

FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Programa Doutoral em Engenharia Biomédica

***Avaliação Acústica e Áudio
Percetiva na Caracterização da Voz
Humana***

Tese de Doutoramento

Aluna: Susana Alexandre Vaz dos Santos Freitas Nº pdb07004
Orientador: Professor Doutor Aníbal João de Sousa Ferreira

Dezembro/2012



Universidade do Porto
Faculdade de Engenharia

FEUP

FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Programa Doutoral em Engenharia Biomédica

Avaliação Acústica e Áudio Percetiva na Caracterização da Voz Humana

Tese de Doutoramento

Aluna: Susana Alexandre Vaz dos Santos Freitas Nº pdb07004
Orientador: Professor Doutor Aníbal João de Sousa Ferreira

Dezembro/2012

Este trabalho foi elaborado pela estudante Susana Alexandre Vaz dos Santos Freitas, número pdb07004, na Faculdade de Engenharia da Universidade do Porto, do Programa Doutoral de Engenharia Biomédica (PRODEB). Foi realizado no âmbito da tese de doutoramento, como parte integrante dos requisitos para a obtenção do grau de Doutor, sob orientação do Professor Doutor Aníbal João de Sousa Ferreira.

Grupo de Acompanhamento do Doutoramento:

- Professor Doutor Diamantino Freitas

- Professora Doutora Isabel Guimarães

- Professora Doutora Lúcia Mourão

RESUMO E PALAVRAS CHAVE

A análise e descrição da voz humana não se assumem como uma tarefa fácil nem consensual. A multidisciplinaridade e multidimensionalidade das tarefas de recolha das amostras de voz e as formas de classificação posteriores refletem a complexidade de todo o processo de avaliação vocal.

A componente empírica da dissertação pressupôs a realização de uma descrição aprofundada de dois dos formatos mais comuns de caracterização da voz humana – a medição através do método acústico e a classificação com base em escalas áudio-percetivas. Pretendeu-se fazer uma revisão e reflexão crítica acerca da dificuldade de associação direta destas duas dimensões.

O trabalho de investigação de campo incluiu a construção de um banco de dados de 90 vozes, com graus de alteração distintos, em adultos falantes do Português Europeu. Estas foram classificadas áudio-percetivamente, através da escala GRBAS, por um painel de 10 peritos nesta forma de avaliação. As amostras foram também analisadas por quatro tipos de *software* distintos de análise acústica: por um *software opensource* (*PRAAT*) e três comercializados (*Multi Dimensional Voice Program* - MDVP, da Kay Elemetrics; *VoiceStudio*, da Seignal; *Dr. Speech*, Tiger Electronics). O processo culminou com a descrição e discussão dos resultados obtidos.

Concluiu-se que apesar de ser cada vez mais fácil aceder a programas de *software* e existirem inúmeras propostas de medidas acústicas, nem todas têm representatividade estatística e similitude numérica entre as diferentes aplicações. No estudo desenvolvido o valor preditivo das mensurações do sinal face aos parâmetros áudio-percetivos assumiu magnitudes desde um nível fraco ($R^2_a=0,17$) a moderado ($R^2_a=0,71$). Registou-se, ainda, que a capacidade de avaliação áudio-percetiva incorre em viés, apesar dos valores dos intervalos de confiança a 95% das correlações entre as classificações médias dos peritos (CCIC) estar acima de 0.9 para os parâmetros G(Grau Geral), R(Rugosidade), e B(Soprosidade). Os parâmetros A(Astenia) e S(Tensão)

tiveram uma consistência inferior. Assim, os dez peritos manifestaram critérios similares de avaliação para os cinco parâmetros da GRBAS em análise, porém não fizeram a sua quantificação de forma estatisticamente semelhante, o que condicionou os resultados apresentados e implicou a definição de critérios para seleção dos peritos mais consistentes na classificação realizada.

A par com a realização do trabalho descrito, construiu-se um *software* de treino e teste em avaliação áudio-percetiva Voice-PE (*Voice – Perceptual Evaluation*) baseado nas classificações dos seis peritos mais consistentes, para atuais e futuros terapeutas da fala e outros utilizadores deste modo de classificação da voz, assim como são apresentadas fórmulas de quantificação/descrição da qualidade vocal que conjugaram os parâmetros da escala GRBAS e os valores das medidas de análise acústica obtidos para os diferentes tipos de *software* analisados no trabalho levado a cabo.

Apesar do estudo do sinal da fala ganhar robustez e eficiência com a possibilidade de concordância entre os seus componentes acústicos e percetivos – desde a fase do diagnóstico até à intervenção propriamente dita do terapeuta da fala – a sua concretização é difícil e variável. Caberá sempre ao avaliador apreender e dominar as diferentes formas de recolha, análise e caracterização do sinal vocal, no sentido de melhor conseguir traçar estimativas do seu processo de produção fisiológica e percepção acústica.

Palavras-Chave: voz, avaliação percetiva, parâmetros acústicos, treino áudio percetivo.

ABSTRACT AND KEY-WORDS

The study and description of the human voice is not either an easy or consensual assignment. The multi-subject and multidimensionality of the assessment of the voice samples' collection and their posterior classification reflects the complexity of the whole process of the vocal evaluation.

The empirical component of the thesis presumes a comprehensive description of the two most common forms of human voice characterization – the measurement using the acoustic method and the classification based on the use of audio-perceptual scales. We aim to make a review and critical understanding about the common difficulties in the direct association between these two dimensions of voice classification.

The field work of the thesis includes the construction of a database of 90 voice samples, with different disturbance levels, of European Portuguese adult speakers. These samples were audio-perceptually classified, using the GRBAS scale, by a panel of 10 experts in this type of assessment. The samples were also analyzed by four different types of software programs that make acoustical evaluation: one open source software (PRAAT) and three commercial ones (*Multi Dimensional Voice Program - MDVP*, by Kay Elemetrics; *VoiceStudio*, by Seegnal; *Dr. Speech*, by Tiger Electronics).

We conclude that even though it's easier to access to the software programs and there are numerous proposals of acoustic measures, not all of them are statistically representative nor have numeric semblance between the different applications. On the developed study the predictive value of the signal measures in relation to the audio-perceptual parameters assumed magnitudes since a low level ($R^2_a=0,17$) to a moderate one ($R^2_a =0,71$). We also registered that the audio-perceptual ability of evaluation incurs on a bias, despite the correlation values of 95% confidence intervals between the average classifications among the experts (CCIC) being above 0.9 for the G(Grade), R(Roughness) and B(Breathiness) parameters. The A(Asthenia) and S(Strain) parameters had an inferior consistency. Considering this, the ten experts revealed

similar evaluation criteria for the five GRBAS parameters in analysis, although they didn't make its quantification in a similar statistical way, which had an effect on the results and implied a criteria definition for the selection of the most consistent experts in this specific classification.

Along with this work, it was designed a training and test software in audio-perceptual evaluation Voice-PE (Voice-Perceptual Evaluation) based in the six most consistent experts' classifications, for actual and future speech therapists and other users of this voice classification, and we also presented quantification/description formulas of voice quality that combine the GRBAS scale parameters with the obtained measures of the acoustic analysis for the different types of software analyzed.

Even though the study of the speech signal had been empowered and gained efficiency with the possibility of agreement between its acoustic and perceptual components – since the diagnosis up until the intervention of the speech therapist *per se* – its concretization is tough and unsteady. It will always be the evaluator responsibility to apprehend and having full knowledge of the different ways of collection, analysis and characterization of the voice signal, in order to achieve a full and coherent comprehension of the physiological production and acoustical perception of voice.

Key-Words: voice, perceptual assessment, acoustic parameters, audio-perceptual training.

DEDICATÓRIA

Para a minha Mãe,
que sempre ouviu a minha Voz e nela acreditou!

AGRADECIMENTOS

À minha Família e Amigos, pela paciência e coragem ao aceitarem os atrasos, as não comparências, as faltas aos compromissos e responsabilidades, muitas vezes em detrimento de alguém ou algo que lhes era estranho, mas que eles sabem que me apaixona.

Ao meu orientador – Professor Doutor Aníbal Ferreira – por nunca deixar de acreditar que uma não-engenheira “original” poderia sê-lo na essência, na vontade de estudar, descobrir e contribuir com a tecnologia para o bem comum. Uma menção a todo o grupo de acompanhamento, especialmente à Professora Doutora Lúcia Mourão, que durante o semestre na UNICAMP possibilitou uma das maiores experiências de aprendizagem e crescimento que alguma vez tive.

À minha diretora de serviço – Dr.^a Cecília Almeida e Sousa – colega e amiga. A minha inspiração e referência para ser uma profissional de excelência e um ser-humano de exceção.

Ao Vítor Almeida e ao Pedro Melo Pestana, por serem o trabalho, o lazer, as viagens, o Norte, o crescimento e a partilha que foram cruciais ao longo destes anos.

Aos meus doentes e alunos por tudo o que me ensinaram e ensinam todos os dias.

A todos os meus Professores, Mestres e Educadores, que ao longo da vida me mostraram que se vive a aprender e se aprende a viver!

A todos... que de alguma forma contribuíram para a realização deste trabalho:
Obrigada!

ÍNDICE

	p.p.
RESUMO E PALAVRAS-CHAVE	v
ABSTRACT AND KEY-WORDS	vii
DEDICATÓRIA	ix
AGRADECIMENTOS	x
ÍNDICE	xi
ÍNDICE DE FIGURAS	xv
ÍNDICE DE TABELAS	xvii
ÍNDICE DE QUADROS	xviii
ÍNDICE DE GRÁFICOS	xxi
Capítulo 1 – INTRODUÇÃO	
1.1. Contextualização	1
1.2. Motivações	5
1.3. Objetivos	8
1.4. Metodologia	10
1.5. Organização da Dissertação	11
Capítulo 2 – A AVALIAÇÃO ÁUDIO-PERCETIVA E ACÚSTICA NA CARACTERIZAÇÃO DA VOZ HUMANA	13
2.1. A Avaliação Áudio-Percetiva	15
2.1.1. Escalas Áudio-Percetivas	18
2.1.2. O Painel de Juízes/Avaliadores Áudio-Percetivos	31
a) Erros de Medida por Fatores Associados à Tarefa de Avaliação	33
b) Erros de Medida por Fatores Associados ao Avaliador	33
c) Erros de Medida por Fatores Associados à Interação entre o Avaliador e a Tarefa de Avaliação	34
2.2. Reprodutibilidade dos Resultados da Avaliação Áudio-Percetiva	41

2.3.	A Avaliação Acústica	45
2.4.	A Avaliação e Análise Acústica da Voz	55
2.5.	Parâmetros Acústicos e Implicações Clínicas	64
	2.5.1. Espectrografia	67
	2.5.2. Frequência Fundamental	69
	2.5.3. Desvio Padrão da Frequências Fundamental (<i>pitch sigma</i>)	74
	2.5.4. <i>Jitter</i> – Perturbação da Frequência Fundamental a curto prazo	75
	2.5.5. <i>Shimmer</i> – Perturbação da Amplitude a curto prazo	80
	2.5.6. <i>HNR</i> – Proporção Harmônico-Ruído	84
2.6.	Reprodutibilidade dos Resultados da Avaliação Acústica	91

Capítulo 3 – CONCORDÂNCIA ENTRE MEDIDAS DE AVALIAÇÃO ACÚSTICA E PARÂMETROS ÁUDIO-PERCETIVOS NA CARACTERIZAÇÃO DA VOZ HUMANA

3.1.	Estudos com referência à relação entre as medidas acústicas e a avaliação áudio-percetiva	94
3.2.	Estudos com recurso a medidas acústicas ou a avaliação áudio-percetiva, associados à avaliação laríngea e/ou com alterações vocais específicas	111

Capítulo 4 – METODOLOGIA

4.1.	Materiais e métodos de recolha dos dados	
	4.1.1. Juízes	114
	4.1.2. Amostras Vocais	115
	4.1.3. Escala de Avaliação Áudio-Percetiva	116
	4.1.4. Medidas de Avaliação Acústica	119
	4.1.5. Análise Estatística	120
4.2.	Resultados	
	4.2.1. Descrição das medidas acústicas de acordo com o tipo de voz (completa ou segmentada) relativamente aos tipos de <i>software</i>	123
	4.2.2. Descrição das medidas acústicas (pares de vozes - MDVP) de acordo com a amplificação das amostras de voz	125
	4.2.3. Descrição das medidas acústicas de acordo com o programa de <i>software</i>	126

4.2.4. Descrição das medidas acústicas de acordo com o programa de <i>software</i> – representação gráfica e coeficiente de correlação	128
4.2.5. Descrição das classificações dos parâmetros áudio-percetivos de acordo com os peritos	136
4.2.5.1. Descrição da avaliação áudio-percetiva (GRBAS)	136
4.2.5.2. Descrição da consistência Inter-Avaliador (CCIC)	138
4.2.5.3. Descrição da diferença entre as vozes “originais” e respetivas “repetições”, de acordo com os peritos (Consistência Intra-Avaliador)	138
4.2.6. Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas	148
4.2.6.1. Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o <i>software Dr. Speech</i>	148
4.2.6.2. Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o <i>software MDVP</i>	150
4.2.6.3. Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o <i>software Praat</i>	152
4.2.6.4. Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o <i>software Voice Studio</i>	154
4.3 Discussão dos Resultados	
4.3.1 Análise das medidas acústicas de acordo com o tipo de voz (completa ou segmentada) relativamente aos tipos de <i>software</i>	156
4.3.2 Análise das medidas acústicas (pares de vozes – <i>MDVP-Multi Dimensional Voice Program</i>) de acordo com a amplificação das amostras de voz	157
4.3.3 Análise das medidas acústicas de acordo com o programa de <i>software</i> e coeficiente de correlação Ró de Spearman	158
4.3.4 Análise descritiva da avaliação áudio-percetiva (GRBAS)	160
4.3.5 Análise da consistência Inter-Avaliador (CCIC)	161
4.3.6 Análise da consistência Intra-Avaliador	162

4.3.7 Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas	162
4.4 Material desenvolvido ao longo da elaboração da dissertação:	
4.4.1. Descrição da contrução e funcionamento do material de treino e teste em termos de avaliação áudio-percetiva (<i>Voice-PE – Voice – Perceptual Evaluation</i>)	164
4.4.2. Estudo sobre práticas profissionais dos terapeutas da fala Portugueses na avaliação da voz humana (Coautoria)	173
4.4.3. Publicações/Estudos/Comunicações Orais sobre medidas acústicas e áudio-percetivas	173
4.4.4. Artigos propostos para publicação decorrentes da Tese	175
Capítulo 5 – CONCLUSÃO E PERSPECTIVAS FUTURAS	176
REFERÊNCIAS BIBLIOGRÁFICAS	180
ANEXOS	
Anexo I – CAPE-V – Consensus Auditory-Perceptual Evaluation of Voice <i>in Protocolo de Avaliação da Qualidade Vocal da Universidade de Aveiro (Jesus et al., 2009).</i>	I
Anexo II – Testes de <i>Kolmogorov-Smirnov</i> para as distribuições das medidas acústicas de acordo com a amplificação da voz	II
Anexo III – Testes de <i>Kolmogorov-Smirnov</i> para as distribuições das medidas acústicas de acordo com o programa de <i>software</i>	III
Anexo IV – Testes de <i>Kolmogorov-Smirnov</i> para as distribuições das medidas acústicas de acordo com o tipo de voz, para cada programa de <i>software</i>	IV
Anexo V – Simetria da distribuição das diferenças D_i	VI
Anexo VI – Regressões Lineares Múltiplas	VII

ÍNDICE DE FIGURAS

p.p.

Figura 1: Aspetto do programa informático criado por Shaheen Awan (in Awan & Lawson, 2009) para avaliação vocal áudio-percetiva	17
Figura 2: Desenho esquemático dos sistemas e órgãos envolvidos na produção do sinal de voz e do modelo linear de fonte-filtro proposto por Fant (1970) (adaptado de Dajer, 2010)	46
Figura 3: Exemplo de Sinal Tipo 1	59
Figura 4: Exemplo de Sinal Tipo 2	59
Figura 5: Exemplo de Sinal Tipo 3	60
Figura 6: Ilustração do sinal de voz captado por um microfone e correspondente à palavra <i>voz</i> . Destaca-se a região vozeada do sinal e a região não-vozeada. Usando técnicas de Processamento Digital de Sinal (PDS) é possível a medição objetiva e precisa de alguns parâmetros de perturbação extraídos diretamente do sinal acústico	64
Figura 7: Espectrograma de um caso com o software Dr. Speech®	67
Figura 8: Espectro de banda larga	68
Figura 9: Espectro de banda estreita	68
Figura 10: Representação esquemática da mudança vocal na adolescência (Probst et al., 2004)	70
Figura 11: Desenho esquemático da direção do fluxo aéreo através da glote e em contacto com a coluna de ar supraglótica (adaptado de Titze, 1994)	85
Figura 12: Sinal de onda complexa – representação isolada da onda sinusoidal e de ruído e a conjugação dos dois componentes	86
Figura 13: Página Web de avaliação áudio-percetiva	118
Figura 14: Primeira janela da página Web de avaliação áudio-percetiva	119
Figura 15: Primeira janela do Voice-PE, em Inglês.	165
Figura 16: Janela da opção de treino com classificação individual das vozes	166
Figura 17: Janela da opção de treino com classificação individual das vozes através da escala visual analógica	167

Figura 18: Janela da opção de treino com classificação individual das vozes através da escala numérica	168
Figura 19: Exemplo do sumário da avaliação áudio-percetiva através da escala visual analógica	169
Figura 20: Exemplo do sumário da avaliação áudio-percetiva através da escala numérica	170
Figura 21: Exemplo parcial do relatório a imprimir sobre a avaliação áudio-percetiva, com uso da escala visual-analógica	171
Figura 22: Janela de pré-seleção do parâmetro áudio-percetivo da escala GRBAS a treinar	171
Figura 23: Janela da avaliação comparativa, do parâmetro áudio-percetivo B(Soprosidade)	172

ÍNDICE DE TABELAS

	p.p.
Tabela 1: Definição dos parâmetros da escala GIRBAS (Behlau, 2001; Pouchoulin, 2008)	22
 Tabela 2: Definição dos parâmetros da <i>Stockholm Voice Evaluation Approach</i> (traduzido pela autora de Hammarberg, 2000)	27
 Tabela 3: Apresentação comparativa das vantagens e desvantagens das escalas percetivas – GRBAS (e adaptações mais recentes), VPAS, SVEA, Bufallo III e CAPE-V	30
 Tabela 4: Fontes da variabilidade intra e interavaliadores áudio-percetivos	38
 Tabela 5: Principais programas de <i>software</i> de análise acústica de voz	49
 Tabela 6: Caracterização dos peritos na avaliação áudio-percetiva através da escala GRBAS	114
 Tabela 7: Definição dos parâmetros áudio-percetivos da escala GRBAS (com base em Hirano, 1981; Dejonckere et al., 1996; Dejonckere et al., 2001; Webb et al., 2004)	116

ÍNDICE DE QUADROS

p.p.

