information about pathological characteristics embedded in the residue signals may explain the fact that Wiener filters are more efficient than Kalman filters for pathology classification.

Table 5 shows Kalman filters as the best discriminators of voice types for phonemes /e/ (mainly) and /i/, despite poor concentration upon a unique pair of coefficients for ARX model. This fact indicates that an acoustical assessment of vocal tract demands the use of different techniques to evidence a given dysphonia.

4.3 Discrimination of voice types

This section presents the discrimination results of all pathological pairs combination possible using the acoustic parameters or the adaptive algorithms. Speech arising from speakers with carcinoma, granuloma, chronic laryngitis, dysarthria, partial or total cordectomy, contact ulceration, leukoplakia and unilateral paralysis, characterized as extremely hoarse, produce high levels of separability, and a large number of pairs discrimination, in all phonemes, algorithms and measurements. That is because these pathologies affect as much the fundamental frequency (harmonic structure of speech) as the formant structure, in a accentuated way (reduction of energy starting from the second formant and an increase of high frequency noise). Pathologies like edema, thick of vocal cords and hypokinetic dysphonia (with breathy characteristics), as well as hyperemia and hyperkinetic dysphonia (with roughness characteristics), follow the same tendency in separability.

In order to verify the potentiality of the procedure, the parameters performance was evaluated in patients with hard perceptual diagnosis. Also, the normal pattern voice takes part in the non-parametric test, and the remaining dysphonias are gathered together in the "others" group. Table 6 presents the number of total cases that the acoustical measurement is able to discriminate a pathologic pair.

	EX	P/A	Jitter	Peak	SFF	SFR	SNR	Total
Cyst x Reinke's E.		1	1			1		3
Cyst x Laryngeal cleft	1	1				1		3
Cyst x Nodule	1			7		1	1	10
Cyst x Polyp	1		9					10
Cyst x Others						2		2
Cyst x Normal	3	13	15			1		32
Reinke's E. x Laryngeal cleft		2	2		4	1		9
Reinke's E. x Nodule	2	1						3
Reinke's E. x Polyp		1	1			1		3
Reinke's E. x Others		2	1	1		2		6
Reinke's E. x Normal	4	19	27		1	10		61
Laryngeal cleft x Nodule	9		1			1		11
Laryngeal cleft x Polyp	4	1	9		2	3	1	20
Laryngeal cleft x Others			2	2	10	5	3	22
Laryngeal cleft x Normal	10	20	21			8		59
Nodule x Polyp	5		8	2				15
Nodule x Others		1		13	1		2	17
Nodule x Normal	13	5	19	3		6		46
Polyp x Others			7					7
Polyp x Normal	12	17	18	1	6	5	2	61
Others x Normal	17	27	27	1	25	8	18	123
Total	82	111	168	30	49	56	27	523

Table 6 - Number of events that the parameter did distinguish the dysphonic pair

It is verified that the Jitter, followed by P/A, presents better discrimination of pathological pairs. That is a direct consequence of the use of Kalman filters, which have proved to be the best estimator from this parameter. The excess coefficient, related to the residue sample amplitude distribution, has a large influence in dysphonia discrimination, concentrating, mainly, on the separation between laryngeal cleft and vocal nodules. The less efficient measurement is SNR, followed by the Peak parameter, probably because the residue signals present similar effects as these parameters. It's important to say that these pathologies are acoustically very similar, and that does not invalidate the acoustical parameters, because they are quite important in multivariate classification.

The ability of pathological pairs discrimination regarding algorithms is shown in tab. 7. The Kalman filters are identified in columns by K letter, followed by ARX model coefficients.