Quadro 1: Características metodológicas da avaliação áudio-percetiva nos 21 estudos com vogais sustentadas usados na meta-análise de Maryn et al., (2009), um de Uloza et al., (2011) e um de Moers et al., (2012)	107
Quadro 2: Características metodológicas da avaliação áudio-percetiva dos 7 estudos com fala encadeada usados na meta-análise de Maryn et al., (2009) e um de Moers et al., (2012)	109
Quadro 3: Comparação da média (desvio padrão) das medidas acústicas de acordo com o tipo de voz (<i>Dr. Speech</i>)	123
Quadro 4: Comparação da média (desvio padrão) das medidas acústicas de acordo com o tipo de voz (<i>Praat</i>)	124
Quadro 5: Comparação da média (desvio padrão) das medidas acústicas de acordo com o tipo de voz (<i>Voice Studio</i>)	125
Quadro 6: Comparação da média (desvio padrão) das medidas acústicas de acordo com a amplificação da voz	126
Quadro 7: Comparação da média (desvio padrão) das medidas acústicas de acordo com o programa de <i>software</i>	127
Quadro 8: Coeficientes de correlação de <i>Spearman</i> entre os valores da medida acústica “ <i>Mean F0Hz</i> ” para os tipos de <i>software</i>	129
Quadro 9: Coeficientes de correlação de <i>Spearman</i> entre os valores da medida acústica “ <i>SD F0Hz</i> ” e os tipos de <i>software</i>	130

Quadro 10: Coeficientes de correlação de <i>Spearman</i> entre os valores da medida acústica “ <i>HNRdB</i> ” e os tipos de <i>software</i>	132
Quadro 11: Coeficientes de correlação de <i>Spearman</i> entre os valores da medida acústica “ <i>Jitterlocal</i> ” e os tipos de <i>software</i>	133
Quadro 12: Coeficientes de correlação de <i>Spearman</i> entre os valores da medida acústica “ <i>JitterPPQ5</i> ” e os tipos de <i>software</i>	134
Quadro 13: Coeficientes de correlação de <i>Spearman</i> entre os valores da medida acústica “ <i>Shimmerlocal</i> ” e os tipos de <i>software</i>	135
Quadro 14: Coeficientes de correlação de <i>Spearman</i> entre os valores da medida acústica “ <i>ShimmerAPQ5</i> ” e os tipos de <i>software</i>	136
Quadro 15: Estatísticas descritivas para a GRBAS de acordo com os peritos	136
Quadro 16: Estatísticas descritivas e Coeficiente de Correlação Intra-Classe para as componentes da GRBAS, de acordo com os peritos	138
Quadro 17: Estatísticas descritivas e Coeficiente de Correlação Intra-Classe para as componentes da GRB de acordo com os peritos	147
Quadro 18: Resultados das regressões lineares múltiplas com seleção de variáveis <i>Stepwise</i> , das medidas acústicas sob as componentes da escala GRBAS para o <i>software Dr. Speech</i> : coeficientes normalizados dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$)	149
Quadro 19: Resultados das regressões Lineares múltiplas com seleção de variáveis <i>Stepwise</i> , das medidas acústicas sob as componentes da escala GRBAS para o <i>software MDVP</i> : coeficientes normalizados	

dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$)

151

Quadro 20: Resultados das regressões Lineares múltiplas com seleção de variáveis *Stepwise*, das medidas acústicas sob as componentes da escala GRBAS para o software *Praat*: coeficientes normalizados

dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$)

153

Quadro 21: Resultados das regressões Lineares múltiplas com seleção de variáveis *Stepwise*, das medidas acústicas sob as componentes da escala GRBAS para o software *Voice Studio*: coeficientes normalizados

dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$)

155

ÍNDICE DE GRÁFICOS

p.p.

Gráfico 1: Representação gráfica do comportamento da medida “Mean F0” para os quatro tipos de <i>software</i> em estudo	129
Gráfico 2: Representação gráfica do comportamento da medida “SD F0” para os quatro tipos de <i>software</i> em estudo	130
Gráfico 3: Representação gráfica do comportamento da medida “HNRdB” para os quatro tipos de <i>software</i> em estudo	131
Gráfico 4: Representação gráfica do comportamento da medida “Jitter local” para os quatro tipos de <i>software</i> em estudo	132
Gráfico 5: Representação gráfica do comportamento da medida “Jitter PPQ5” para os quatro tipos de <i>software</i> em estudo	133
Gráfico 6: Representação gráfica do comportamento da medida “Shimmer local” para os quatro tipos de <i>software</i> em estudo	134
Gráfico 7: Representação gráfica do comportamento da medida “ShimmerAPQ5” para os quatro tipos de <i>software</i> em estudo	135
Gráfico 8: Valores médios da GRBAS para cada perito	137
Gráfico 9: Box-Plot das 10 vozes “originais” e respetivas “repetições” do parâmetro G-Grau da GRBAS, para os 10 peritos	140
Gráfico 10: Box-Plot das 10 vozes “originais” e respetivas “repetições” do parâmetro R-Rugosidade da GRBAS, para os 10 peritos	140

Gráfico 11: *Box-Plot* das 10 vozes “originais” e respetivas “repetições”
do parâmetro B-Soprosidade da GRBAS, para os 10 peritos 141

Gráfico 12: *Box-Plot* das 10 vozes “originais” e respetivas “repetições”
do parâmetro A-Astenia da GRBAS, para os 10 peritos 141

Gráfico 13: *Box-Plot* das 10 vozes “originais” e respetivas “repetições”
do parâmetro S-Tensão da GRBAS, para os 10 peritos 142

Capítulo 1 – INTRODUÇÃO

1.1. Contextualização

A voz é um fenómeno que comporta grandes variações (Le Huche e Allali, 2005) e depende de uma complexa e interdependente atividade de todos os músculos que servem à sua produção, além da integridade dos tecidos do aparelho fonador (Behlau, 2001). É produzida pelo trato vocal, a partir da laringe, com a passagem do ar pelas pregas vocais e o movimento das estruturas fonoarticulatórias (Akpinar et al., 2011).

Essencialmente, a voz é um fenómeno percetivo em resposta a um estímulo acústico (Eadie & Baylor, 2006; Shrivastav, 2003; Shrivastav et al., 2005). Uma voz patológica é, rotineiramente, descrita em termos de definições percetivas, o reflexo do modo como “soa ao ouvido” de quem a escuta (Webb et al., 2004).

O conceito de qualidade vocal está diretamente relacionado com preceitos fisiológicos, percetivos e acústicos (Behlau, 2001; Guimarães, 2007; Camargo e Madureira, 2010; Akpinar et al., 2011). A disfonia ou patologia vocal é definida como um distúrbio da comunicação oral, no qual a voz não cumpre a transmissão da mensagem verbal e emocional, representando uma dificuldade ou alteração na emissão, condicionando a sua produção (Behlau, 2001).

Como tal, é compreensível que seja difícil usar um único método que avalie de forma abrangente e precisa a qualidade vocal ou, correlativamente, a sua deterioração. Desta forma, é indispensável recorrer a análises multifatoriais que permitam um conhecimento amplo, adequado e eficaz da função laríngea e da qualidade vocal (Behrman, 2005; Wertzner et al., 2005; Guimarães, 2007; Behlau, 2010).

A elaboração deste trabalho torna-se pertinente na medida em que evidencia a relevância do processo de avaliação na voz falada (Gerratt et al., 1993; Carding et al., 2000; Camargo e Madureira, 2009), área de intervenção do terapeuta da fala.

Por sua vez, estando o âmbito da Voz em constante estudo, pretende-se organizar e descrever de forma aprofundada os principais processos referentes à avaliação vocal, em especial a áudio-percetiva e o método acústico, bem como as relações existentes entre eles em particular. As correspondências entre a avaliação percetivo-auditiva e as medidas acústicas da qualidade vocal são variáveis e existem estudos com resultados contraditórios (Dejonckere et al., 1996; Morsomme et al., 2001; Bhuta et al., 2004). Apesar de algumas referências a correlações positivas (Eskenazi et al., 1990; Feijoo e Hernandez, 1990; Toner et al., 1990; Rabinov et al., 1995; Scherer et al., 1995; Bielamowicz et al., 1996; Dejonckere et al., 1996; Wuyts et al., 1996), foram apontadas limitações ao estabelecimento de relações diretas entre estas duas dimensões da avaliação da voz patológica (Askenfelt e Hammarberg, 1986; Eskenazi et al., 1990; Feijoo e Hernandez, 1990; Bielamowicz et al., 1996; McAllister et al., 1996; Kent & Ball, 2000; Yu et al., 2001; Mackenzie-Beck, 2005; Zhang et al., 2005b; Maryn et al., 2009; Oates, 2009; Camargo & Madureira, 2010; Ma & Yiu, 2011).

A avaliação áudio-percetiva é considerada *gold-standard* por alguns investigadores, principalmente aqueles que a usam na rotina clínica e a veêm como soberana em relação às outras (Hammarberg e Gaufin, 1995; Camargo, 1996; Hammarberg, 2000; Behram, 2005; Blaj et al., 2007; Bunton et al., 2007; Kreiman et al., 2007; Patel & Shrivastav, 2007; Oates, 2009; Eadie et al., 2010a), embora a consistência intra e interavaliadores seja entre ligeira-moderada (Eskenazi et al., 1990; Kempster et al., 1990; Kreiman et al., 1990; Wolfe et al., 1991; Dejonckere et al., 1993; De Bodt et al., 1996; Lee et al., 2008), e o ouvido humano considerado mais confiável a avaliar vozes normais ou extremamente alteradas do que as ligeira a moderamente perturbadas (Kreiman et al., 1993; Rabinov et al., 1995; Cannito et al., 2004; Dedivitis et al., 2004). Por ser um fenómeno essencialmente auditivo, depende do treino (Bassich & Ludlow, 1986; De Bodt et al., 1997), tipo de estímulo (Bele, 2005; Zraick et al., 2005; Eadie &

Baylor, 2006; Gould et al., 2012), da instrução da tarefa (Awan & Lawson, 2009) e experiência do avaliador (Kreiman et al., 1992; Kreiman et al., 1993; Behlau, 2001; Patel & Shrivastav, 2007). Por outro lado, as pessoas com uma dificuldade vocal procuram o terapeuta da fala motivados pela forma como escutam a sua voz e é a melhoria da qualidade da voz que permite um julgamento acerca da eficácia da Terapia (De Bodt et al., 1996; Kent & Ball, 2000; Ma & Yiu, 2011). A avaliação áudio percetiva é um modo de avaliação rápido, não invasivo, acessível de imediato e que não requer o uso de qualquer equipamento eletrónico (Orlikoff et al, 1999). O terapeuta consegue, assim, transmitir a outros, inclusive ao utente, os resultados da avaliação e intervenção terapêutica, de forma sucinta e significativa (Hammarberg et al., 1986; Carding et al., 2000), bem como sustentar os seus objetivos terapêuticos e monitorizar a evolução da terapia eficazmente (Gerratt et al., 1993; Kreiman et al., 1993; De Bodt et al, 1996; Carding et al., 2000).

A análise acústica tradicional, por seu lado, baseia-se na linearidade da Teoria Acústica da Produção de Fala, modelo fonte-filtro para a produção de vogais (Fant, 1970). Esta é uma das mais importantes limitações deste método citada na literatura – o facto de assumir a independência da fonte e do filtro, uma vez que ambos interagem. Acredita-se que os músculos e tecidos se influenciam de formas não lineares e as propriedades biomecânicas são modeladas. Deste modo, pressupõe-se que a sua caracterização seja mais precisa com recurso a modelos que consideram a não-linearidade (Shrivastav & Sapienza, 2003; Dajer, 2010; Dejonckere, 2010; Shrivastav & Camacho, 2010; Dejonckere et al., 2012). Nas últimas duas décadas a dinâmica não-linear tem introduzido uma nova perspetiva para a análise de sistemas que evoluem no tempo e são sensíveis às condições iniciais, ou sejam, embora determinísticos, apresentam comportamentos aperiódicos (Repp, 1988; Klatt & Klatt, 1990; Shrivastav & Sapienza, 2003; Awan et al., 2010). A partir dos primeiros estudos dos anos 90 têm sido aplicadas na análise dos sinais da voz diversas ferramentas baseadas em dinâmica não-linear (Baken, 1990; Teager & Teager, 1990; Titze et al., 2003; Dajer, 2010; Gelzinis et al., 2008; Awan et al., 2010).

Refira-se então que, atualmente, o método acústico beneficia com o apoio das tecnologias de Engenharia de Processamento de Sinal. As tecnologias de processamento de sinal proporcionam funcionalidades a nível de *software* e de *hardware* que constituem o princípio de funcionamento do equipamento e lhe conferem uma determinada utilidade prática (Maryn et al., 2009; a,b), ao facilitarem a tarefa de análise e classificação de uma voz (alterada ou não), ao longo das várias fases do processo terapêutico (Martens et al., 2007; Awan & Lawson, 2009; Dejonckere et al., 2012).

1.2. Motivações

As razões que justificam a realização deste trabalho prendem-se com dificuldades práticas percecionadas pela doutoranda ao longo do seu quotidiano profissional, no âmbito do qual se deparou com a necessidade de resposta às questões de partida que se seguem:

- Quais as formas de avaliação, percetiva e acústica, da qualidade e das alterações da voz que existem? Quais as mais eficazes? Quais as que deverão ser incluídas/excluídas na avaliação da qualidade da voz humana? E porquê?
- Que alterações vocais permitem uma descrição fiável com recurso ao método acústico e de avaliação áudio percetiva?
- Qual a relação existente entre as escalas? Correlacionam-se? São independentes? Ou interagem? E de que forma, isto é, com que magnitude?
- Portanto, quais os parâmetros da avaliação acústica que se associam, corroboram e reforçam ou complementam a caracterização áudio-percetiva?

Como motivação mais forte para a realização desta dissertação aponta-se a necessidade de investigar o grau de proximidade entre a engenharia de processamento de sinal (método acústico) e a terapêutica da fala (método áudio-percetivo) na avaliação da voz humana. Isto porque, isoladamente, cada uma destas formas de descrição e classificação da voz e do grau de disfonia é francamente parcial e potencialmente pobre (Behlau, 2010; Akpinar et al., 2011; Ma & Yiu, 2011). Por outro lado, persistem muitas lacunas no domínio e uso de instrumentação biomédica na área da terapêutica da fala e, mais especificamente, no âmbito da patologia vocal, tal como são questionados os dados resultantes das escalas percetivas comumente usadas, devido aos índices moderados de consistência no que concerne à reproduzibilidade dos mesmos (Maryn et al., 2009; Dejonckere et al., 2012; Moers et al., 2012).

O trabalho desenvolvido foi articulado com alguns dos objetivos de investigação do projeto “ARTTS – Tecnologia de Apoio em Tempo-Real ao Canto” (FCT PTDC/SAU-BEB/104995/2008) que visa o desenvolvimento de tecnologias interactivas de apoio ao ensino e aprendizagem do canto, bem como para a monitorização preventiva da voz, seja na forma cantada ou falada.

Devido à sua natureza, o trabalho assume ainda uma motivação pedagógica com suporte tecnológico. Isto porque se pretende criar um programa de *software* associado ao banco de dados de vozes usado ao longo da dissertação, devidamente analisadas acusticamente (por mais do que um *software* de análise de sinal permitindo caracterizar um mesmo conjunto de parâmetros acústicos) e classificadas percetivamente por juízes devidamente treinados nesse campo, que se colocará à disposição de futuros e atuais terapeutas que, deste modo, terão à sua disposição uma aplicação de treino e preparação para uma mais consistente e eficaz avaliação (Awan & Lawson, 2009; Iwarsson & Petersen, 2012; Gould et al., 2012) e intervenção terapêutica.

Assim sendo, procurar-se-á selecionar um grupo ilustrativo de vozes patológicas portuguesas para criar um banco de dados de referência (“âncoras” auditivas) para a caracterização do tipo e severidade de cada parâmetro áudio percetivo – que será cotado por uma amostra significativa e válida de juízes – complementada, também, pelo estudo da consistência interavaliações de diferentes parâmetros acústicos, levadas a cabo por distintos programas de *software opensource* e comerciais.

Em síntese, na área de engenharia de processamento de sinal encontram-se problemas relacionados com a forma de análise de sinal de voz que pressupõem a definição, estudo e melhoria constante dos algoritmos de processamento de sinal. Com efeito, devido a dependerem na prática de uma abordagem matemática simplificada, que não é capaz de reproduzir a atitude analítica inerente à percepção do ouvido humano são, na maioria dos casos, de fidedignidade condicionada, não

possibilitando uma caracterização completa das vozes, em especial as patológicas (Frohlich et al., 1997; Finizia et al., 1999; Parsa & Jamieson, 2001; Maryn et al., 2009; a,b). Contudo, de acordo com a literatura, a sua concordância com a classificação áudio percetiva varia de razoável a forte para alguns parâmetros (Dejonckere, et al., 1996; Giovanni et al., 1996; Wolfe & Martin, 1997; Piccirillo et al., 1998; Wuyts et al., 2000; Bhuta et al., 2004; Eadie & Baylor, 2006; Ma & Yiu, 2006; Maryn et al., 2009; Choi et al., 2012b; Moers et al., 2012). Há também a necessidade urgente – e prévia – de enriquecimento e corroboração quantitativa dos dados percetivos, usados amplamente pelos terapeutas da fala, também eles de reproduzibilidade variável intra e interavaliadores (Oates, 2009; Eadie et al., 2010a; Kreiman & Gerratt, 2011; Choi et al., 2012a).

Para tal, e por último, prevê-se a possibilidade de criar medidas de quantificação/classificação da qualidade vocal que conjuguem os parâmetros da escala GRBAS, usada ao longo do trabalho de dissertação de doutoramento.

1.3. Objetivos

A análise e classificação da qualidade vocal são áreas abrangentes que congregam competências e conhecimentos da engenharia de processamento de sinal e da terapêutica da fala.

Assim, pretende-se ter uma noção atual das áreas do saber relacionadas com o tema, através do estudo das mais recentes técnicas de recolha, análise e classificação da voz, tanto com base no uso do método acústico, como no do áudio percetivo.

Especificamente, pretende-se conhecer cada um dos métodos, descrever e analisar diferentes estudos levados a cabo até à data sobre a correspondência detetada entre eles.

Este processo permitiu uma maior consciencialização para a temática, com vista a uma mais eficaz definição do plano de trabalho que procurou, em síntese:

1. Recolher, caracterizar e analisar a consistência interavaliadores e concordância entre os valores da cotação percetiva de um banco de dados de 90 vozes – levada a cabo por 10 terapeutas da fala/fonoaudiólogos.
2. Realizar, descrever e analisar a consistência e concordância entre as avaliações através do método acústico do banco de dados de vozes referido no ponto anterior – pelo *software opensource* (PRAAT) e três comercializados (*Multi Dimensional Voice Program* - MDVP, da Kay Elemetrics; *VoiceStudio*, da Seegnal; *Dr. Speech*, Tiger Electronics).
3. Descrever e discutir os resultados da correspondência entre a avaliação acústica e percetiva do banco de dados de vozes utilizado desde o ponto 1.

4. Construir um conjunto de fórmulas que relacionem as medidas acústicas de cada *software* estudado, com os parâmetros percetivos caracterizados pelos terapeutas da fala/fonoaudiólogos mais consistentes.
5. Encontrar um conjunto de vozes cotadas pelos peritos que representem os diferentes graus de qualidade vocal (e severidade de alteração), por forma a construir um material de escuta e treino áudio percetivo, a disponibilizar a atuais e futuros profissionais para avaliação áudio percetiva e caracterização da voz.

1.4. Metodologia

No que concerne à componente de revisão bibliográfica, esta tarefa baseou-se na leitura de bibliografia recomendada e de artigos científicos publicados recentemente e/ou obras de referência pela sua influência intemporal nos profissionais que estudam e lidam diariamente com a caracterização da voz, alterada ou não.

Mais especificamente, debruçamo-nos sobre publicações do âmbito da terapêutica da fala, da otorrinolaringologia e da engenharia biomédica/engenharia de processamento de sinal, tendo ainda participado em congressos europeus e ibero-americanos em 2009-2012, assim como realizado estágios (KTH, Estocolmo (2009); Universidade Metodista de São Paulo, na mesma cidade (2009), e Universidade Estadual de Campinas – UNICAMP, Campinas-São Paulo (2010/2011) e especializações em voz (BWizer/CEFAC, 2008), onde se expuseram e foram discutidos de forma construtiva, com profissionais de renome, os problemas atuais da avaliação e diagnóstico da qualidade vocal.

A construção do banco de dados de vozes patológicas – referência para diferentes tipos e graus de severidade de cada parâmetro de perturbação da voz – é da responsabilidade da autora da dissertação, que recolheu e armazenou amostras de vozes de casos acompanhados num serviço de otorrinolaringologia de um hospital central, ao longo dos últimos 10 anos de exercício profissional. O uso das mesmas foi autorizado pela Comissão de Ética do Hospital em causa. As mesmas foram analisadas e classificadas percetivamente por um painel de juízes (10 terapeutas da fala, 3 deles a exercer funções no Brasil, e os restantes 7 em Portugal Continental). As amostras foram classificadas acusticamente por programas de *software opensource* e *freeware versus* comerciais. Todo o trabalho culmina com a descrição e interpretação das correspondências encontradas entre a avaliação acústica e áudio percetiva das vozes. Por último, apresenta-se um *software* didático com o banco de dados de vozes de referência caracterizado ao longo de toda a dissertação (*Cf. Capítulo 4, Secção 4.4.1.*).