	Grad.	RLS	K1/1	K3/3	K6/4	K6/8	K8/4	K8/6	K12/6	Total
Cyst x Reinke's E.	1	1	1							3
Cyst x Laryngeal cleft	1					1			1	3
Cyst x Nodule	2	1	1	1	1			1	3	10
Cyst x Polyp		1	1	2	2	1	1	1	1	10
Cyst x Others					1				1	2
Cyst x Normal	5	5	3	2	5	3	2	3	4	32
Reinke's E. x Laryngeal cleft	1	1			1	1	2	1	2	9
Reinke's E. x Nodule	1	1		1						3
Reinke's E. x Polyp		1		2						3
Reinke's E. x Others		1	1		2		1		1	6
Reinke's E. x Normal	7	10	7	7	5	7	7	6	5	61
Laryngeal cleft x Nodule	1	2	1	1	1	1	1	1	2	11
Laryngeal cleft x Polyp	4	2	2	1	1	2	4	2	2	20
Laryngeal cleft x Others	2	1	2	4	2	5	1	4	1	22
Laryngeal cleft x Normal	10	9	6	8	5	7	5	5	4	59
Nodule x Polyp	1	1	3	2	1	2	1	3	1	15
Nodule x Others	2	2	1	2	1	3	2	3	1	17
Nodule x Normal	5	9	6	3	3	7	3	4	6	46
Polyp x Others	1		1	1	1	1		1	1	7
Polyp x Normal	6	7	8	6	6	8	8	6	6	61
Others x Normal	17	15	16	11	13	13	12	14	12	123
Total	67	70	60	54	51	65	50	55	54	523

Table 7 - Number of events that the algorithm did distinguish the dysphonic pair

Table 7 shows that the RLS algorithm presents the best results, indicating that this estimator keeps important information of the residue signal. Also, Kalman filters (mainly 6/8), although inferior, present similar results to Wiener filters. Probably, the lack of correlation accomplished on the residue signal transfers the pathological information to the coefficients of exogenous part in ARX model estimate.

As far as the pathologies are concern, the Reinke's edema presents low number of discrimination, jointly with the pair laryngeal cleft and cyst. As these pathologies are characterized as breathy, it's necessary to use a large number of parameters, in way to identify more accurately these dysphonias.

The laryngeal cleft is a pathology in that, normally, the vocal cords are in perfect condition, with air escaping by an anterior or posterior aperture in the glottis. That dysphonia presents good distinction among the combined pairs, once the remaining present some modification in vocal cords. In conclusion, even pathologies with quite similar acoustical characteristics are susceptible to be discriminated.

The adaptive method used shows good results when one considers its ability to determine if a speaker presents or not an illness in the larynx, since that for all acoustical measurements, phonemes and algorithms considered it was possible to distinguish the normal pattern from the pathological ones.

5. Conclusion

The results obtained demonstrate that the acoustical measurements represent a feasible technique in speech pathology discrimination. Similar dysphonias like cysts and nodules present distinct characteristics when this technique is used.

These methods allow robust acoustical measurements to be performed by using a large number of voice signal samples in residue signal estimation, along with robust statistical measurements, like the median instead of the average. This guarantees that spurious noise in parameters do not influence negatively the pathology estimate.

Another point of concern is the convergence rate of these algorithms and which is the starting point of the signal to initiate the determination of the acoustical measurements. Through the statistical characteristics and initial estimate of optimum filters coefficients by autocorrelation (Levinson) or autocovariance (Cholesky) techniques, a decrease of 15 ms in the starting point of acoustical parameters calculation has been obtained.

Regarding the acoustical measurements, Jitter and P/A, obtained from residue signal, present good results in pathological discrimination, whereas Peak and SFR are parameters with low capability to separate voice types pairs. The influence of fundamental frequency perturbation has strong importance in pathologic identification despite the fact that measurements of noise level like SFF, SNR and EX present equivalent performance in this task. Thus, it is possible to obtain an efficient classifier with the joint use of these measurements.

The phoneme /e/ allows a better discrimination of pathologies by measurements than phonemes /a/ and /i/, probably because, for this phoneme, the vocal cords are in the intermediate tension condition. On the other hand, the measurements EX and P/A present better results when applied on phoneme /a/, Jitter and SFR when are used over /i/, and Peak, SFF and SNR over phoneme /e/. Since the tension requirements on vocal cords are greater for the phoneme /i/, its adduction is too large, decreasing the air turbulence, and consequently, the noise embedded in the residue signal.

Due to a more relaxed condition of vocal cords, the phoneme /a/ allows a better evaluation of the noise quantity in the residue signal and the assessment of breathy pathologies, through the EX and P/A measurements.

Related to the estimation algorithms of residue signal, it may be concluded that for phonemes /e/ and /i/, the use of Kalman filter presents better results, despite it having no predominant coefficients pair. When the phoneme /a/ is chosen, better results are achieved by the use of Wiener filter, mainly RLS.