1.5. Organização da Dissertação

A dissertação de doutoramento está estruturada da seguinte forma:

- Capítulo 1 - INTRODUÇÃO: apresentação das motivações e objetivos da dissertação de doutoramento.
- Capítulo 2 – A AVALIAÇÃO ÁUDIO-PERCETIVA E ACÚSTICA NA CARACTERIZAÇÃO DA VOZ HUMANA: descrição e discussão dos conceitos inerentes ao tema a ser investigado – a avaliação acústica e áudio-percetiva na classificação da voz humana – com a análise da reproduzibilidade dos resultados obtidos com o uso de cada uma.
- Capítulo 3 – CONCORDÂNCIA ENTRE MEDIDAS DE AVALIAÇÃO ACÚSTICA E PARÂMETROS ÁUDIO-PERCETIVOS NA CARACTERIZAÇÃO DA VOZ HUMANA: contextualização do tema no domínio científico – congregam-se os dados sobre os trabalhos mais recentes e significativos de investigação acerca dos resultados da associação e correspondência entre diferentes medidas de avaliação acústica e formas de classificação áudio-percetiva da voz humana.
- Capítulo 4 - METODOLOGIA: descrição da metodologia de investigação desenvolvida para a recolha e análise dos dados da dissertação, explicação do trabalho de campo, exposição dos resultados obtidos e, por fim, análise e discussão dos mesmos. Apresentação das tarefas concomitantes ao decorrer do doutoramento e, também, a descrição do material pedagógico de treino áudio-percetivo construído.
- Capítulo 5 - CONCLUSÕES: conclusões e reflexões acerca dos resultados e principais dificuldades, assim como propostas para o futuro, decorrentes do processo de construção da dissertação.

Finaliza com a bibliografia consultada e de suporte a toda a dissertação, assim como os anexos, onde se encontram os testes da estatística descrita no capítulo 4.

Capítulo 2 – A AVALIAÇÃO ÁUDIO-PERCETIVA E ACÚSTICA NA CARACTERIZAÇÃO DA VOZ HUMANA

A avaliação da voz humana implica uma complexa combinação de dados que permitem a construção de um quadro conceitual que possibilita e/ou suporta a explicação do “como” e “porquê” de uma perturbação vocal apresentada (Uloza et al., 2011).

No sentido lato, a perturbação da voz é caracterizada por toda e qualquer dificuldade ou alteração na emissão vocal que impede a produção natural de voz (Verdolini, 1994; Behlau e Pontes, 1995; Verdolini et al., 2006), condicionando a comunicação oral. A voz desviada ou alterada engloba não só as perturbações dos parâmetros vocais (intensidade, altura tonal e timbre), como também as alterações cinestésicas (da dinâmica músculo-esquelética) que podem existir sem uma manifestação auditiva específica. O uso do termo voz disfónica engloba uma série ilimitada de sintomas, como: desvios na qualidade vocal, esforço na emissão do som, fonoastenia (fadiga), perda de potência/projeção vocal (diminuição do volume), variações descontroladas da frequência fundamental, quebras da intensidade vocal, baixa resistência vocal e sensações desagradáveis na emissão. Em suma, reconhece-se a existência de perturbação vocal quando: a altura tonal e a sensação de intensidade e/ou a qualidade vocal são desagradáveis ou inadequadas para a idade e sexo do indivíduo; do ponto de vista social e/ou familiar o indivíduo refere desconforto ou dor ao falar; ou se causa alguma preocupação e limitação ao próprio (Verdolini, 1994; Guimarães, 2007; Tavares & Martins, 2007).

As informações obtidas sobre a alteração da voz ao longo das diversas provas de avaliação terapêuticas existentes são usadas para o terapeuta da fala:

- a) Compreender melhor a condição vocal, por comparação com valores padronizados;
- b) Conseguir explicar as alterações (laríngeas e vocais) ao falante;
- c) Selecionar de modo adequado os métodos e técnicas de intervenção;
- d) Determinar a eficácia da Terapia implementada (Carding & Horsley, 1992; Ramig & Verdolini, 1998; Freeman & Fawcus, 2004).

Em contexto clínico e suportado numa constante articulação entre terapeutas da fala e otorrinolaringologistas – em consonância com o preconizado pelo Comité de Foniatria da Sociedade Europeia de Laringologia (ELS) – a avaliação da voz implica a recolha de informações através de:

- a) Anamnese ou entrevista clínica;
- b) Avaliação da fisiologia laríngea (laringoscopia e/ou estroboscopia e/ou electroglotografia, entre outros);
- c) Avaliação áudio-percetiva;
- d) Exame funcional (avaliação músculo-esquelética e aerodinâmica);
- e) Análise acústica; e,
- f) Autoavaliação do impacto psicossocial da voz (Dejonckere et al., 2001).

O protocolo promovido pela ELS foi elaborado com base na revisão exaustiva da literatura, experiência dos membros da comissão e discussão em plenário da sociedade, utilizando princípios básicos na enumeração e definição dos itens de avaliação anteriormente citados: a natureza multidimensional da voz, os requisitos mínimos necessários que serviram de base para meta-análises comparativas no tratamento da voz e as medidas básicas de medição nas perturbações vocais (Brasolotto & Rehder, 2011; Kreiman & Gerratt, 2011).

Da revisão da literatura sobre a temática da avaliação da qualidade da voz verificamos que existe um debate acérrimo sobre a mesma, com linhas de investigação que visam dar maior segurança e fiabilidade à análise áudio-percetiva e evidenciar o mérito do método acústico, com vista à integração dos resultados áudio percetivos e acústicos, em especial para as vozes com alterações (Awan & Roy, 2009; Jesus et al., 2009; Maryn et al., 2009; Brasolotto & Rehder, 2011; Shrivastav, 2011; Moers et al., 2012).

De seguida serão descritos os principais procedimentos de recolha, registo e análise dos dados obtidos através da avaliação áudio-percetiva e acústica da voz humana, considerados também na literatura como a análise percetiva e a objetiva (Pouchoulin, 2008), respetivamente.

2.1. A AVALIAÇÃO ÁUDIO-PERCETIVA

A avaliação áudio-percetiva pressupõe que o profissional julgue uma amostra vocal produzida pelo falante, que refere (ou não) queixas de utilização da voz. Esta forma de avaliação é amplamente usada na clínica quotidiana dos profissionais que lidam com estes casos por variadas razões, nomeadamente o facto da qualidade vocal ser percetível na sua natureza, pelo que as respetivas características assumem um valor intuitivo e passível de partilha entre distintos ouvintes (Oates, 2009; Shrivastav, 2011). Geralmente, a(s) prova(s) pedida(s) implica(m) a sustentação de vogais, a fala encadeada, entre outras, e ficam registadas em áudio e/ou vídeo (preferencialmente) para posterior análise e suporte demonstrativo (ao paciente) e comparativo com avaliações futuras.

Note-se que existem problemas inerentes à utilização da avaliação áudio-percetiva da qualidade vocal (Freeman e Fawcus, 2004; Pouchoulin, 2008; Oates, 2009; Brasolotto & Rehder, 2011; Kreiman & Gerratt, 2011), dos quais se destacam a:

1. Baixa consistência intra e interavaliadores;
2. Não disponibilização de mensurações objetivas;
3. Não existência de uma escala universal de avaliação percetiva.

A revisão da literatura demonstra que a avaliação percetiva faz depender a sua validade de vários fatores (Sederholm *et al.*, 1992; Eadie & Baylor, 2006; Ma & Yiu, 2006; Guimarães, 2007; Awan & Lawson, 2009; Awan & Roy, 2009; Jesus *et al.*, 2009; Oates, 2009; Brasolotto & Rehder, 2011; Kreiman & Gerratt, 2011), como sejam: as tarefas de avaliação, o tipo de escala usada e o formato das respostas, a qualidade vocal das amostras de voz em análise, a preparação e experiência prévias do avaliador bem como a existência de parâmetros vocais externos (por exemplo, fenómenos de coarticulação, características suprasegmentais) que funcionem como ajudas (“âncoras”) ao ouvinte (Shrivastav, 2011).

Diversos estudos mostram que a variabilidade de classificações de vozes individuais é maior para as ligeira-moderadamente alteradas, do que as dos extremos (normais ou

severamente perturbadas) (Kreiman & Gerratt, 1998; de Krom, 1994; Rabinov et al., 1995; Yu *et al.* 2001; Dedivitis et al., 2004; Bele, 2005; Eadie et al. 2005; Ma & Yiu 2006; Martens et al., 2007; Awan & Lawson, 2009; Eadie et al., 2010a).

Awan & Lawson (2009) realizaram uma investigação em que 40 alunos de terapia da fala, sem experiência de avaliação percetiva, classificavam 36 vozes (com diferentes graus de severidade e parâmetros de alteração vocal). Antes da sessão de avaliação cada juiz recebia uma breve informação sobre a tarefa (15-20 minutos) e os dados eram apresentados e recolhidos num programa informático construído pelo primeiro autor (*Cf.* Figura 1). Existiam 4 grupos de estudo, que beneficiaram de diferentes tipos de ajudas:

- Grupo 1= nenhuma ajuda (ou “âncora”);
- Grupo 2= “âncora” auditiva;
- Grupo 3= “âncora” visual;
- Grupo 4= combinação das duas “âncoras”.

Concluíram que o grupo 4 era o que obtinha mais alta consistência intra (0.85) e interavaliadores (0.78). O uso de “âncoras” auditivas diminui a variabilidade da classificação das vozes, inclusive no grupo das ligeiramente alteradas. O uso das pistas não implicou uma diferença significativa no tempo gasto para a classificação das vozes.

Este trabalho servirá como referência para a construção do *software* didático (*Cf.* Capítulo 4, Secção 4.4.1.) com o banco de dados de vozes caracterizado ao longo de toda a dissertação, destinado ao treino áudio-percetivo de futuros e atuais terapeutas da Fala ou outros profissionais que beneficiem com competências em classificação áudio-percetiva. O mesmo resulta, também, das conclusões da equipa da Universidade de La Trobe, na Austrália, liderada pela Professora Jennifer Oates, que construiu outro *software* conhecido para treino áudio-percetivo (Oates & Russel, 1997; Vertigan et al., 2008).

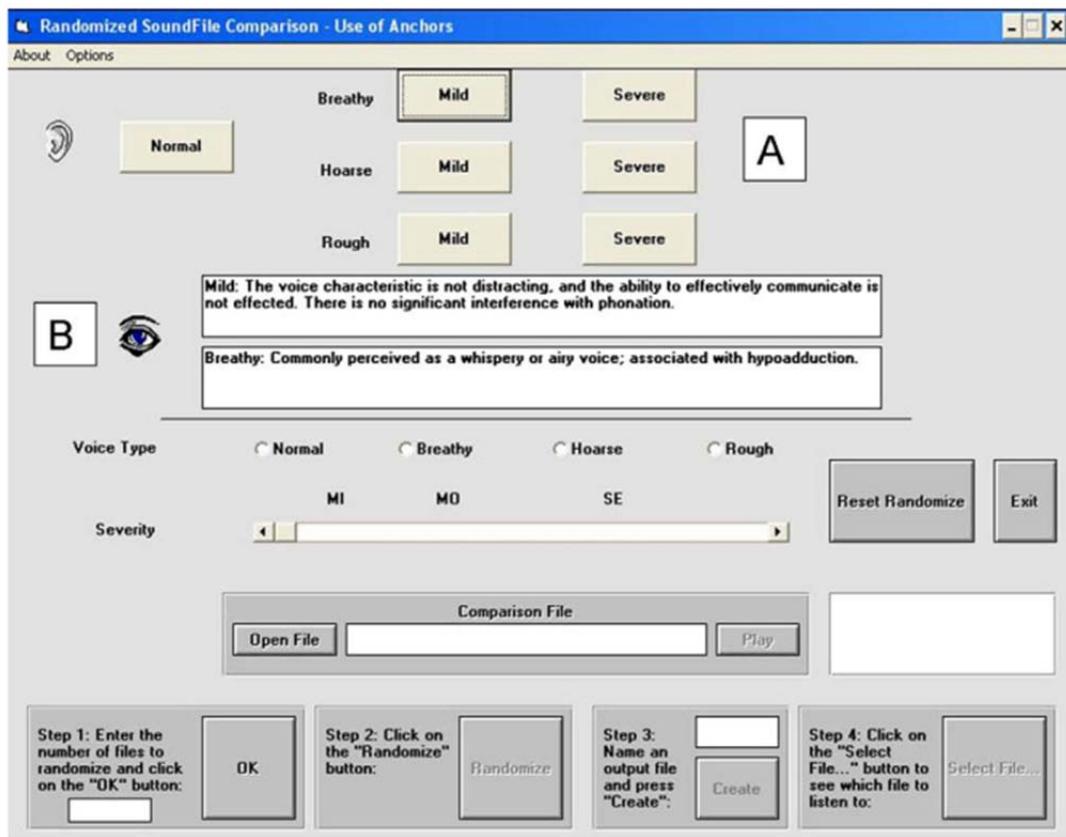


Figura 1: Aspetto do programa informático criado por Shaheen Awan (*in* Awan & Lawson, 2009) para avaliação vocal áudio-perceptiva.

Para alguns autores a avaliação perceptiva da voz patológica é uma componente central do processo de caracterização das alterações vocais (Behlau, 2001; Araújo et al., 2002; Cummings, 2008; Gould et al., 2012) e, de longe, aquela que é mais usada para descrever a voz em contexto clínico (Hammarberg, 2000; Shrivastav, 2011), atendendo à sua rapidez de uso, a ser não invasiva e confortável para o falante, à eficiência dos resultados obtidos e ao facto de implicar poucos recursos materiais (custo reduzido) (Kent, 1996; Orlikoff et al., 1999; Schiavetti & Metz, 2002; Yamaguchi et al., 2003; Webb et al., 2004; Awan & Lawson, 2009). Acrescente-se ainda que o facto da voz ser um fenómeno perceptivo por natureza faz com que seja fácil e imediato partilhar as suas características entre profissionais, pacientes e outros. Assim, a descrição perceptiva da voz assume-se como potencialmente intuitiva, significativa e interpretável para qualquer elemento que partilhe os resultados da avaliação de uma voz (Kreiman et al., 1993; Oates, 2009).

A recolha e análise áudio percetiva de amostras vocais pode ser implementada de forma formal (com recurso a escalas protocoladas) ou informal (pela análise das características da voz do paciente, indicando os diferentes sistemas intervenientes na produção da voz – respiratório, fonatório, articulatório e de ressonância). É um processo amplamente estudado, integrado e integrador entre as componentes auditiva, articulatória e fonológica da produção e percepção da fala (Liberman, 1957; Fowler et al., 1980; Stevens, 1989; Nearey, 1990; Massaro & Stork, 1998; Diehl, 2000). Será encarado ao longo desta dissertação, sucintamente, como o processo de ouvir e descrever uma voz, caracterizando-a em termos globais ou com referência a parâmetros específicos, passíveis de associação a características psicoacústicas e fisiopatológicas (Carding et al., 2000; Mathieson & Greene, 2001; Shrivastav, 2011). Existem inúmeras escalas e protocolos de avaliação áudio-percetiva, no entanto descreveremos com mais pormenor as mais referidas na literatura.

2.1.1. ESCALAS ÁUDIO PERCETIVAS: GRBAS (*Grade, Roughness, Breathiness, Astheny, Strain*) (Hirano, 1981), GIRBAS (*Grade, Instability, Roughness, Breathiness, Astheny, Strain*) (Dejonckere, Remacle & Fresnel-Elbaz, 1996), RASAT (*Rouquidão, Aspereza, Soprosidade, Astenia, Tensão*) (Pinho & Pontes, 2002), RASATI (*Rouquidão, Aspereza, Soprosidade, Astenia, Tensão, Instabilidade*) (Pinho & Pontes, 2008), GRBASH (*Grade, Roughness, Breathiness, Astheny, Strain, Harshness*) (Nemr & Lehn, 2010), (I)INFVo (*Impression, Intelligibility, Noise, Fluency, Voicing*) (Moerman et al., 2006; a,b), VPAS (*Vocal Profile Analysis Scheme*) (Laver, 1980), BVP (*Buffalo III – Voice Screening Profile*) (Wilson, 1987), SVEA (*Stockholm Voice Evaluation Approach*) (Hammarberg, 1986, 2000), e o CAPE-V (*Consensus Auditory Perceptual Evaluation of Voice*) (Kempster et al., 2009).

A análise áudio percetiva da voz patológica tem como essência o princípio de que voz e audição são indissociáveis (Pouchoulin, 2008; Souza, 2010). Deste modo, é compreensível que uma das formas mais expeditas de classificação vocal se baseie na escuta impressionista (Orlikoff et al., 1999), mesmo com o reconhecimento mundial do pouco consenso em torno da definição dos parâmetros percetivos que descrevem a qualidade da voz (Awan & Roy, 2009; Kreiman & Gerratt, 2011), o que condiciona o

seu uso para efeitos médico-legais. Esta é a vertente qualitativa das escalas áudio percetivas. A escala de medida dos parâmetros assume-se como a vertente quantitativa (Awan & Lawson, 2009). As três principais formas de quantificação são:

- A escala bipolar, que preconiza a dicotomia “sim/não” (ou ainda as noções de “presença/ausência”) de pares de parâmetros de qualidade vocal. A título de exemplo, os pares hiperfuncional/hipofuncional ou estável/instável.
- A escala de classe numérica ou EAI (*Equal-Appearing Interval*) propõe múltiplos níveis equidistantes entre 1 e n, geralmente entre 4 a 7 níveis, para quantificar separadamente cada um dos parâmetros qualitativos (Laver, 1980; Hirano, 1981; Dejonckere et al, 1993; de Krom, 1994; Nawka et al., 1994; De Bodt et al., 1996; Wolfe & Martin, 1997; Behlau, 2001; Dedivitis et al., 2004; Webb et al., 2004; Martens et al., 2007; Moers et al., 2012). Kreiman et al. (1993) consideraram que a escala com 7 níveis é a mais apropriada para avaliar vozes patológicas do que as que incluem 4 possibilidades de classificação. Awan & Roy (2009) ratificam as conclusões do estudo anterior.
- A escala analógica visual (VAS – Visual Analogue Scale) permite atribuir visualmente um grau de severidade do parâmetro vocal analisado, ao indicar com uma cruz sobre uma linha, geralmente de 10cm, sendo que quanto mais à direita, mais alterada a qualidade vocal (Takahashi, 1976; Kreiman et al., 1993; Behlau, 2001; Cannito et al., 2004; Carding, 2004; Awan & Lawson, 2009). A distância medida da esquerda ao local da cruz indica o grau de severidade estimado (Chan, 2011). Simberg et al. (2000) utilizaram a VAS para discriminar entre vozes normais e alteradas, definindo como ponto de corte os 34mm. Um estudo brasileiro (Yamasaki et al., 2008) demonstrou que esta é uma forma de classificação mais sensível às alterações vocais, robusta para triagem, que não sofre influência de aspectos culturais, com pontos de corte bastante próximos ao do estudo anterior: 34,5mm (Yamasaki et al., 2007) e 35,5mm (Yamasaki et al., 2008).

Note-se que, em 1999, Wuyts et al. desenvolveram uma investigação em que 29 avaliadores caracterizaram através da escala GRBAS (Hirano, 1981) 14 vozes patológicas, com uma escala ordinal (de 0 a 3) e uma visual-analógica (100mm). A escala ordinal foi a que obteve maior consistência interavaliadores. Esta também

refletiu de forma mais eficaz as diferenças de interpretação entre os parâmetros da escala. Yiu & Ng (2004) e Yiu et al., (2007) ratificam estes achados. Contrariamente, Karnell et al. (2007) concluíram que a classificação com a VAS usada na escala CAPE-V era mais sensível a pequenas diferenças de qualidade vocal que a EAI da GRBAS. Yu et al. (2002) também concluíram que a concordância entre a classificação da GRBAS através da VAS se correlacionava de forma mais forte com as medidas de avaliação acústica, do que as classificações derivadas de níveis de EAI (88% *versus* 64% de concordância). Recentemente, Moers et al. (2012) não obtiveram valores distintos, tanto com o uso da VAS como da EAI, na avaliação por 5 juízes experientes (mais de 10 anos de prática profissional) do parâmetro qualidade vocal global (“*overall quality*”).

A investigação de Yamasaki et al. (2008) estabeleceu uma correspondência entre a VAS e a EAI de 4 pontos na avaliação áudio-percetiva da voz, com base na classificação – por 4 avaliadores – de 211 vozes de adultos, com e sem queixa vocal. A determinação dos valores de corte para os diferentes graus (parâmetros G-Grade) de alteração vocal mostrou-se possível através da análise das curvas ROC¹. Assim, de 0-35,5mm estão valores de variabilidade normal da qualidade vocal (VNQV), o que inclui o desvio leve, denominado grau 1 na escala EAI (ou VNQV – variabilidade normal da qualidade vocal). A VNQV pode expressar tanto manifestações de estilo vocal, preferência de uso da voz, caracterização profissional, ou um desvio vocal de grau leve. O grau 2 representou a região entre desvio leve e moderado (35,5-50,5mm). O grau 3 correspondeu ao intervalo para a categoria de desvio moderado a severo (50,5-90,5mm) e, uma faixa exígua para as alterações severas (acima dos 90,5mm), representativa do grau 4. Note-se que os resultados mostraram que as diferenças entre estas quatro categorias não obedecem à partição matemática de 25mm intervalos iguais (Madazio & Behlau, 2009).

¹ A análise da área sob a curva ROC (*Receiver Operating Characteristic*) é uma medida da capacidade de qualquer modelo estatístico descriminar os sujeitos com a característica de interesse *versus* sujeitos sem a característica de interesse. Baseia-se na teoria de deteção de um sinal na presença de ruído, e corresponde à representação gráfica da probabilidade de se detetar o sinal verdadeiro (i.e. sensibilidade) e o sinal falso (1 – especificidade) (Maroco, 2010).

São várias as escalas usadas desde a década de 80 do século passado, as quais serão descritas de seguida e discutidas nos estudos publicados com o uso das mesmas. De notar que os protocolos de classificação da voz humana não estão sustentados por um modelo teórico de qualidade vocal e – porque o constructo a ser medido não está claramente definido – não é possível estabelecer se um determinado protocolo ou escala ou medida quantitativa é a “correta” para este efeito (Kreiman & Gerratt, 2011). Esta limitação relacionada com a validade do instrumento, o que se está a medir e o porquê, tem reforçado a noção de subjetividade inerente ao uso de escalas áudio-perceptivas e condicionado a dificuldade (e até impossibilidade) de determinar se uma escala ou protocolo de avaliação é necessário e/ou suficiente.

A escala mais conhecida e usada mundialmente é a GRBAS, de Hirano (1981) (Hammarberg, 2000; Cummings, 2008; Wirz & Beck, 1995 *in* Pouchoulin, 2008). Esta foi desenvolvida e implementada em 1969 pelo Comité para Testes da Função Fonatória da Sociedade Japonesa de Logopedia e Foniatria, baseada nos trabalhos de Isshiki et al. (1969). Cada um dos cinco grafemas é o acrônimo de G – avaliação global da disfonia (*grade*); R – aspereza/rugosidade (*roughness*); B – soprosidade (*breathiness*); A – astenia (*astheny*); S – tensão (*strain*) (Hirano, 1981 *in* Guimarães, 2007). É uma escala compacta e de simples utilização, eficaz para a triagem vocal, independente da língua usada, que avalia a fonte glótica, pelo que depende da sonorização do sopro pulmonar durante a produção de vogais sustentadas (/a/ ou /ε/), leitura ou fala encadeada (Pinho, 2003; Pouchoulin, 2008). Os parâmetros avaliados são classificados numa escala de quatro pontos (níveis de gravidade): 0= normal ou ausência de alterações; 1= ligeiro ou discretas modificações; 2= moderado ou alterações evidentes; 3= severo/grave ou com variações extremas.

Em 1996 Dejonckere, Remacle & Fresnel-Elbaz et al. propuseram acrescentar o fator Instabilidade (I) à escala GRBAS original. Os parâmetros desta escala estão caracterizados na Tabela 1.

<i>Parâmetros</i>	<i>Definição</i>
<i>G – Grade</i>	Grau de alteração vocal – impressão global da voz, identificação do grau de alteração vocal, como um todo
<i>I – Instability</i>	Instabilidade, flutuação na F0 e/ou na qualidade vocal
<i>R – Roughness</i>	Irregularidade na vibração das pregas vocais, sensação de “rugosidade” (<i>raucité</i> em Francês) que corresponde a flutuações irregulares da F0 e/ou da amplitude do som glótico
<i>B – Breathiness</i>	Soprosidade, relativo a turbulência audível, escape de ar importante através da glote, sensação de ar na voz
<i>A – Asteny</i>	Astenia, fraqueza vocal, falta/perda de potência na voz relacionada com intensidade fraca, energia vocal reduzida, harmónicos pouco definidos e pouco elevados em relação ao nível de ruído
<i>S – Strain</i>	Tensão, impressão de estado hiperfuncional, frequência anormalmente aguda (ruído nas frequências altas do espectro e/ou riqueza de harmónicos agudos marcados)

Tabela 1: Tradução da definição dos parâmetros da escala GIRBAS (Behlau, 2001; Pouchoulin, 2008; Kreiman & Gerratt, 2011)

Existem críticas ao uso destas escalas, nomeadamente porque não contemplam a classificação da adequação da frequência fundamental e outras características da voz (Kreiman & Gerratt, 2011). Estudos identificaram ainda a baixa fiabilidade dos parâmetros *Asteny-Astenia* e *Strain-Tensão* (Dejonckere et al., 1993; De Bodt et al., 1997; Dejonckere et al., 2001; Webb et al., 2004; Gould et al., 2012). Contudo, este é um método que se correlaciona moderadamente com outras formas de classificação vocal, nomeadamente os questionários de avaliação do impacto da disfonia na qualidade de vida (Hogikyan & Sethuraman, 1999; Hogikyan et al., 2000; Jones et al., 2006; Karnell et al. 2007; Franco & Andrus, 2009).

Em 2002 Pinho (Pinho, 2003) sugere a adoção, no Brasil, da escala RASAT a partir da GRBAS. As siglas indicadas correspondem, do ponto de vista anatomo-fisiológico e áudio percutivo, a:

Rouquidão (grau de): irregularidade (aperiodicidade) de vibração das pregas vocais durante a fonação e a fase de fechamento durante a adução das pregas vocais é

incompleta (Andrews, 1995). Assim, a voz é percecionada com ruídos produzidos a baixa frequência (Hirano et al, 1990; Hammarberg e Gauffin, 1995; Hammarberg, 2000; Pinho, 2008) que mascaram os harmónicos inferiores no traçado spectrográfico, ou surgem sub-harmónicos (Martens et al., 2007). Este parâmetro verifica-se em casos de: fenda glótica maior ou igual a 0,5mm², presença isolada de uma alteração orgânica ou fenda de qualquer dimensão com alterações da mucosa vibrante (Isshiki, 1980; Pinho, 2003) das pregas vocais (por exemplo: nódulos, hiperemias ou edemas²) (Colton & Casper, 1996).

Aspereza: rigidez da mucosa que também causa alguma irregularidade vibratória, especialmente se associada a fenda glótica ou outras alterações laríngeas como, por exemplo, edema das pregas vocais (Edema de Reinke³). A voz é seca, sem projeção, com ruídos nas altas frequências pela diminuição da onda mucosa (Hirano et al., 1986; Hammarberg & Gauffin, 1995; Hammarberg, 2000) que implica um maior dispêndio de energia e consequente gasto de ar para desencadear a vibração de uma mucosa rígida – notando-se a proeminência dos harmónicos inferiores. A aspereza é percebida nas vozes produzidas com esforço excessivo, tensão laríngea e constrição do trato vocal (Oliveira, 2010). O exemplo clássico de rigidez é a voz dos casos com sulco vocal⁴, quistos⁵, retracções cicatriciais e lesões neoplásicas⁶.

² Os nódulos são as lesões benignas mais comuns das pregas vocais, e normalmente resultam de hiperfuncionamento vocal, sendo protruções bilaterais, não obrigatoriamente simétricas, edematosas e/ou de colagénio. A laringite é uma inflamação/hiperemia que pode ser aguda ou crónica, associada a agentes infeciosos, alérgicos, tóxicos, térmicos ou físicos. Na laringite aguda normalmente surge edema e inflamação, por sua vez, o caráter crónico provoca hipertrofia ou metaplasia da mucosa e fibrose das camadas mais profundas (Verdolini, 1994; Sataloff, 1997; Pinho, 1998; Guimarães, 2007).

³ No edema de Reinke observa-se uma acumulação dispersa de fluido, ou edema, ao longo da zona membranosa das pregas vocais, concretamente na camada superficial da lámina própria, também designado como espaço de Reinke. É habitualmente bilateral e assimétrico, sendo mais frequente no sexo feminino com hábito tabágico marcado. Para além do tabaco a sua etiologia encontra-se associada a alergias e sinusite crónica (Verdolini, 1994; Sataloff, 1997; Pinho, 1998; Guimarães, 2007).

⁴ O sulco glótico ou sulcus glottidis, por sua vez, caracteriza-se por um sulco longitudinal, uni ou bilateral, decorrente do aumento de fibras de colagénio ao longo da zona membranosa das pregas vocais. Esta situação provoca rigidez na prega vocal, diminuindo a amplitude de vibração e a onda da mucosa durante a fonação. As causas são pouco claras, no entanto, aponta-se para fatores congénitos e de

Soprosidade: presença de ruído de fundo, audível e visível em diferentes regiões do espectro (Martens et al., 2007), que corresponde fisiologicamente e de forma proporcional à fenda glótica, com voz de *loudness* reduzido (Andrews, 1995; Colton & Casper, 1996; Oliveira, 2010). Salienta-se que, em casos excepcionais, pode encontrar-se este parâmetro evidenciado face a rigidez extrema da mucosa na ausência de fenda glótica (Pinho et al., 2008). A soprosidade pode estar presente nos casos de fendas glóticas isoladas, assimetria do padrão vibratório das pregas vocais, lesões do tipo nódulos, quistos, pólipos, papilomas⁷ e edemas, e em pacientes com doenças neurológicas degenerativas (Stemple et al, 1995). O ruído é o parâmetro mais presente nos casos com disfonia (Butha, et al., 2005).

Astenia: relacionada com o mecanismo de hipofunção das pregas vocais e reduzida energia de emissão do som. É esperada uma menor definição dos harmónicos de altas frequências (Martens et al., 2007). Exemplo: *miastenia gravis* ou outras perturbações neurológicas do controlo vocal (Behlau, 2001).

Tensão: associada a esforço vocal por aumento da adução glótica (hiperfunção), geralmente inerente ao aumento da atividade da musculatura extrínseca da

desenvolvimento, efeitos secundários a processos crónicos de inflamação repetidos ou de longa duração, assim como a sobrecarga vocal (Verdolini, 1994; Sataloff, 1997; Pinho, 1998; Guimarães, 2007).

⁵ Os quistos e pseudo-quistos mucosos são geralmente unilaterais, contudo podem causar inflamação na prega vocal contralateral devido ao contacto. São protrusões (sacos de epitélio cheios de líquido) na camada superficial da lâmina própria, no bordo livre das pregas vocais. Podem ser epidermóides ou de retenção. O quisto de retenção tem como causa provável o bloqueio glandular, enquanto que o epidermóide tem normalmente origem congénita (Verdolini, 1994; Sataloff, 1997; Pinho, 1998; Guimarães, 2007).

⁶ As neoplasias são neoformações tecidulares de caráter maligno, granulosas, brancas, que não invadem a superfície dos tecidos, mas penetram no corpo das pregas vocais. A etiologia é desconhecida, mas normalmente está relacionada com o consumo de tabaco e álcool, substâncias irritantes ambientais, químicos, alterações metabólicas e causas desconhecidas (Guimarães, 2007).

⁷ Os papilomas são uma lesão causada, normalmente, por uma infecção viral, sendo mais prevalentes em crianças. São tumores epiteliais benignos pequenos, com coloração variável entre vermelho e esbranquiçado, com a textura semelhante a uma amora, ocorrendo essencialmente na região glótica e supraglótica, no entanto pode acontecer o envolvimento de outras regiões, como a nasofaringe, subglote, traqueia e brônquios (Verdolini, 1994; Sataloff, 1997; Pinho, 1998; Guimarães, 2007).

laringe, com elevação da posição desta (Behlau, 2001). São evidenciados harmónicos em altas frequências (Martens et al., 2007). Exemplo: disfonia espasmódica em adução e síndromes de abuso vocal (síndrome de tensão musculo-esquelética), com consequente alteração de massa (i.e. nódulos ou pólipos).

A escala RASAT obedece ao mesmo sistema de classificação da GRBAS, em quatro níveis, mas são também contemplados valores intermédios (Pinho et al., 2008). Baseia-se em amostras de vogais sustentadas (/a/ ou /ε/) ou fala encadeada (Pinho et al., 2008). Assim, por exemplo, um falante com disfonia de rouquidão moderada, aspereza moderada-severa, soprosidade ligeira, sem astenia e com tensão ligeira-moderada, seria classificado como R₂A_{2,5}S₁A₀T_{1,5}.

Em 2008, Pinho & Pontes acrescentaram o parâmetro I (Instabilidade) à escala previamente explicada, traduzindo a flutuação das qualidades da voz avaliada. Mantiveram os procedimentos de recolha de dados e caracterização da escala RASAT, de 2002. A RASATI continua a centrar-se na avaliação do nível laríngeo, sendo um procedimento de triagem vocal rápido, compacto e fiável, com elevado grau de fiabilidade (Pinho & Pontes, 2008).

Em 2010 foi apresentada a GRBASH (*Grade, Roughness, Breathiness, Astheny, Strain, Harshness*) (Nemr & Lehn, 2010). As colegas que a utilizam na prática clínica lidam, especialmente, com casos oncológicos de cabeça e pescoço. A definição dos parâmetros e respetiva cotação é a mesma que a original GRBAS, apenas com o H (*Harshness*) relativo à aspereza, preconizada por Pinho (2008) na escala RASAT e RASATI.

Saliente-se que para estes casos – de vozes de substituição – em que o parâmetro G da GRBAS nunca assume o valor zero (Dejonckere, 2010) foi criada a escala (I)INFVo (Moerman et al., 2006; a,b) desenvolvida especialmente para classificar vozes cujo som é gerado numa neoglote, decorrente de ablações cirúrgicas (de grande parte ou a totalidade da laringe). Deste modo, cada sigla do nome da escala é o acrônimo de:

I – impressão geral da qualidade da voz, aceitabilidade e adequação à comunicação diária;

I – impressão de inteligibilidade;

N – ruído extra não intencional, o que reflete a perturbação de ruídos disruptivos incontrolados (por exemplo: voz molhada, clicks articulatórios ou de injeção de ar);

F – fluência, o que reflete a suavidade percecionada da produção da fala, e contabiliza todo o tipo de interrupções indesejadas;

Vo – vozeamento; correspondente a frases escutadas como vozeadas ou desvozeadas quando assim se espera que os sons sejam produzidos fonemicamente.

A sua cotação é a mesma que a original GRBAS – em 4 pontos (níveis de gravidade): 0= normal ou ausência de alterações; 1= ligeiro ou discretas modificações; 2= moderado ou alterações evidentes; 3= severo/grave ou com variações extremas; ou, através de uma escala visual analógica (100mm).

A escala *Vocal Profile Analysis Scheme* (VPAS) é muito usada pelos terapeutas da fala do Reino Unido (Cummings, 2008), baseando-se nos trabalhos de Laver et al. (1980). Este protocolo é uma aproximação à medição da qualidade da voz suportada numa abordagem teórica que caracteriza a voz de acordo com o modo fisiológico como ela é produzida (Kreiman & Gerratt, 2011). Pressupõe a descrição fonética da qualidade vocal partindo da referência a uma posição articulatória neutra, definida pela fisiacústica, na análise de provas de leitura ou fala espontânea. Assim, possibilita a descrição de características laríngeas e supraglóticas (trato vocal) relativamente a 31 parâmetros distribuídos por três sub-categorias específicas:

- A qualidade vocal: as qualidades laríngeas e supralaríngeas do som.
- A qualidade prosódica: o *pitch* (média, gama tonal e variabilidade), tremor e o *loudness* (média, extensão e variabilidade).
- A organização temporal: o ritmo respiratório, a continuidade, o débito, a cadênci a.

A classificação final nesta escala é apresentada em 6 graus. Conseguir levar a cabo esta tarefa pressupõe um treino prévio do avaliador ao longo de 12 horas de visualização de cassetes áudio e vídeo. Estudos revelam uma taxa de concordância inter

avaliadores de 65-75% (Wirz & Beck, 1995). Infelizmente a compreensão dos termos descritivos da qualidade vocal não é facilmente traduzida para um modelo psicoacústico de caracterização da voz humana, já que os modelos de produção não predizem os parâmetros que se salientam áudio-percetivamente, e sob que circunstâncias (Kreiman et al., 2005). Ou seja, o perfil de análise vocal especifica informação sobre o local onde a qualidade da voz deve estar a ser produzida, mas não é capaz de modelizar o comportamento do ouvinte que a classifica. Este protocolo resulta como um modelo apenas e só quando se pressupõe que cada característica fisiológica é percetivamente significativa, e que o ouvinte consegue separar os vários parâmetros descritivos usados.

Outra escala comumente usada na avaliação percetiva da voz é a *Buffalo III – Voice Screening Profile* (Wilson, 1987). Esta classifica numa escala de cinco pontos (em que 1=normal e 5=muito grave) os seguintes parâmetros: tônus laríngeo, timbre, *loudness*, *pitch*, ressonância nasal, ressonância oral, suporte respiratório, musculatura de abuso/esforço vocal, o débito da fala, a ansiedade de fala, inteligibilidade da fala e eficiência vocal geral. Baseia-se em provas de leitura, conversação espontânea, vogal sustentada e contagem de números. Os resultados da sua aplicação possibilitam a classificação da gravidade da disfonia, o seu impacto sobre a comunicação e, ainda, orientação no que concerne à intervenção terapêutica a levar a cabo.

Ao longo das últimas duas décadas do Séc. XX o departamento de Terapia da Fala do Hospital de Huddinge (Suécia) desenvolveu e aperfeiçou a *Stockholm Voice Evaluation Approach* (SVEA) (Hammarberg, 1986; Hammarberg & Gaufin, 1995). Esta escala partiu da análise de correlações entre 28 variáveis (baseada em 50 termos percetivos usados pelos clínicos), de que resultaram 13 parâmetros qualitativos propostos para a avaliação percetiva de vozes normais e patológicas (ver Tabela 2), em cinco níveis quantitativos (em que 0=normal e 4=muito grave), baseada na análise da voz durante a leitura de um texto com a duração de aproximadamente 40 segundos. Esta escala provou ser muito útil para fins clínicos (tanto com população infantil como adulta), assim como na formação/treino de futuros profissionais na avaliação percetiva da voz (Hammarberg, 2000).

Parâmetros	Tentativa de Definição
Afonia/Afonia intermitente	A voz tem, de forma permanente ou intermitente, ausência de sonoridade. i.e., existem momentos de sussurro ou perda de voz
Soprosidade	A voz é produzida com encerramento glótico insuficiente, as pregas vocais vibram mas abduzidas o que cria um ruído de turbulência audível na glote
Hiperfuncional/Tensa	A voz soa esforçada, devido à compressão/constricção das pregas vocais e da laringe durante a fonação com sopro insuficiente
Hipofuncional/Laxo	É o oposto da hiperfuncional. Existe tensão vocal insuficiente e pouca atividade da musculatura laríngea, resultando em voz fraca
Vocal fry/Crepitante	Vibração periódica/aperiódica em baixas frequências, as pregas vocais estão muito próximas e apenas uma secção livre para vibrar; também conhecido como registo pulsátil
Rugosidade	Aperiodicidade em baixas frequências, presumivelmente relacionada com algum tipo de irregularidade vibratória
Aspereza ("scrapiness")	Aperiodicidade em altas frequências, presumivelmente relacionada com algum tipo de irregularidade vibratória
Instabilidade vocal/ de <i>pitch</i>	A voz varia em termos de <i>pitch</i> ou de qualidade vocal ao longo do tempo
Quebras de voz	Quebras intermitentes entre o registo modal e o registo falsete
Diplofonia	Dois <i>pitches</i> diferentes podem ser percecionados ao mesmo tempo
Registo Modal/Falseto	Modos de fonação: o registo modal é o modo fonatório normal, padrão, com encerramento glótico durante a fonação; em oposição com o falsete, que usualmente está no topo do alcance de frequência fundamental, como resultado do alongamento e afunilamento das pregas vocais, com encerramento insuficiente
<i>Pitch</i>	Correlação auditiva com a frequência fundamental
<i>Loudness</i>	Correlação auditiva com nível de pressão sonora da fala

Tabela 2: Tradução da definição dos parâmetros da *Stockholm Voice Evaluation Approach* (Hammarberg, 2000)

Os autores reforçam que desenvolveram um material para melhorar a performance dos avaliadores nos testes de escuta, associado ao SVEA. Assim, foi construído um programa de avaliação percetiva suportado por computador (VISOR – *Visual Sort and Rate*) que provou ser um método eficaz para a classificação numa escala analógica visual das cotações percetivas de vozes patológicas (Granqvist et al., 2003).