The use of algorithms is also influenced by acoustical parameters. In measurements of magnitude and fundamental frequency perturbation (Peak and Jitter, respectively), the Kalman filters present better results. Because of mathematical considerations on which the method is based (noise in estimate system, correlation of model excitation signal, robust technique), part of residue signal noise is removed, and makes easy the synchronous determination of the fundamental frequency period. Intrinsically this method minimizes the correlation between the sample of residue signal. That minimization makes the Kalman filter present worse results in discrimination than Wiener filter regarding spectral measurements as EX, P/A and SFR. As the applied method produces similar residue signal for all dysphonias, the Kalman filter is less sensitive to discrimination.

Probably, the information removed from residue signal by Kalman filter, that allows one to distinguish between voice patterns, is stored in exogenous coefficients of the estimate ARX model. The analysis of these coefficients may permit the construction of new parameters that allow the pathological discrimination using only the Kalman filter.

Comparing the separability performance of the algorithms, the Wiener filter estimate by RLS was the best discriminator of the voice types pairs. The use of Kalman filter demands several combinations of ARX model coefficients, and consequently, a large computational effort in the voice signal evaluation. The faster algorithm was descent gradient, 5 times faster than RLS. One notices in the tables shown that this technique presents goods results, being useful for low capacity equipments, since the memory extent and processing capability demanded are reduced.

In conclusion the technique presented allows the pathological discrimination, even for those with very similar characteristics as polyps, nodules and laryngeal clefts. Classifiers can be constructed based on presented measurements. The efficiency of these classifiers, however, must be improved with the use of other acoustical parameters, extracted from residue signal or voice signals themselves (time and frequency domain).

6. Bibliography

- [1] P. Lieberman, "Perturbation in vocal pitch", *Journal of the Acoustical Society of America*, 33, 597-603, 1961.
- [2] P. Lieberman, "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges", *Journal of the Acoustical Society of America*, 35, 344-353, 1963.
- [3] M. H. L. Hecker and E. J. Kreul, "Description of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency", *Journal of the Acoustical Society of America*, 49, n. 4, part 2, 1275-1282, 1970.
- [4] R. E. Deal and F. W. Emanuel, F. W. "Some waveform and spectral features of vowel roughness", *Journal of Speech and Hearing Research*, 21, 250-264, 1978.
- [5] Y. Horii, "Fundamental frequency perturbation observed in sustained phonation", *Journal of Speech and Hearing Research*, 22, 5-19, 1979.
- [6] Y. Horii, "Jitter and shimmer differences among sustained vowel phonations", *Journal of Speech and Hearing Research*, 25, 12-14, 1982.