Por último, o CAPE-V (*Consensus Auditory Perceptual Evaluation of Voice*) é uma escala de avaliação percetiva que classifica seis parâmetros vocais (Severidade Global, Rugosidade, Soprosidade, Tensão, *Pitch* e *Loudness*) através de uma escala visual analógica de 100mm (complementada por outros descritores: consistência/inconsistência) assim como dois dados vocais adicionais, como a classificação da ressonância ou tremor, que foi desenvolvida em 2002 depois de uma conferência da ASHA (*American Speech-Language-Hearing Association*) e traduzida/adaptada para o português brasileiro por Behlau (2004) e para o europeu na Universidade de Aveiro (em fase de aferição) (*Cf. Anexo I*).

Baseia-se em provas de vogal sustentada (/ a / e / i /, durante 3-5seg), leitura de frases pré-definidas e conversação espontânea. A escala foi, em parte, suportada nos trabalhos de Gerratt et al. (1993) que comprovaram a consistência dos resultados obtidos com escalas analógicas *versus* as ordinais ou com referência a intervalos. As instruções para o seu uso e cotação estão disponíveis *online* no site da *American Speech-Language-Hearing Association's Division 3 for Voice and Voice Disorders* (Kempster et al., 2009; Chan, 2011).

Optamos pela descrição destas escalas uma vez que são as mais encontradas na literatura da área. Como forma de compreensão rápida e fácil das suas características, foi construída pela autora da dissertação a Tabela 3 (Apresentação comparativa das vantagens e desvantagens das escalas percetivas – GRBAS (e adaptações mais recentes), VPAS, Bufallo III, SVEA e CAPE-V).

ESCALA	Vantagens	Desvantagens
GRBAS/ GIRBAS/ RASAT/ RASATI/ GRBASH/ IINFVo	<ol style="list-style-type: none"> 1. Baseado em provas clínicas simples (fala espontânea e/ou vocal sustentada) 2. Classificação final em níveis (0-3) 3. Classifica parâmetros laríngeos – cinco – pertinentes 4. Define a terminologia 5. Relaciona com a função fisiológica 6. Baseado na teoria acústica 7. Fácil/rápido de usar (<5min)/aprender 	<ol style="list-style-type: none"> 1. Classifica só o nível laríngeo (i.e. sem parâmetros supraglóticos) 2. Não classifica parâmetros usados noutras escalas, como o <i>pitch</i> e <i>loudness</i> 3. Sem treino formal pré-definido
VPAS	<ol style="list-style-type: none"> 1. Análise detalhada de configurações do trato vocal (31 parâmetros) 2. Classificação final em seis níveis 3. Caracteriza fatores vocais individuais 4. Adequado a vozes normais e patológicas 5. Relaciona com a função fisiológica 6. Implica programa de treino de dois dias 	<ol style="list-style-type: none"> 1. É necessária a prática regular das competências de escuta 2. Aprendizagem morosa 3. Com provas específicas (leitura com frases foneticamente equilibradas) para a língua inglesa 4. Maior dispêndio de tempo para a recolha de dados e análise/classificação (>10min)
Buffalo III	<ol style="list-style-type: none"> 1. Analisa uma vasta gama de categorias (12 parâmetros) glóticas, supraglóticas e do comportamento vocal 2. Sugere intervenção terapêutica 3. Classificação final em níveis (1-5) 4. Fácil/rápido de usar (5-10min)/aprender 	<ol style="list-style-type: none"> 1. Inclui parâmetros não associados à voz 2. Sem treino formal pré-definido 3. Com provas específicas (leitura com frases foneticamente equilibradas) para a língua inglesa
SVEA	<ol style="list-style-type: none"> 1. Classifica parâmetros laríngeos – treze – pertinentes 2. Classifica alterações vocais 3. Classificação final em níveis (0-4) 4. Define a terminologia 5. Relaciona com a função fisiológica 6. Baseado na teoria acústica 7. Fácil/rápido de usar (5-10min)/aprender 8. Com software de suporte ao treino (VISOR) 	<ol style="list-style-type: none"> 1. Sem treino formal pré-definido 2. Com características vocais influenciadas pela cultura (ex. Fry), que tornam generalização difícil 3. Com provas específicas (leitura com frases foneticamente equilibradas) para a língua sueca
CAPE-V	<ol style="list-style-type: none"> 1. Analisa categorias (6 parâmetros) glóticas e supraglóticas 2. Classificação final numa escala analógica visual 3. Possibilidade de acrescentar parâmetros adicionais 4. Define a terminologia e possui um site de orientação para o seu uso e cotação 	<ol style="list-style-type: none"> 1. Sem treino formal pré-definido 2. Maior dispêndio de tempo para a recolha e análise de dados (5-10min) 3. Com provas específicas (leitura com frases foneticamente equilibradas) para a língua inglesa

Tabela 3: Apresentação comparativa das vantagens e desvantagens das escalas percetivas – GRBAS (e adaptações mais recentes), VPAS, SVEA, Bufallo III e CAPE-V.

As semelhanças entre os protocolos e a falta de mudanças significativas ao longo das últimas décadas (e/ou até a aglutinação de parâmetros transversais às diferentes escalas) reforça a (falsa) impressão de que as principais dificuldades e desafios inerentes à descrição da qualidade da voz estão resolvidos e, como tal, os clínicos e

investigadores da área concordam acerca da melhor forma de avaliação da voz humana. O uso da expressão “*consensus model*” associado a protocolos como o CAPE-V ou o SVEA são exemplo deste facto (Kreiman & Gerratt, 2011). Porém, esta aparente convergência de meios de caracterização vocal pode também ser vista como uma permanente dificuldade na resposta a questões fundamentais relativas à seleção e interpretação das escalas.

Até à data não existe uma escala universal, pelo que para a realização da dissertação a autora selecionou para estudo os parâmetros GRBAS (*Grade, Roughness, Breathiness, Astheny* e *Strain*) para classificar as vozes-alvo usadas ao longo da sua elaboração. As principais razões de suporte a esta decisão são o facto da qualidade da voz ser, essencialmente, um fenómeno áudio-percetivo (Kreiman & Gerratt, 2011; Shrivastav, 2011) e a escala adotada apresentar valores de validade de constructo moderadas a altas, ser de fácil uso (Gould et al., 2012) e do conhecimento/aplicação profissional na amostra de juízes usada.

2.1.2. O PAINEL DE JUÍZES/AVALIADORES ÁUDIO-PERCETIVOS:

Na clínica quotidiana espera-se que qualquer protocolo de avaliação seja simples, reproduzível e prático. Os resultados da avaliação áudio percetiva servem para caracterizar a voz do falante (com ou sem alteração vocal), para termo de comparação (após a terapia e/ou plano de aperfeiçoamento vocal), como facilitador da troca de informações entre os profissionais da equipa e com o próprio falante (Carding et al., 2000; Awan & Roy, 2009; Shrivastav, 2011).

Porém, existirá sempre um teor subjetivo (nem que seja inconscientemente) das classificações realizadas. Isto porque estas não correspondem diretamente à percepção auditiva do juiz/avaliador, já que estão subjacentes a dois processos distintos: primeiro o estímulo vocal evoca uma sensação, a qual corresponderá a uma decisão. Ou seja, a avaliação é uma medida indireta da percepção (Shrivastav et al., 2005), pelo que a variabilidade pode estar associada a diferenças na percepção ou na tomada de decisão sobre a sensação auditiva.

Kent (1996) sugere que o uso das escalas percetivas está sujeito a erros e variabilidade porque:

- a) As escalas usadas no contexto clínico e de investigação, por vezes, são impróprias para medir os atributos da qualidade vocal, nomeadamente pelas características da magnitude dos atributos a avaliar;
- b) Os avaliadores nem sempre têm o mesmo entendimento sobre os parâmetros – resolução da escala de medida;
- c) Os avaliadores nem sempre conseguem discriminar uma só dimensão da escala num estímulo sonoro complexo, pois é difícil isolar atributos individuais no padrão complexo do sinal vocal;
- d) Os avaliadores têm dificuldade em manter a consistência das classificações, intra e interjuízes, devido à flutuação das referências internas inerentes a diferentes parâmetros vocais (Kreiman & Gerratt, 2011).

O modelo matemático que modeliza a variabilidade da avaliação percetiva foi postulado por Thurstone (1927 *in* Shrivastav et al., 2005) como a “lei dos julgamentos comparativos”, que considerava que quando um estímulo físico é aplicado a um observador, desencadeia um “processo de discriminação”. O mesmo estímulo aplicado a um avaliador em dois momentos distintos pode pressupor dois processos de discriminação diferentes. O autor ainda postula que este processo segue uma distribuição normal, logo, é possível estabelecer modelos matemáticos para caracterizar a resposta dos juízes aos estímulos vocais. Esta forma de descrição da avaliação áudio-percetiva pode ser também considerada uma consequência direta da lei dos grandes números, que diz que a probabilidade de um resultado se vai aproximar da frequência relativa da sua ocorrência na natureza, se uma mesma experiência for repetida inúmeras vezes, sob as mesmas condições.

A constituição de painéis de juízes para a avaliação áudio percetiva visa reunir um conjunto próximo de opiniões, que reduzam a subjetividade (i.e. os erros de medida) deste tipo de avaliação, através do controle dos fatores associados ao avaliador; à tarefa de avaliação; ou características resultantes da interação dos dois parâmetros

anteriores (Eadie & Baylor, 2006; Kreiman et al, 2007b). Assim, os constituintes do painel podem ser de profissões distintas, que audio-percetivamente escutam e diferenciam amostras de vozes para, por último, as classificarem. Se a sua constituição for cuidada permite-se, deste modo, um conjunto de dados mais fiáveis e reproduutíveis acerca da qualidade vocal em análise (Kreiman et al., 1990, 1992, 1993; de Krom, 1994; De Bodt et al., 1997; Eadie & Baylor, 2006; Eadie et al., 2010a).

a. ERROS DE MEDIDA POR FATORES ASSOCIADOS AO AVALIADOR

Estes fatores incluem as referências internas do avaliador relativamente à qualidade vocal a ser percecionada, aos seus hábitos percetivos e erros/distorções próprias e a sensibilidade geral face à qualidade de vozes a escutar. O treino e a exposição a uma vasta panóplia de alterações vocais ajudam a modelar estes fatores (Eadie & Baylor, 2006; Sofranko & Prosek, 2012). Os antecedentes de preparação musical também podem afetar a consistência intra-avaliador na classificação de vozes disfónicas, especificamente do parâmetro soprosidade (Eadie et al., 2010a). Nesta categoria estão ainda incluídos o cansaço do avaliador, os lapsos de atenção e os erros de transcrição, assumidos como erros aleatórios.

b. ERROS DE MEDIDA POR FATORES ASSOCIADOS À TAREFA DE AVALIAÇÃO

Estes dizem respeito a características da própria avaliação como por exemplo uma má definição dos parâmetros a avaliar pela escala ou a falta de proximidade com a realidade percetiva (Eadie et al., 2010a). Inclui-se ainda o contexto percetivo, pois pode alterar a referência interna do avaliador (Kreiman et al., 1992). Por exemplo, ao escutar uma voz classificada como moderada após várias amostras de vozes ligeiramente desviadas, então o avaliador poderá tender a agravar a sua percepção da voz moderada e a alterar a sua referência.

c. ERROS DE MEDIDA POR FATORES ASSOCIADOS À INTERAÇÃO ENTRE O AVALIADOR E A TAREFA DE AVALIAÇÃO

A sensibilidade do avaliador pode interagir com a resolução da escala usada, no sentido de que a sua percepção face a um determinado parâmetro (e.g. Aspereza) possa ser mais fácil, por ter contactado com mais casos com essa característica evidente (Gould et al., 2012). Contudo, o avaliador pode não estar tão sensível a outros parâmetros (Eadie et al., 2010 (a)).

Outro erro está associado à especificidade da escala usada. Isto é, se a escala é multidimensional e está descrita de modo unidimensional, os avaliadores podem focar-se apenas num parâmetro individual, reduzindo os valores de consistência interavaliadores (Kreiman et al., 1993; Kreiman & Gerratt, 1998; Kreiman & Gerratt, 2000).

Dois júris podem ainda divergir no modo como usam os diferentes níveis de uma mesma escala. Ou seja, um pode usar apenas os valores centrais de uma escala de 0 a 4, enquanto outro usa toda a gama de cotação possível (Eadie & Baylor, 2006; Gould et al., 2012).

Deste modo, a constituição de um painel de juízes pode depender da experiência dos avaliadores (Kreiman et al., 1990; Anders et al., 1988), as suas estratégias de escuta e classificação específicas (Kreiman et al., 1992) ou da coerência e homogeneidade de treino do grupo (Shrivastav et al., 2005; Eadie & Baylor, 2006; Patel & Shrivastav, 2007; Maryn et al., 2008; Oates, 2009). Note-se que a experiência do juiz em avaliação de vozes pode ter pressuposto o desenvolvimento de um referencial prévio e de consistência crescente na graduação da qualidade vocal (Kreiman et al., 1993; Eadie & Baylor, 2006); e/ou a sua associação a sensações cinestésicas (ou físicas); e/ou a capacidade de rapidamente correlacionar características psicofísicas com o funcionamento do aparelho fonador, competências que um avaliador inexperiente não conseguiu ainda construir (Moses, 1954 *in* Pouchoulin, 2008).

Existem alguns trabalhos de investigação que refletem acerca das características de um bom grupo de juízes e cuidados a ter para a sua constituição (Kreiman et al., 1992; Gerratt et al., 1993; Wolfe et al., 2000; Shrivastav et al., 2005; Eadie & Baylor, 2006; Kreiman et al., 2007 (a,b); Patel & Shrivastav, 2007). Note-se que há estudos que apontam para resultados contraditórios e ratificam que a inconsistência interavaliadores é muitas vezes o resultado dos métodos usados para a recolha da avaliação e não inerente a problemas nas capacidades percetivas e inconsistência entre os juízes (Kreiman & Gerratt, 2000; Kreiman et al., 2007; Patel & Shrivastav, 2007; Kreiman & Gerrat, 2010; Kreiman & Gerratt, 2011). Anders et al. (1988) referem que a experiência na temática não influencia a avaliação, uma vez que encontrou diferenças pequenas (mas não significativas) entre grupos de avaliadores com experiência *versus* com treino em avaliação áudio percetiva. Dejonckere et al. (1993) concluíram que a escala GRBAS obtinha valores mais consistentes inter avaliadores com experiência, principalmente o parâmetro G(grau) ($p=0.7$), em oposição com o A(astenia) e S(tensão), ambos não consistentes. Eadie & Baylor (2006) encontraram esta tendência de maior consistência inter avaliadores face ao parâmetro G(grau), tanto em tarefas de vogal sustentada ($r=0.905$) como de leitura ($r=0.922$), e concluíram que aumentava ligeiramente com duas horas de treino percetivo ($r=0.938$ e $r=0.961$, para as tarefas indicadas, respetivamente). Dejonckere et al. (1993) demonstraram que a escala GRBAS obtinha adequados valores de consistência intra e interjuízes, embora mais fortes para o parâmetro já referido. De Bodt et al. (1997) encontraram valores de consistência teste-reteste moderados ($k=0.43$), na avaliação dos 23 juízes (12 experientes e 11 não experientes) que usaram a escala GRBAS, sem diferenças estatisticamente significativas entre as medianas das classificações dos avaliadores de diferentes profissões/experiência de avaliação percetiva. Estes autores concluíram ainda que os parâmetros G, R e B são os menos ambíguos, com valores de teste-reteste entre 0.5 e 0.7, embora salientem que apenas usaram vozes severamente alteradas. Heman-Ackah et al. (2002) realizaram um estudo com 2 juízes que classificaram com a escala GRBAS a voz de 38 casos com paralisia da prega vocal. Defrontaram-se com valores mais consistentes da correlação bivariada de Pearson para os parâmetros Grau ($r=0.828$, $p=0.003$), Rugosidade ($r=0.718$, $p=0.019$) e Soprosidade ($r=0.939$, $p<0.001$). Martens et al. (2007) encontraram um aumento

significativo da consistência inter avaliadores com o uso de pistas (ou “âncoras”) visuais de apoio à avaliação (espectrograma), para os parâmetros G-Grau geral de alteração vocal ($p<0.05$), R-Rugosidade ($p<0.001$) e B-Soprosidade ($p<0.001$).

Devido à fraca reproduzibilidade dos parâmetros Instabilidade, Tensão e Astenia em diversos estudos enumerados (Dejonckere et al., 1993; De Bodt et al., 1997; Oates, 2009; Gould et al., 2012; Moers et al., 2012), estes foram retirados do protocolo de avaliação da função vocal da *European Laryngological Society* (Dejonckere et al., 2001) e do protocolo de avaliação vocal Alemão (*German RBH evaluation scheme*) (Nawka et al., 1994).

Kreiman et al. (1990) demonstraram que os avaliadores com e sem experiência/treino usam estratégias de classificação distintas. Neste estudo, o grupo de juízes sem experiência apresentou menor acordo sobre os parâmetros de qualidade vocal. Hammarberg (1986) concluiu que os juízes sem experiência de avaliação tendem a apoiar a decisão em aspectos dominantes da qualidade vocal, mais evidentes da amostra, como o *pitch* e *loudness*. Bunton et al. (2007) compararam a classificação de 47 amostras de fala disártica (com a duração de 40seg.) – por dois tipos de avaliadores, experientes (10) /não experientes (10) – através de uma escala específica para esta patologia composta por 38 parâmetros, em 7 níveis distintos, e não encontraram diferenças entre os dois grupos de juízes. Bele (2005) comparou a consistência da avaliação de 71 vozes normais – em tarefa de vogal sustentada e leitura – caracterizadas em 15 parâmetros, por 7 avaliadores experientes e 3 não-experientes, concluindo que o primeiro grupo tendeu a ser mais consistente.

Pouchoulin (2008) ratifica que a constituição de um júri de escuta não pode ser encarada como uma tarefa trivial, já que pressupõe um conjunto de cuidados para evitar erros que atuam sobre o resultado da avaliação. Refere, por exemplo, considerações acerca das referências – culturais e clínicas – dos avaliadores; a sua formação académica de base; a duração das amostras a escutar; o número de sessões para realizar os estudos de investigação; a forma como se apresentam as amostras,

entre outras variáveis de confusão, inerentes a fatores intrínsecos e/ou extrínsecos ao avaliador.

Acrescentem-se, então, fatores estudados como: atenção/concentração, fadiga, memória de estímulos vocais prévios, treino e experiência anterior com a escala/tarefa de avaliação (Poulton, 1989 *in* Shrivastav et al., 2005). Martin & Wolfe (1996) estudaram os efeitos do treino na capacidade de avaliadores inexperientes classificarem os parâmetros de Rouquidão, Aspereza e Soprosidade em vozes patológicas. Concluíram que houve uma melhoria após treino, contudo 40% das amostras ainda eram classificadas incorretamente. Bassich & Ludlow (1986) concluíram com o seu estudo que seriam necessárias 8 horas de treino a avaliadores não treinados para conseguir um valor de 80% de consistência interavaliadores, no uso de uma escala com 13 parâmetros. Os autores sugerem que este tempo de treino poderia ter sido reduzido se tivessem sido proporcionados “estímulos-âncora” e usada conversação espontânea (em oposição à vogal sustentada). Ma & Yiu (2006) apresentam valores de consistência inter avaliadores entre 0.86 e 0.91 ($p=0.0001$) na classificação do G (grau geral de alteração vocal) por juízes que tiveram treino prévio e beneficiavam de “estímulos-âncora” com vozes sintéticas ao longo do teste. Chan & Yiu (2006) reiteraram que avaliadores classificam o parâmetro Soprosidade em amostras de vozes sintéticas, com 80% de consistência, se lhes forem proporcionados “estímulos-âncora” e com o uso do método de comparação emparelhada. Com esta metodologia de estudo são apenas necessárias 2 horas de treino. Awan & Roy (2009) realizaram um estudo com 176 vozes, representativas de diferentes graus de alteração vocal, em que o painel de 10 juízes apenas beneficiou de uma sessão de 20 minutos para apresentação do estudo, material de classificação percetiva (*software* informático específico) e definição de conceitos (e.g. G=grau geral de severidade da alteração vocal). Tinham acesso a um “estímulo-âncora” auditivo e obtiveram um forte grau médio de correlação interavaliadores (r de Pearson=0.83) e de acordo interavaliadores (média de acordo=81.03%).

Os efeitos do treino podem ainda afetar as pistas acústicas e visuais usadas pelos avaliadores para realizarem a avaliação percetiva (Shewell, 1998; Chan & Yiu, 2002; Yiu

et al., 2002; Chan & Yiu, 2006; Awan & Lawson, 2009; Chan, 2011; Gould et al., 2012). Martin & Wolfe (1996) concluíram que os ouvintes classificam as vozes principalmente com base na frequência fundamental (F0) antes do treino, contudo esta tarefa foi também influenciada após o treino pelas medidas de ruído (HNR ou SNR).

A análise áudio-percetiva apresenta aspectos ainda desconhecidos e com grande variabilidade. Por isso é que avaliadores não experientes concordam acerca das dimensões da qualidade vocal que pareçam mais evidentes/importantes, tanto para vozes patológicas como normais; já os ouvintes experientes/treinados tendem a não concordar tão facilmente (Kreiman et al., 1990; Kreiman et al., 2007; Patel & Shrivastav, 2007; Iwarsson & Petersen, 2012). Do mesmo modo, as discordâncias podem ser o reflexo das diferenças de treino e experiência profissional (Wolfe & Martin, 1990; Kreiman et al., 1990; Kreiman & Gerratt, 1998; Bele, 2005; Sofranko & Prosek, 2012). O sistema de referência interno dos avaliadores é idiossincrático, e varia intra e interjuízes, de acordo com a experiência prévia e o contexto acústico no qual as análises são levadas a cabo (Kreiman et al., 2007; Choi et al., 2012; Iwarsson & Petersen, 2012; Gould et al., 2012).

Kreiman et al. (2007) estudaram a influência de alguns destes fatores na variabilidade de classificações intra e interavaliadores, através da manipulação dos mesmos em investigações experimentais, tal como apresentado na Tabela 4.

<i>Razão teórica de Variabilidade</i>	<i>Fator experimental</i>
Magnitude do atributo a medir	O significado teórico de qualidade vocal
Resolução da escala de medida	Escala Visual Analógica <i>versus</i> Escala Intervalar
Dificuldade em isolar atributos individuais no padrão complexo do sinal vocal	Estímulos de comparação emparelhados <i>versus</i> não emparelhados
Instabilidade dos <i>standards</i> internos inerentes a diferentes qualidades vocais	Presença <i>versus</i> ausência de estímulos de comparação

Tabela 4: Fontes da variabilidade intra e interavaliadores áudio-percetivos.

Concluíram que estes quatro fatores eram responsáveis por 84,2% da variância da probabilidade de acordo na classificação dos juízes. Tal como postulado pelos

investigadores, a probabilidade de qualquer dupla de avaliadores concordar nas suas avaliações de uma dada voz é quase perfeita (96% de acordo) quando todos os fatores são controlados. Sugerem que um protocolo de classificação da qualidade vocal deveria:

- a) Não depender da seleção/definição de parâmetros de qualidade da voz;
- b) Evitar a dependência de referências internas e ajudar os avaliadores a focar a atenção áudio-percetiva;
- c) Ter uma escala de medida com resolução adequada (Kreiman & Gerratt, 2011).

Os estudos sugerem que a variabilidade interavaliador não é aleatória mas altamente previsível e evitável através do uso de instrumentos de medida adequados. Aliás, parece contraintuitivo que a capacidade áudio-percetiva incorpore uma grande variabilidade aleatória, já que a qualidade vocal na evolução filogenética da espécie humana e a importância dos sistemas de signos verbais-orais na comunicação em comunidade (e.g. a deteção do género, da idade, do estado emocional, do papel social, personalidade, entre outros) reforçam que esta competência apresenta uma elevada eficiência, pelo que parece mais lógico que os valores de consistência pobre derivem dos protocolos adotados e não das diferenças de percepção dos ouvintes (Kreiman & Sidtis, 2011; Kreiman & Gerratt, 2011; Balasubramanium et al., 2012).

Em suma, a análise percetiva beneficia em termos de reproduzibilidade e consistência inter avaliadores com a constituição de grupos de classificação que sejam especialistas nesta tarefa, realizando-a em múltiplas sessões de escuta (Shewell, 1998; Chan e Yiu, 2006), com recurso a “estímulos-âncora” ou a métodos de emparelhamento comparativo (e.g. análise-por-síntese) e *feedback* auditivo, como forma de recalibração frequente das referências internas dos avaliadores (Gerratt et al., 1993; Chan & Yiu, 2002; Eadie & Baylor, 2006; Kreiman et al., 2007; Awan & Lawson, 2009; Brasolotto & Rehder, 2011; Gould et al., 2012; Iwarsson & Petersen, 2012; Sofranko & Prosek, 2012).

Porém, a aplicabilidade prática de protocolos de avaliação da qualidade vocal mais complexos – devido a limitações financeiras, de disponibilidade temporal, entre outras

– é muitas vezes condicionada. Reforça-se ainda a necessidade de um planeamento metodológico (experimental) e estatístico adequado das investigações sobre avaliação áudio percetiva da voz (Kreiman & Gerratt, 2011), com várias avaliações de cada estímulo para cada ouvinte e escalas padronizadas de avaliação (Webb et al., 2004; Brasolotto & Rehder, 2011), por forma a conseguir um grau elevado de consistência interavaliadores.

2.2. REPRODUTIBILIDADE DOS RESULTADOS DA AVALIAÇÃO ÁUDIO-PERCETIVA

O uso clínico de um determinado teste ou escala depende da reprodutibilidade dos seus parâmetros. Este aspeto depende de três fatores:

- i. A variabilidade interobservador, isto é, a variação de julgamentos entre diferentes avaliadores/terapeutas.
- ii. A variabilidade intraobservador, ou seja, a variação de julgamentos de um mesmo avaliador/terapeuta ao longo do tempo.
- iii. A variabilidade intrassujeitos, entendida como a variação da qualidade vocal do falante em diferentes momentos temporais da recolha e análise da mesma. Este aspeto pode estar associado a fatores intrínsecos (emocionais, cansaço) ou extrínsecos, que se não forem controlados condicionam os resultados da avaliação áudio-percetiva.

Esta secção resulta de uma pesquisa na base de dados bibliográfica Pubmed que possibilitou a revisão da literatura publicada (1950-2/julho/2012) sobre os resultados da avaliação áudio-percetiva, formas de uso e respetiva validade dos resultados. Os termos de pesquisa usados foram (voice quality OR voice disorders) AND (reproducibility of results OR retest OR variability OR variation OR repeatability OR intra-subject OR intrasubject).

No motor de busca da Pubmed as pesquisas podem ser filtradas (“translated”) através de uma opção (“query translation”). Aqui os termos a procurar podem ser extendidos, o que resultou em (“voice disorders” [MeSH Terms] OR “voice disorders” [All Fields]) OR (“voice quality” [MeSH Terms] OR “voice quality” [All Fields]) AND (“reproducibility of results” [MeSH Terms] OR “reproducibility” [All Fields]) OR “reproducibility of results” [All Fields] OR “retest” [All Fields] OR “variability” [All Fields] OR “repeatability” [All Fields] OR “intrasubject” [All Fields] OR “intra-subject” [All Fields]).

Foram estabelecidos os seguintes critérios de inclusão:

- Escritos em inglês e publicados até 2 de julho/2012.
- Relativos a estudos com a qualidade vocal de humanos (objetiva e subjetiva).
- Artigos com referência a teste-reteste. Este tipo de estudo é definido com a aplicação múltipla (duas ou mais vezes) de um mesmo teste em dois momentos temporais distintos, a uma dada amostra.

Número total de artigos encontrados: 108

Total de artigos sobre avaliação áudio percetiva: 43

No total foram encontradas 43 referências publicadas ao longo de 25 anos (1986-2012) sobre a validade da avaliação áudio-percetiva, enquadráveis nos critérios de inclusão da pesquisa levada a cabo. Note-se que em nenhum dos artigos foi caracterizada a variabilidade intrassujeito. Todos faziam referência à validade interobservador e em 32 focavam-se os resultados da validade intraobservador. Para melhorar estes dois aspectos saliente-se que, em 10 dos estudos analisados, as amostras de voz foram caracterizadas em comparação com outra. Por vezes, esta era uma amostra de referência, considerada como uma “amostra âncora”. Podia provir de vozes naturais (Chan & Yiu, 2002; Eadie & Doyle, 2002; Yiu & Ng, 2004; Moers et al., 2012) pré-selecionadas, ou vozes sintetizadas (Chan & Yiu, 2002; Granqvist, 2003; Kreiman et al., 2007; Yiu et al., 2007). Em alguns estudos as diferentes amostras eram comparadas entre elas (Granqvist, 2003; Kreiman et al., 1994; Shrivastav, 2006; Wolfe et al., 2000; Yamauchi et al., 2010), e numa das investigações as comparações foram realizadas entre duas amostras do mesmo sujeito (Webb et al., 1992).

Foram usadas várias escalas, sendo que as mais referenciadas foram já previamente descritas nesta dissertação. A GRBAS está presente em 26 dos artigos. A classificação através de escalas categóricas é usada em 21 estudos (Murry et al., 2004; Martens et al., 2007; Webb et al., 2004; Lee et al., 2005; Karmell et al., 2007; Feijoo & Hernandez, 1990; De Bodt et al., 1997; Yamauchi et al., 2010). É feita referência à escala analógica visual em 11 artigos (Chan & Yiu, 2002; Yiu & Ng, 2004; Granqvist, 2003; Kreiman et al., 1994; Kreiman & Gerratt, 1998; Kreiman et al., 1993; Dejonckere et al., 1996; Langeveld et al., 2000; Ma & Yiu, 2001; Morsomme et al., 2001; Moers et al., 2012).

Em 19 das investigações analisadas todos os casos possuíam uma alteração vocal; em 20 artigos participaram indivíduos com/sem perturbações da voz. Em apenas um estudo foram analisados falantes sem características vocais alteradas (Bele, 2005). Em dois dos artigos estas amostras eram de vozes sintetizadas (Titze et al., 2003; Bergan et al., 2004).

Existe uma variabilidade marcada entre os instrumentos escolhidos (encontramos 13 escalas diferentes, no total). Em alguns casos foi desenvolvido e implementado um novo protocolo de avaliação áudio-percetiva, ajustado a “clusters” específicos de pacientes (por exemplo, a disfonia espasmódica⁸, a disartria⁹ ou a nasalidade) (Langeveld et al., 2000; Erickson, 2003; Sweeney e Sell, 2008).

Em síntese, para além da variabilidade interobservador, intraobservador e intrassujeito, a reprodutibilidade da avaliação áudio percetiva pode ainda ser influenciada pelo sistema de avaliação usado, a forma de cotação adotada e a amostra vocal escutada (leitura, fala espontânea, vogal sustentada ou uma combinação das opções anteriores).

Na literatura descrita e revista previamente são várias as escalas disponíveis, com distintos critérios de recolha, análise e classificação dos dados. A comparação entre as mesmas não é, portanto, viável, embora existam dois estudos que o tenham feito (Jesus et al., 2009; Yamauchi et al., 2010).

A maioria dos estudos que usa avaliação áudio percetiva possui validade interobservador e intraobservador, de moderada-boa. Por vezes são feitas tentativas de aumento destes valores através da otimização dos sistemas de cotação. Uma vez que cada observador possui o seu *gold standard* interno de qualidade vocal, é

⁸ Quadros de alterações da voz, mais comuns em mulheres, e na terceira década de vida, sem qualquer evidência de lesão neurológica, com diferentes naturezas etiológicas (Sapir, 1995; Behlau & Pontes, 1997; Behlau, 2001; Murdoch, 2005; Pinho e Pontes, 2008).

⁹ Acometimento da voz, articulação, ressonância, prosódia e até da respiração decorrente de uma afecção congénita ou adquirida do sistema nervoso central (Murdoch, 2005; Pinho et al., 2006; Guimarães, 2007; Pinho & Pontes, 2008).

compreensível que os investigadores pretendam a sua substituição por um referencial externo (uma amostra de voz de referência como padrão para a classificação). Aqueles que usam este tipo de referencial concluem um aumento da validade (Gerratt et al., 1993; Chan e Yiu, 2002; Yiu et al., 2002; Eadie & Baylor, 2006; Awan & Lawson 2009).

2.3. A AVALIAÇÃO ACÚSTICA

A avaliação acústica – quando utilizada no âmbito da avaliação e intervenção terapêutica nas perturbações vocais – permite quantificar, de modo não invasivo, as características da voz humana (Kent & Ball, 2000; Behlau, 2001; Pinho, 2003; Pinho et al., 2006; Awan & Roy, 2009; Ma & Yiu, 2011; Choi et al., 2012). O método acústico propicia as designadas medidas objetivas, ou seja, dados que são extraídos automaticamente, ou não, através de um processamento computacional adequado. Aquele possibilita a integração de dados fornecidos pela avaliação áudio-percetiva com o plano fisiológico, pois permite a especificação detalhada do processo de geração do sinal sonoro: fornece, então, uma estimativa indireta dos padrões vibratórios das pregas vocais, bem como dos formatos do trato vocal supraglótico e das respetivas modificações (Hirano & Bless, 1997; Vogel, 2011). Para tal, recorre a diferentes parâmetros acústicos que compõem o sinal – periodicidade, amplitude, duração e composição espectral (Murdoch, 2005; Guimarães, 2007) – que caracterizam os atributos físicos da voz nos domínios do tempo, da frequência e da intensidade, para além de outras medidas complexas que conjugam o cruzamento daqueles domínios (Buder, 2000; Camargo, 2000; Vogel, 2011; Madill & McCabe, 2011; Ma, 2011).

O século XX foi um marco na implementação do método acústico como forma de avaliação da voz. Durante a década de 20 foram desenvolvidos inúmeros estudos com a análise das características vocais suportadas nos resultados dos oscilogramas e as representações gráficas resultantes – que possibilitam a correlação da amplitude do som em função do tempo. Na década de 40 teve início a utilização da transformada de Fourier para obter a análise espectral. O grande avanço nesta tecnologia ocorreu na década de 50, com o aparecimento do espectrógrafo que fornece o spectrograma – onde é possível observar a variação da concentração de energia acústica ao longo do tempo (Araújo, 1999) – facultando o registo tridimensional do sinal sonoro, que integra os aspetos do tempo (já referido), a frequência e a intensidade num único gráfico de dois eixos (Kent e Read, 1992). Mais tarde Gunnar Fant (1970) postulou a teoria da fonte-filtro linear (ou modelo de fonte-filtro) que explica matematicamente através de moldes simplificados a produção da voz e da fala (Johnson, 2003). O modelo

fonte-filtro pressupõe três principais sub-sistemas acústicos para a produção de voz: de modulação do impulso glótico; de caracterização do trato vocal e da radiação labial (Kent, 1993, 1997). Na Figura 2 estão ilustrados (à esquerda) o modelo fisiológico dos sistemas envolvidos na produção de voz, assim como a correspondência com o modelo linear de fonte-filtro proposto por Fant.

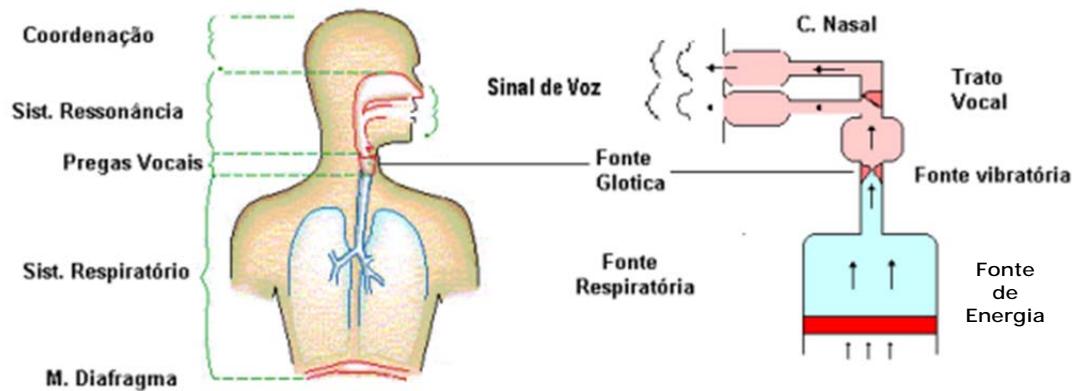


Figura 2: Desenho esquemático dos sistemas e órgãos envolvidos na produção do sinal de voz e do modelo linear de fonte-filtro proposto por Fant (1970) (adaptado de Dajer, 2010).

Sumariamente, o som gerado nas pregas vocais é chamado de fonte vibratória ou vocal. Esta é semelhante para todos os sons vocálicos produzidos na mesma frequência e intensidade fonatória (Sundberg, 1987). Porém, quando se ouvem as vogais perceciona-se uma enorme distinção entre elas, mesmo sendo de igual intensidade ou frequência fundamental. Deste modo, os sons das vogais não são consequência apenas da fonte glótica. Esta sofre algum tipo de mudança que contribui para o produto final escutado. O responsável por este fenômeno é o trato vocal, que atua como ressoador, modificando o seu formato de modo a conceber padrões particulares de ressonância, caracterizados pela função de transferência ou, mais especificamente, resposta em frequência.

Assim, atendendo a que a fonte vocal produz um som que é composto por uma série de frequências, o trato vocal amplia ou atenua seletivamente (i.e. filtra) as compatíveis com o seu formato, as quais são radiadas (efeito de radiação) pela boca e cavidade nasal (Sundberg, 1987; Kent e Read, 1992; Titze, 1994). Estas frequências de

ressonância no trato vocal designam-se por formantes. Geralmente existem quatro ou cinco mais relevantes.

No processo de radiação labial ocorre uma nova filtragem do tipo passa-alto (Kent e Read, 1992), o que se associa ainda a uma tendência acústica de as altas frequências se propagarem em linha reta, enquanto as baixas frequências se espalham (difratam).

As décadas de 60 e 70 marcaram o desenvolvimento do processamento digital de sinais, com a divulgação dos primeiros programas de *software* de análise da voz, com definições mais precisas e claras das medidas a usar. As técnicas de processamento de sinal permitem recolher e caracterizar as particularidades de vibração das pregas vocais (Pereira e Montagnoli, 1999; Kent & Ball, 2000). O processamento digital do sinal permite a análise, transformação ou interpretação de sinais através de algoritmos computacionais incluindo a FFT (*Fast Fourier Transform*), a LPC (*Linear Predictive Coding*), ou técnicas de filtragem e *Cepstrum* (Davis, 1979; Read et al., 1992; Childers et al., 1997; Araújo, 1999; Pereira e Montagnoli, 1999; Baken & Orlikoff, 2000). Deste modo, as medidas obtidas na análise acústica correspondem a parâmetros físicos definidos. O sinal glótico, sinal da fonte, sofre efeitos ao longo do trato vocal supraglótico até à saída deste para o meio externo, o que configura uma ação de filtro (*Cf. Figura 2*). Acontece um somatório das ondas sonoras provenientes da fonte glótica com outras refletidas ao longo do trato vocal, sendo o resultado final (sinal de saída) o sinal irradiado pelos lábios (Fant, 1970; Kent & Read, 1992; Coleman, 1960 e Lieberman, 1961 *in* Buder, 2000).