- [7] D. Sorensen and Y. Horii, "Frequency and amplitude perturbation in the voice of female speakers", *Journal of Communication Disorders*, 16, 57-61, 1983.
- [8] D. Sorensen and Y. Horii, "Directional perturbation factors for jitter and for shimmer", *Journal of Communication Disorders*, 17, 143-151, 1984.
- [9] F. Klingholz and F. Martin "Quantitative spectral evaluation of shimmer and jitter", *Journal of Speech and Hearing Research*, 28, 169-174, 1985.
- [10] K. D. Hall, "Variations across time in acoustic and electroglottographic measures of phonatory function in women with and without vocal nodules", *Journal of Speech and Hearing Research*, 38, 783-793, 1995.
- [11] Y. Koike "Vowel amplitude modulations in patients with laryngeal diseases", *Journal of the Acoustical Society of America*, 45, 839-844, 1969.
- [12] S. Iwata, "Periodicities of pitch perturbations in normal and pathologic larynges", *Laryngoscope*, 82, 87-96, 1972.
- [13] N. B. Pinto and I. R. Titze, "Unification of perturbation measures in speech signals", *Journal of the Acoustical Society of America*, 87, 1278-1289, 1990.
- [14] R. W. Wendahl, "Laryngeal analog synthesis of harsh voice quality", *Folia Phoniatrica*, 15, 241-250, 1963.
- [15] R. W. Wendahl "Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness", *Folia Phoniatrica*, 18, 98-108, 1966.
- [16] S. N. Awan and M. L. Frenkel "Improvements in estimating the harmonics-to-noise ratio of the voice", *Journal of Voice*, 8, 255-262, 1992.
- [17] T. Fukazawa, A. El-Assuooty and I. Honjo, "A new index for evaluation of the turbulent noise in pathological voice", *Journal of the Acoustical Society of America*, 83, 1189-1193, 1988.
- [18] N. Hiraoka, Y. Kitazoe, H. Ueta, S. Tanaka and M. Tanabe, "Harmonic-intensity analysis of normal and hoarse voices", *Journal of the Acoustical Society of America*, 76, 1648-1651, 1984.
- [19] H. Kasuya, S. Ogawa, K. Mashima and S. Ebihara "Normalized noise energy as a acoustic measure to evaluate pathologic voice", *Journal of the Acoustical Society of America*, 80, 1329-1334, 1986.
- [20] G. de Krom "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals", *Journal of Speech and Hearing Research*, 36, 254-266, 1993.
- [21] H. Muta, T. Baer, K. Wagatsuma, T. Muraoka and H. Fukuda, "A pitch-synchronous analysis of hoarseness in running speech", *Journal of the Acoustical Society of America*, 84, 1292-1301, 1988.
- [22] E. Yumoto, W. J. Gould, and T. Baer "Harmonics-to-noise ratio as an index of the degree of hoarseness", *Journal of the Acoustical Society of America*, 71, 1544-1549, 1982.
- [23] G. de Krom "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments", *Journal of Speech and Hearing Research*, 38, 794-811, 1995.
- [24] K. Shoji, E. Regenhagen, J. D. Yu and S. M. Blaugrund "H-Index: A new measure of glottal efficiency for the pathologic voice", *Laryngoscope*, 102, 1113-1117, 1992.
- [25] Emanuel, F. W. e Whitehead, R. L. (1979). Harmonic levels and vowel roughness. *Journal of Speech and Hearing Research*, 22, 829-840.
- [26] Lively, M. A. e Emanuel, F. W. (1970). Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. *Journal of Speech and Hearing Research*, 13, 498-502.
- [27] Sansone, F. E. e Emanuel, F. W. (1970). Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. *Journal of Speech and Hearing Research*, 13, 472-488.
- [28] Trullinger, R. W., Emanuel, F. W., Skenes, L. L. e Malpass, J. C. (1988). Spectral noise level measurements used to track voice improvement in one patient. *Journal of Communication Disorders*, 21, 447-457.
- [29] S. B. Davis, "Preliminary results using inverse filtering of speech for automatic evaluation of laryngeal pathology", *Journal of the Acoustical Society of America*, 58, suppl. 1, S111-S112, 1970.
- [30] S. B. Davis, "Acoustic characteristics of normal and pathological voices", In: Lass, N. J. (Ed.), *Speech and Language: Advances in Basic Research and Practice*, vol. 1, New York Academic Press, 271-335, 1979.
- [31] L. Eskenazi, D. G. Childers, and D. M. Hicks, "Acoustic correlates of vocal quality", *Journal of Speech and Hearing Research*, 33, 298-306, 1990.
- [32] Y. Koike and J. Markel "Application of inverse filtering for detecting laryngeal pathology". *Annals of Otolaryngology*, 84, 117-124, 1975.
- [33] R. A. Prosek, A. A. Montgomery, B. E. Walden and D. B. Hawkins, "An evaluation of residue features as correlates of voice disorders", *Journal of Communication Disorders*, 20, 105-107, 1987.
- [34] J. N. Holmes, "Low-frequency phase distortion of speech recordings", *Journal of the Acoustical Society of America*, 58, 747-749.

- [35] M. P. Gelfer, "Fundamental frequency, intensity and vowel selection: effects on measures of phonatory stability", *Journal of Speech and Hearing Research*, 38, 1189-1198, 1995.
- [36] Kay, S. M., *Modern Spectral Estimation: Theory and Application*, Prentice Hall Inc., Englewoods Cliff, 1988.
- [37] Deller Jr., J. R., J. G. Proakis and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., New York, 1993.
- [38] Ahmed, N, "Adaptive Filtering", In: Elliot, D. F. (Ed.), *Handbook of Digital Signal Processing - Engineering Applications*, Academic Press Inc., San Diego, 1987.
- [39] Honig, M. L and Messerschmitt, D. G, *Adaptive Filters: Structures, Algorithms, and Applications*, Kluwer Academic, Boston, 1984.
- [40] Santina, M. S., A. R. Stubberud and G. H. Hostetter, *Digital Control System Design*, Saunders College Publishing, Orlando, 1994.
- [41] Davis, M. H. A. and R. B. Vinter, *Stochastic Modelling and Control*, Chapman and Hall, New York, 1985.
- [42] S. Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Kogakusha Ltd., Tokyo, 1956.