A partir da década de 90 é que surgiram, em Portugal, as primeiras soluções de *software* e *hardware* para medição e análise do sinal sonoro. Atualmente as propostas – marcas e equipamentos – disponíveis no mercado são inúmeras. Contudo, a sua seleção e uso está dependente das características de recolha de dados, edição, análise e interpretação dos mesmos. É ainda de acrescentar o gasto monetário envolvido e a futura rentabilização do material.

Sabendo que na prática profissional do terapeuta da fala em Portugal este equipamento é cada vez mais habitual (Sousa, Vaz-Freitas e Ferreira, 2011), salientam-se de seguida os seus principais objetivos e vantagens de utilização (Buder, 2000; Behlau, 2001; Barros & Carrara-De Angelis, 2002; Murdoch, 2005; Pinho et al., 2006; Camargo & Madureira, 2010):

- a) Oferece uma maior compreensão acústica do *output* vocal e aproxima formas distintas de avaliação da voz, nomeadamente a análise áudio-percetiva e a acústica ou a laringo-estroboscópica e a acústica;
- b) Proporciona – de modo expedito e de uso fácil – dados normativos para realidades vocais distintas (culturais, profissionais e/ou patológicas);
- c) Propicia informação importante sobre o impacto do sinal vocal no ouvinte;
- d) Oferece a documentação – gráfica e numérica – necessária para descrever a qualidade vocal de um indivíduo, seja ele um utilizador profissional da voz ou um paciente em tratamento, por disfonia, auxiliando e ratificando pareceres judiciais ou outros atestados com caráter legal;
- e) Proporciona imagens e gráficos de análises acústicas, com fácil compreensão por parte do paciente/falante em avaliação ou acompanhamento terapêutico, favorecendo um melhor prognóstico associado ao maior envolvimento e consequente motivação para o processo de mudança vocal;
- f) Monitoriza a eficácia de um tratamento e permite comparar resultados vocais de diferentes metodologias de intervenção, em fases distintas do processo terapêutico ou cirúrgico/medicamentoso;
- g) Acompanha o desenvolvimento de uma voz profissional, e orienta a sua adequação ao longo do tempo, inclusive com a possibilidade de sistemas de feedback-análise acústica em tempo-real;
- h) Assume-se como um instrumento de deteção precoce de problemas vocais e laríngeos, por exemplo em campanhas de triagem, pela deteção de níveis de perturbação fonatória acima dos valores de referência.

Os procedimentos (medidas) possíveis dependem do *software* usado pelo terapeuta da fala. Existem vários dispositivos na Internet e no mercado, que diferem em termos

de velocidade, compatibilidade, mensurações disponíveis e custo associado (ver Tabela 5).

<i>Opensource Software</i>	<i>Comerciais</i>
Audacity 1.3.12 (Audacity Team)	CSL, model 4500 (Kay Elemetrics, 1996)
AMPEX.DVA – Disordered Voice Analyser V.2 (Ghent University – DSSP-ELIS, 2010)	Dr. Speech 4.0 (Tiger Electronics, 1999)
BioVoice V.1 (Università degli Studi di Firenze – DET, 2008)	IVANS (Avaaz Innovation, 2000)
EMU Speech Data System (LMU, 2009)	Lx Speech Studio, SPEAD – Speech Pattern Element Acquisition and Dysplay (Laryngograph)
Praat (Boersma & Weenink, 2003)	MDVP, model 5105 (Kay Elemetrics, 1996)
SFS (Speech Filing System, 1999)	MediVoz (TGH Endoscopia SL.)
Speech Analyser (SIL International, 2007)	Speech Station2 (Sensimetrics)
Sound Forge 10.0b (Sony Creative Software Inc., 2010)	Speech Studio (Laryngograph Ltd.)
Sky Voice Analysis Program (UCLA, Kreiman et al., 2006)	TF32 (Milenkovic, 2003) e CSpeech Software (Milenkovic & Read, 1987)
Wasp (Waveform Annotations Spectrograms and Pitch)	VoiceStudio (Seignal, 2008)
Wavesurfer (Sjolander & Beskow, 2003)	Voxmetria (CTS Informática, 2003)

Tabela 5: Principais programas de *software* de análise acústica de voz referidos na literatura

A fiabilidade da informação recolhida e analisada é passível da influência de um conjunto de variáveis de confusão, por vezes difíceis de controlar no contexto clínico, salientando a sua dependência das condições e procedimentos de captação, armazenamento, edição e análise do sinal sonoro (Buder, 2000; Guimarães, 2007; Oates, 2009; Vogel, 2011; Madill & McCabe, 2011). Estes fatores de interação são críticos porque os parâmetros acústicos refletem a complexa interação entre a fonte glótica e as cavidades de ressonância do trato vocal. Assim, dependem das forças biomecânicas e aerodinâmicas da laringe e estruturas supraglóticas (Mendes et al., 2003; Olszewski et al., 2011; Akpinar et al., 2012), assim como do complexo controlo neuro-motor cortical. Se estas componentes apresentam características anatómicas

e/ou fisiológicas anormais, então os resultados obtidos vão ser desviantes face ao esperado e, assim, assumem-se como indicadores de patologia vocal e da sua respetiva severidade (Giovanni et al., 1999; Landini et al., 2002; Zhang et al., 2005a,b; Jiang et al., 2006; Rahn et al., 2007; Zhang & Jiang, 2008; MacCallum et al., 2009). Em suma, as medidas objetivas da qualidade vocal (Buder, 2000; Oates, 2009; Vogel, 2011; Moon et al., 2012) são também suscetíveis de variáveis confusionais que incluem: as condições de captação do sinal, as características específicas do sistema de *hardware* e *software*, os protocolos de recolha e análise dos dados, a variabilidade individual em termos acústicos e aerodinâmicos e o grau de severidade e tipo de alteração vocal (Holmberg et al., 1994; Titze, 1995; Wolfe et al., 1995; Hillman et al., 1997; Campisi et al., 2000; Zraick et al., 2000; Carson et al., 2003; Holmberg et al., 2003; Cornwell et al., 2004; Behrman, 2005; Zraick et al., 2005; Deliyski et al., 2006; Ma & Yiu, 2006; Zraick et al., 2006; Kreiman et al., 2007; Martens et al., 2007; Choi et al., 2012; Moers et al., 2012). Neste trabalho consideraremos o sinal de fala como o produto de fontes de voz, ruído e ressonância causados pelo trato vocal, efeitos esses que servem de *input* para o sistema auditivo e para o sistema nervoso central. Portanto, remetem para o conhecimento das bases fisiológicas, acústicas e cognitivas implicadas na produção e percepção da fala (Camargo e Madureira, 2010).

Tradicionalmente, as vogais sustentadas produzidas com níveis confortáveis de amplitude e frequência (Zraick et al., 2000; Awan, 2011; Pinho et al., 2006; Choi et al., 2012; Moon et al., 2012) são usadas como amostra a analisar acusticamente e, assim, caracterizar a qualidade vocal de um indivíduo (Maryn et al., 2009). A justificação para este corpus prende-se com a sua facilidade de compreensão e produção, registo e análise (Hirano, 1988; de Krom, 1994; Parsa & Jamieson, 2001; Martens et al., 2007; Pouchoulin, 2008). Guimarães (2007) e Maryn et al. (2009) acrescem a sua estabilidade, o facto de não conterem variações de entoação e efeitos de coarticulação, ou seja, elimina os fatores de confusão inerentes às influências supraglóticas e linguísticas. Martens et al. (2007) sugerem que seja analisada uma porção relativamente estável da vogal sustentada (Titze, 1995), eliminando os primeiros e últimos 250ms do sinal, o que inclui o *onset* e *offset*. Boucher (2008) e Moon et al. (2012) também suportam os seus estudos neste princípio. Pecam contudo

por não representarem a comunicação verbal e, por isso, moderarem a gravidade de uma disfonia (Bassich & Ludlow 1986; Klingholtz, 1990; Laver et al., 1992; Revis et al., 1999; Yiu et al., 2000; Fourcin, 2000; Bele, 2005; Guimarães, 2007; Parsa e Jamieson, 2001; Revis et al., 2002; Pouchoulin, 2008; Maryn et al., 2009). O prolongamento de vogais pode ser usado para obter informação sobre a duração fonatória, o controle pneumofônico, níveis de *pitch* e *loudness* confortáveis, a presença de alterações e a capacidade de manter uma articulação estável. Realizar variações em escala de frequência para uma mesma vogal pode ser usado para determinar a faixa de conforto e uso tonal habitual do sujeito (Zraick et al., 2000; Ma, 2011). O mesmo raciocínio aplica-se à variação da intensidade vocal (Murdoch, 2005). Note-se que a escolha da vogal afeta as medidas de perturbação (Laver, 1980; Camargo & Madureira, 2010; Moon et al., 2012), tal como discutido ao longo deste capítulo. De Krom (1994) usou a escala analógica visual no seu estudo em que comparou a classificação percetiva em diferentes tipos de estímulos vocais (vogal sustentada *versus* fala encadeada) e concluiu que o *onset* e/ou toda a vogal possuía uma resolução superior, refletindo de modo mais preciso a gravidade da alteração vocal, em especial os parâmetros Rugosidade e Soprosidade. Pelo contrário, Hammarberg (1986) considerou que as vogais sustentadas não permitem uma avaliação percetiva completa e correta das características fonatórias, uma vez que os falantes podem sentir-se pouco confortáveis nesta prova, culminando em dados derivados de um comportamento vocal não natural e, assim, incaracterístico (Klingholz, 1990; Qi & Hillman, 1997; Parsa & Jamieson, 2001).

O uso da leitura de um texto foneticamente equilibrado¹⁰ é sugerido por alguns autores como a prova que no laboratório de voz mais se aproxima do discurso espontâneo (Sakata et al., 1994; Bocklet et al., 2009; Moon et al., 2012). Para o Português Europeu foi realizada a tradução do texto “A História do Rato Artur”, por ser uma leitura equilibrada de prosa, com duração adequada (dois minutos, 335 palavras) e esta ser uma tarefa fonatória com uma consistência teste-reteste elevada, possibilitando a comparação de dois momentos temporais, para um mesmo falante

¹⁰ Entenda-se como aquele que é representativo dos fonemas de uma dada língua, em todas as suas combinações possíveis. (Pouchoulin, 2008)

(Guimarães, 2007). Este tipo de amostra pode ser usado para obter informações para além das exclusivamente fonatórias, tais como: precisão articulatória, consistência e manutenção, influências coarticulatórias, padrões de entoação, velocidade de fala, características e duração, nomeadamente das pausas (Murdoch, 2005; Parsa & Jamieson, 2001; Moon et al., 2012).

O recurso a amostra de discurso espontâneo é, claro, o mais realista (Eadie & Baylor, 2005; Maryn et al., 2009; Dejonckere, 2010; Moers et al., 2012) – em termos fonéticos e prosódicos – embora contenha uma variabilidade marcada, inerente ao contexto comunicativo e características psicoemocionais do sujeito (Bickley & Stevens, 1986; Gobl, 1988; Gobl & Ní Chasaide, 1988; Löfqvist & McGowan, 1991; Rammage et al., 1992; de Krom, 1994). Os testes de fala encadeada são linguisticamente mais complexos, e também usados por muitos investigadores (Hammarberg et al, 1980; Klingholz, 1990; Koike, 1967; Muta et al., 1988; Schoentgen, 1989; de Krom, 1994; Sataloff, 1997; Qi et al., 1999; Heman-Ackah et al., 2002; Halberstam, 2004; Eadie & Doyle, 2005; Eadie & Baylor, 2006; Ma & Yiu, 2006; Moers et al., 2012) já que oferecem informação sobre o desempenho ao longo do tempo, a manutenção/estabilidade das qualidades vocais, a variedade e os efeitos das influências coarticulatórias (Löfqvist & McGowan, 1991; de Krom, 1994; Bele, 2005; Eadie & Baylor, 2005; Murdoch, 2005; Dejonckere, 2007). Sabendo que a fala encadeada pressupõe uma maior complexidade de produção fisiológica comparativamente com as vogais sustentadas, os aspetos de alteração vocal estão mais evidenciados e, como tal, poderão ser melhor classificados percetivamente (Bassich & Ludlow, 1986; Moers et al., 2012). Saliente-se um estudo que argumenta o contrário: de Krom (1994) encontrou uma consistência interavaliadores da classificação áudio-percetiva mais baixa neste tipo de amostra vocal. O autor alega que tal ficou a dever-se a dois fatores: à maior variabilidade vocal ao longo da amostra de fala, o que dificulta a especificidade da classificação; à associação que os avaliadores tendem a realizar com parâmetros extraglóticos (por exemplo, de dialeto ou velocidade do discurso) que podem ser pistas percetivas que distorcem e dificultam a classificação final. Esta complexidade da fala encadeada, ao aplicar o método acústico, pressupõe a análise com técnicas mais sofisticadas (Buder, 2000; Klingholtz, 1990;

Laver et al., 1992 *in* Guimarães, 2007; Moers et al., 2012). Klingholtz (1990) analisou uma bateria de vogais sustentadas e fala encadeada de 50 vozes normais e 74 alteradas, sobre as quais aplicou a medida SNR (*signal-to-noise ratio*) e concluiu que esta quando aplicada em vogal sustentada obtinha 2.5% de taxa de erro de classificação das vozes patológicas, e na fala encadeada esse erro subia para 5.6%. Por seu lado, Moers et al. (2012) analisaram um grupo de 73 vozes disfónicas, em que as medidas cepstrais (*cepstral peak prominence* e *smoothed cepstral peak prominence*) aplicadas a amostras de leitura de texto indicavam, ainda que apenas moderadamente, o grau de alteração da voz. Todas as medidas aplicadas às vogais falhavam este objetivo.

Em suma, a avaliação acústica da voz humana oferece medições válidas mas ainda insuficientes para os clínicos, que a encaram como um resultado complementar de diagnóstico não invasivo (Rabinov et al., 1995; Buder, 2000; Campisi et al., 2000; Awan, 2005; Behrman, 2005; Eadie & Doyle, 2005; Vieira et al., 2005; Ma e Yiu, 2006; Sweeney e Sell, 2008; Dejonckere, 2010; Sousa et al., 2011; Brasolotto & Rehder, 2011) que permite o registo e oferece a possibilidade de maior detalhe analítico do processo de geração do sinal sonoro, o qual corresponde a eventos das porções glóticas e supraglóticas do aparelho fonador (Kent & Read, 1992). Note-se que Weber (2002) estudou um conjunto de patologias laríngeas, associando-as ao grau de disfonia, e reitera que em casos com alterações muito marcadas, o programa de análise acústica não é capaz de realizar a avaliação. Estas conclusões foram ratificadas por Carding et al. (2004), Ma & Yiu (2005), Choi et al. (2012), Manfredi et al. (2012) e Moers et al. (2012).

Os principais entraves (Titze & Liang, 1993; de Krom, 1994; Rabinov et al., 1995; Buder, 2000; Buttha, 2004; Behram, 2005; Pouchoulin, 2008; Vogel, 2011; Dejonckere et al, 2012; Moers et al., 2012) à generalização do uso do método acústico parecem ser o facto de:

1. A maioria das análises derivar do estudo de vogais sustentadas, material fonético considerado controverso na literatura, tal como explicado anteriormente.

2. A análise e descrição dos seus resultados estar associada a medidas estatísticas (análise discriminativa, análise de regressão, correlação, ...) dependentes de variáveis individuais, inerentes ao sujeito, tanto quantitativa como qualitativamente.
3. A aquisição de certas medidas estar dependente de um material específico, com custos monetários associados.
4. Ser questionável o uso da análise acústica nos casos de vozes muito irregulares, em especial a aplicabilidade destas medidas a sinais que não pertençam ao Tipo I postulado por Titze (1995) (*Cf. Capítulo 2.4., Secção c*)).

2.4. A AVALIAÇÃO E ANÁLISE ACÚSTICA DA VOZ

A avaliação acústica realiza medições do sinal sonoro vocal (Kent & Read, 1992). A aplicação de diversas formas de análise está dependente da correta digitalização do sinal acústico. Vários autores oferecem sugestões ou normas para obter um adequado registo áudio de vozes (Behlau, 2001; Dejonckere et al., 2011; Mendes et al., 2003; Pinho, 2003; Murdoch, 2005; Smits et al., 2005; Guimarães, 2007; Dejonckere, 2010; Vogel, 2011; Madill & McCabe, 2011; Ma, 2011; Olszewski et al., 2011) e conseguir uma correta análise do ponto de vista visual (espectrografia) e quantitativa (extração de medidas temporais, de periodicidade, amplitude, perturbação, ruído ou outras).

Para a sua correta interpretação deve ter-se uma série de cuidados (Camargo, 2000; Behlau, 2001; Mendes et al., 2003; Maryn et al., 2009; Camargo & Madureira, 2010; Dejonckere, 2010; Vogel, 2011; Madill & McCabe, 2011; Ma, 2011), agrupados em quatro categorias: (a) quanto às condições de registo; (b) quanto aos protocolos de gravação; (c) quanto aos sistemas de gravação e análise; e, (d) quanto à análise qualitativa do sinal acústico.

a) Quanto às condições de registo:

A revisão da literatura sugere:

- A captação direta da voz para o computador é ideal usar no mínimo 16 bit de resolução por amostra, para reduzir o chamado ruído de quantização gerado pela placa de som (Smits et al., 2005; Choi et al., 2012);
- A taxa de amostragem, ou seja, a precisão da escala de medição usada para recolher amostras da onda acústica para representação digital, deve ser ligeiramente superior a 20000 amostras por segundo para que, de acordo com o teorema de Nyquist, frequências até 10.000Hz possam ser preservadas no sinal digitalizado (Titze, 1995; Baken & Orlikoff, 2000; Waugh, 2000; Guimarães, 2007). Se a extensão em frequência da audição humana varia entre 15 a 20 000 Hz, então uma frequência de amostragem de 44100Hz (2 x 20 KHz = 40 kHz) por segundo é suficiente para representar todo o espectro de interesse e evitar o risco de saturação (Deliyski et al, 2005; Pinho et al., 2006; Choi et al., 2012).

- As gravações devem ser realizadas em ambientes silenciosos, idealmente em cabine insonorizada e protegida contra sinais elétricos, com ruído de fundo inferior a 50 dB (Deliyski et al., 2005; Dejonckere, 2010; Akpinar et al., 2011; Moon et al., 2012), em circunstâncias controladas e passíveis de reprodução (Dejonckere, 2010);
- Qualquer método de aquisição e armazenamento de sinal pode afetar, substancialmente, a qualidade do mesmo (Spielman et al., 2006; Vogel, 2011);
- Vozes importadas – seja de que sistema for – mesmo digitais, podem ter sido significativamente alteradas por ruídos, adicionados aquando da sua captura (Vogel & Maruff, 2008);
- Uma verificação parcial do nível de ruído introduzido no sinal pelo sistema de gravação ou captura de voz pode ser realizada, comparando-se dois regtos: gravação com microfone desligado e outra com ele ligado (Vogel & Morgan, 2009);
- O indivíduo deve estar, preferencialmente, de pé, para melhor controlo postural e aproximação ao uso habitual da voz;
- O microfone deve ser instalado num apoio fixo, num ângulo de 45 a 90º da boca do falante, a uma distância inferior a 10 cm (entre 3 a 4 cm) – para recolha de vogais sustentadas (Titze, 1995; Batalla et al., 2004; Vieira et al., 2005; Camargo e Madureira, 2009; Hamdam et al., 2009; Dejonckere, 2010; Madill & McCabe, 2011; Uloza et al., 2011; Choi et al., 2012); ou de 10 cm para a fala encadeada – para evitar interferências no sinal e manter uma proporção sinal-ruído elevada (Behlau, 2001; Ma & Yiu, 2006; Moon et al., 2012). Outros autores propõem 15 cm (6 pol.) da comissura labial direita, preferencialmente controlados com um microfone de cabeça (Murdoch, 2005; Carmo et al., 2006; Eadie & Baylor, 2006; Torre III & Barlow, 2009; Dejonckere, 2010). Há estudos que usam 30 cm de distância, face a microfones de mesa e de cabeça (de Krom, 1994; De Bodt et al., 1997; Wuyts et al., 1999; Shrivastav et al., 2005; Bele, 2005; Jesus et al., 2009; van Gogh et al., 2012). Para manter constante a distância da boca ao microfone, especialmente com crianças, é preferível o uso de microfones de cabeça (Madill & McCabe, 2011);

- Deve ser verificada a impedância e sensitividade do microfone, para não haver distorções na gravação (Titze & Winholtz, 1993; Vogel & Morgan, 2009);
- Caso a emissão tenha excesso de ar (soprosidade) o microfone deve ser posicionado lateralmente à boca do indivíduo, podendo atingir os 90° de deslocamento, para reduzir o ruído aerodinâmico que distorce o sinal acústico (Titze & Winholtz, 1993; Vogel & Morgan, 2009; Dejonckere, 2010; Madill & McCabe, 2011).

b) Quanto aos protocolos de gravação:

As amostras de fala usadas na prática clínica e na investigação diferem (1) no tipo – vogais sustentadas, leitura, conversação, canto ou outra; (2) na forma de produção – voz suave, habitual, projeção vocal ou outra; e, (3) na duração – tempo ou unidade de fala, com consequências na validade e fiabilidade das medições (Dejonckere et al., 2001; Awan, 2001; Pinho et al., 2006; Guimarães, 2007). Assim, deve-se tentar diminuir esta variabilidade contemplando:

- Sempre o mesmo protocolo de gravação, incluindo no registo de cada indivíduo a identificação de, pelo menos, uma vogal sustentada (de Krom, 1994) e uma sequência automática de fala encadeada (Moers et al., 2012);
- A análise da frequência fundamental e os seus índices de perturbação são, geralmente, realizadas com base nas vogais /i/, /u/ e /a/ (Awan, 2001; Behlau, 2001; Dejonckere, 2010; Moon et al., 2012);
- A análise da qualidade vocal suporta-se numa amostra de fala encadeada (Hammarberg, 2000; Heman-Ackah et al., 2002; Halberstam, 2004; Eadie & Doyle; 2005; Eadie & Baylor, 2006; Ma & Yiu, 2006; Martens et al., 2007; Moers et al., 2012);
- A necessidade de um maior número de análises em casos em que o desvio vocal é mais acentuado (Ma & Yiu, 2005). Scherer et al., 1995 (*in* Pinho et al., 2006) sugerem que as medidas de perturbação (vogais sustentadas) devem ser baseadas em, pelo menos, 360 ciclos glóticos, o que equivale a 3,0 seg. para a voz masculina e 1,6 seg. para a feminina. Porém, outros autores referem valores mínimos inferiores: 110 a 190 ciclos como mínimo fiável (Karnell et al., 1991);

- A verificação imediata da qualidade da gravação, logo após a sua realização, antes que o paciente saia do consultório.

c) Quanto aos sistemas de gravação:

O principal aspeto a ter em atenção é a escolha do microfone para a captação de voz (Vogel, 2011). Assim, este deve ser selecionado quanto ao tipo (dinâmicos), direcionalidade (omnidirecional), frequência de resposta (linear), impedância (baixa), modo de uso e distância da fonte (Guimarães, 2007; Vogel & Morgan, 2009). Contudo, atente-se a que:

- Ao usar um gravador este deve ser profissional, equipado com um microfone de condensador, mono, unidirecional, com sensibilidade mínima de -60dB, deslocado do corpo da unidade de gravação para evitar a captação de ruído do *hardware*;
- Alguns formatos ou suportes (e.g. *minidisc*) revelam uma diferença reduzida da qualidade de gravação face à sua realização direta para o computador. Assim, deve ser evitada a edição e registo das amostras, assim como a sua passagem para outros formatos ou suportes, que implicam perdas devido ao uso de algoritmos de compressão que alteram os parâmetros mais sensíveis;
- O armazenamento em CD-R (*recordable compact disc*) ou CD-RW (*compact disc rewritable*) é adequado em termos de durabilidade e fiabilidade da qualidade dos dados;
- As vozes registadas através de gravadores ou câmaras de vídeo apenas devem ser analisadas áudio percetivamente, uma vez que o microfone está acoplado ao corpo da câmara e isso acrescenta ruídos extrínsecos ao sinal acústico, além de interferências acústicas como ecos e reverberações.

d) Quanto à análise qualitativa do sinal acústico:

Esta característica está dependente da qualidade da gravação do som. Neste sentido Titze (1995) publicou um documento baseado no *Workshop on Acoustic Analysis* em que sugere a existência de três tipos de sinais sonoros:

- **Sinal Tipo 1:** periódico com pequenas perturbações aleatórias – é um sinal quase periódico, que não apresenta alterações qualitativas no segmento a ser analisado;

se existirem modulações ou sub-harmónicos a sua energia tem magnitude inferior à da frequência fundamental (F0). Os sinais deste tipo permitem medições de F0, *jitter*, *shimmer* e índice sinal-ruído. As vozes normais ou ligeiramente alteradas produzem um sinal acústico deste tipo (*Cf.* Figura 3).

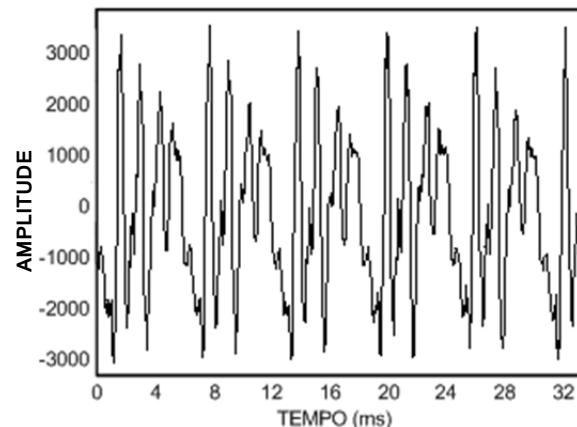


Figura 3: Exemplo de Sinal Tipo 1 (representação temporal).

- **Sinal Tipo 2:** periódico e com estrutura sub-harmónica e modular – é um sinal acústico que apresenta alterações qualitativas no segmento analisado, ou seja, bifurcações, intermitênicas, sub-harmónicos (*Cf.* Figura 4) e modulações. A frequência dos sub-harmónicos ou modulações é de energia considerável e, por isso, não existe uma frequência fundamental única no segmento em questão. Embora se possam selecionar as amostras mais estáveis, tal não corresponde às características globais da emissão do falante, sendo necessários pelo menos 100 ciclos para uma medição confiável (Guimarães, 2007). Apenas permite a análise visual do traçado, uma vez que a alteração sobreposta não possibilita uma caracterização fiável. A maioria das vozes patológicas produz um sinal deste tipo.

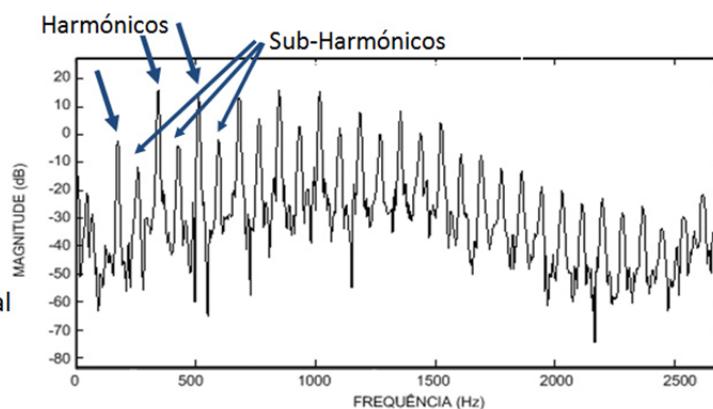


Figura 4: Exemplo de Sinal Tipo 2 (representação espectral).

- **Sinal Tipo 3:** sinal não periódico, não tem estrutura periódica estável (Cf. Figura 5), ou seja, é instável (“caótico” – Titze, 1995; Parsa & Jamieson, 2001) e, por isso, os investigadores consideram que fica condicionada a sua mensuração fiável, mesmo pela análise visual (Karnell et al., 1997; Yiu, 1999; Buder, 2000; Carding et al., 2004; Ma & Yiu, 2005). Giovanni et al. (1999) propõem alguns sistemas de medição não linear como a análise fractal, atractores ou o expoente de Lyapunov para a análise acústica não linear deste tipo de sinal glótico (Yiu et al., 2000; Behlau, 2001; MacCallum, et al., 2009; Awan et al., 2010; Dejonckere, 2010; Vaziri et al., 2010), a qual salienta o estudo pormenorizado das bifurcações ou mudanças súbitas da qualidade do padrão vibratório.

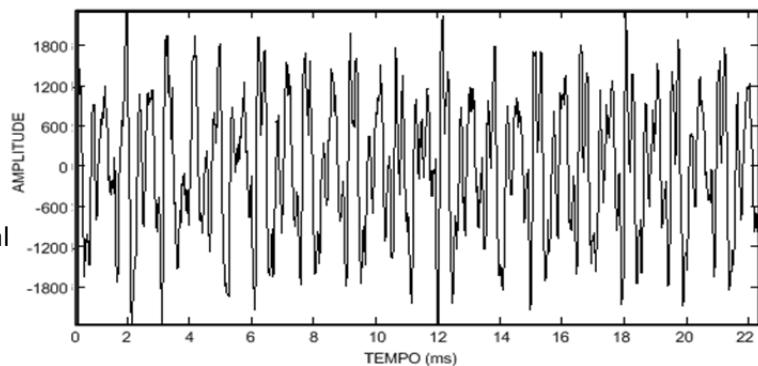


Figura 5: Exemplo de Sinal Tipo 3 (representação temporal).

Recentemente Sprecher et al. (2010) propuseram um sinal do tipo 4:

- **Sinal Tipo 4:** é um sinal de ruído estocástico e é, por isso, incompatível com medições lineares ou através da utilização de medidas de perturbação (Sprecher et al., 2010). É prevalente nas vozes dominadas por componentes de ruído (Michaelis et al., 1998). As únicas formas de avaliação deste tipo de voz são as qualitativas com recurso a espectrogramas, análise áudio-percetiva e dimensões de correlação infinitas, tal como postulado por Milenkovic & Read (1992), Titze & Liang (1993) e Titze, em 1995. Este tipo de sinal é altamente complexo, pelo que ainda estão em estudo formas objetivas de medição e análise das mesmas (Awan et al., 2010; Choi et al., 2012).

Saliente-se que os fracos resultados experimentais observados nos métodos LPC (*Linear Predictive Coefficients*) são o reflexo dos desafios que se impõem aquando da modelização do espectro auditivo. A aperiodicidade e as bifurcações não

lineares do sinal glótico invalidam o uso das medidas de perturbação que recorrem aos métodos LPC, pois são vulneráveis à extração precisa da frequência fundamental (Titze, 1995; Choi et al., 2012). É assim cada vez mais evidente o esforço das tecnologias da engenharia de processamento de sinal para ir de encontro à pouca uniformidade da banda de frequência do espectro audível (Hermansky, 1990; Hajaiej et al., 2006; Pouchoullin, 2008), aspecto reforçado no contexto das perturbações da voz, em que os desvios da periodicidade podem ser aleatórios ou correlacionados, mas na maioria dos casos são tão extremos que condicionam a identificação de um período regular para análise (Zhang & Jiang, 2008; Maryn et al., 2009; Sprecher et al., 2010). Neste domínio há particularidades temporais e espectrais do sinal diretamente relacionadas com a disfonia que não podem ser nem descartados, nem excessivamente predominantes no processo de parametrização (Yumoto et al., 1982; Milenovic et al., 1987; de Krom et al., 1993; Herzel et al., 1994; Qi et al., 1997; Buder, 2000; Sataloff & Hawkshaw, 2001 *in* Kent, 2004; Maryn et al., 2009; Awan et al., 2010).

Deste modo, outra dificuldade adicional na análise acústica do sinal vocal é o sistema empregue na extração dos parâmetros de análise, uma vez que ainda não existe uma padronização do mesmo (Vaziri et al., 2010). Por outro lado, as medidas acústicas fornecem valores médios para os parâmetros acústicos do sinal de fala que, na maioria dos casos, dependem da determinação da frequência fundamental e tornam-se inválidos em amostras de voz com instabilidade temporal (superior a 5%) significativa (Titze, 1995; Parsa & Jamieson, 2001; Choi et al., 2012). A aperiodicidade e as bifurcações não lineares do sinal invalidam o uso das medidas de perturbação baseadas no domínio do tempo (Milenovic & Read, 1982; Yumoto et al., 1982; Milenovic, 1987; Titze & Liang, 1993; Choi et al., 2012). No estudo de Espanhol (2004) foi demonstrado que os “determinadores” disponíveis (*jitter*, *shimmer* e *Signal-to-Noise Ratio*) não apresentam a robustez pretendida. Umaphaty et al. (2005) comentam que embora esta seja uma das ferramentas mais usadas, ainda não existe consenso quanto à sua utilidade na discriminação entre vozes normais e patológicas (Eadie & Baylor, 2005; Ma & Yiu, 2006). A título de exemplo, Klingholtz (1990) realizou um estudo em que fez a

análise estatística da correlação entre sinal-ruído (*Signal-to-Noise Ratio* – SNR) em tarefas de fala encadeada *versus* vogais sustentadas, para uma amostra de 50 vozes patológicas e 74 normais. Concluiu que a medida acústica estudada era mais sensível – capaz de discriminar entre vozes normais e alteradas – em tarefas de fala encadeada (5,6% de erro) do que nas amostras de vogal sustentada (22,6% de erro). Qi et al. (1999) analisaram a capacidade do algoritmo estimador de SNR discriminar entre disfonia-voz normal, com uma correlação de 0.78, e uma capacidade de concordância de 95% quando se consideravam vozes altamente disfónicas ou próximas do normal. Parsa & Jamieson (2001) realizaram um estudo semelhante e concluíram que as medidas de perturbação não eram suficientemente discriminativas – em tarefas de fala encadeada – para distinguir entre voz normal *versus* patológica, contudo a medida espectral média de longo termo (*tilt espectral*) e o SFR (*spectral flatness ratio*) explicavam 96% dos resultados. Recentemente, Moon et al. (2012) compararam as medidas acústicas do domínio do tempo de 202 sujeitos com vozes normais, na produção de vogal sustentada *versus* leitura de um texto, e concluíram da existência de tendências diferenciadoras – embora não estatisticamente significativas – consoante o género.

Tipicamente são encontradas na literatura grandes diferenças nos valores das medidas para os diferentes tipos de *software* disponíveis. Tal compromete a viabilidade e fiabilidade dos resultados e limita fortemente a possibilidade de comparação inter estudos (Carson et al., 2003; Smits et al., 2005; Martens et al., 2007; Lopes et al., 2008; Maryn et al., 2009; Choi et al., 2012a). Esta foi uma das considerações tidas ao longo da dissertação e, por isso, foram comparados os resultados fornecidos para as mesmas vozes, por quatro aplicações de *software* distintas (*Cf.* Capítulo 4).

Embora as investigações das últimas duas décadas e a sua aplicação clínica tenham trazido inúmeras contribuições, importa ressaltar que este é um procedimento complexo (Brasolotto & Rehder, 2011; Vogel, 2011; Madill & McCabe, 2011) que requer um suporte teórico e treino para o manuseio dos

recursos computorizados disponíveis atualmente. Note-se que muitos investigadores consideram, consensualmente, que a análise acústica da voz patológica – baseada em medidas que dependem da extração prévia da periodicidade do sinal, i.e. em parâmetros de perturbação – não tem fiabilidade, pois o sinal vocal possui uma aperiodicidade de grau extremamente variável, que muitos dos programas de *software* não são capazes de estimar de forma eficaz (Titze, 1995; Bielamowicz et al., 1996; Yiu, 1999; Yiu et al., 2000; Buder, 2000; Ma & Yiu, 2005; Choi et al., 2012). Assim, recomenda-se cautela na utilização deste recurso de avaliação vocal isolado (Buder, 2000; Welham et al., 2003; Hsiung et al., 2004; Eadie & Baylor, 2005; Ma & Yiu, 2006; Pinto et al., 2007; Maryn et al., 2009; Camargo & Madureira, 2010; Dajer, 2010; Madill & McCabe, 2011; Vogel, 2011) por forma a conseguir que ele tenha um verdadeiro significado clínico, principalmente em casos de vozes com muitos sub-harmónicos e modulações, ou com um comportamento de sinal acústico caótico (Buder, 2000; Batalla et al., 2004; Choi et al., 2012a), o que corresponde à grande maioria dos pacientes que são acompanhados clinicamente pelos terapeutas da fala.

2.5. PARÂMETROS ACÚSTICOS E IMPLICAÇÕES CLÍNICAS

Tal como foi previamente descrito, o sinal acústico decorrente do vozeamento de uma vogal sustentada (tipicamente o /a/) é captado por um microfone, digitalizado e, posteriormente, analisado através de um procedimento computacional para medição de fatores de perturbação objetivos, associados à forma de onda do vozeamento (*Cf. Figura 6*).

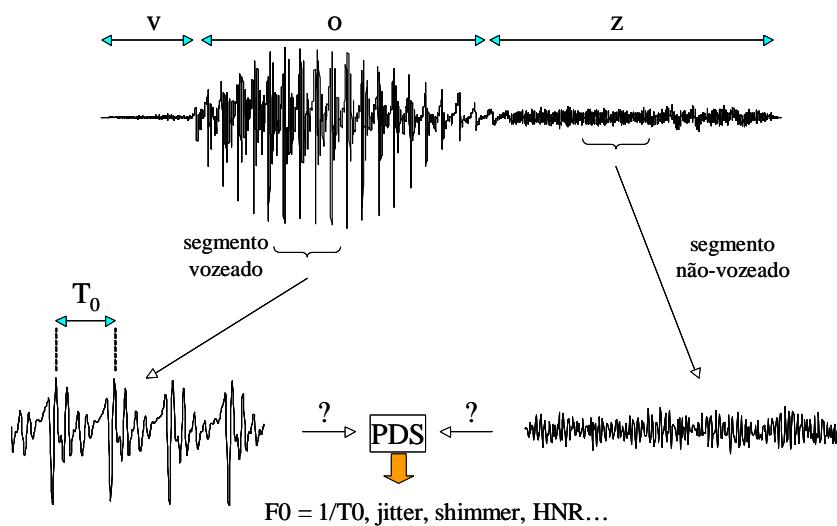


Figura 6: Ilustração do sinal de voz captado por um microfone e correspondente à palavra *voz*. Destaca-se a região vozeada do sinal e a região não-vozeada. Usando técnicas de Processamento Digital de Sinal (PDS) é possível a medição objetiva e precisa de alguns parâmetros de perturbação extraídos diretamente do sinal acústico.

O sinal da voz pode ser analisado de formas distintas: diretas e indiretas.

- **Formas diretas do sinal:** são aquelas em que a análise do sinal se baseia na onda temporal de pressão acústica que irradia da boca, isto é, consiste na técnica que interpreta o sinal vocal como um fenômeno de pressão sonora (Fant, 1960).
- **Formas indiretas do sinal:** pressupõem uma versão modificada do sinal vocal, como por exemplo o espectro, o *cepstrum* ou o sinal residual decorrente de análise LPC. Podem incluir a separação – através de fórmulas matemáticas de modelagem do trato vocal – entre a fonte glótica e o filtro (Rothenberg, 1973; Wong et al., 1979). Há técnicas específicas de filtragem inversa que retiram todo o efeito do trato. Existem

dois métodos a destacar: técnica de filtragem inversa residual e técnica de filtragem inversa glotal (Davis, 1979; Pereira e Montagnoli, 1999; Buder, 2000).

A técnica de filtragem inversa residual remove os efeitos acústicos do trato vocal sobre as características do sinal de voz, de forma a especificar, aproximadamente, aquele que resulta da excitação glótica. Ou seja, inclui um conjunto de técnicas que subtraem um modelo fixo dos sub-sistemas glótico e supraglótico, mantendo as informações do primeiro (Buder, 2000). As técnicas de filtragem inversa glotal removem os efeitos acústicos da radiação labial e do trato vocal sobre as características do sinal da voz, preservando as do sub-sistema glótico. Assim, o sinal obtido é o glótico e não o residual (Davis, 1979; Pereira e Montagnoli, 1999).

Matematicamente é possível a separação entre a radiação labial e o trato vocal, uma vez que cada um resulta em diferenças acústicas conhecidas (Buder, 2000). O trato vocal é responsável pelas ressonâncias (formantes), enquanto a radiação labial condiciona, principalmente, o acoplamento com o meio de propagação do som.

Na bibliografia são referidas muitas medidas (CPP – *Cepstral Peak Proeminence* (Hilenbrand et al., 1994); GNE – *Glottal-to-Noise Excitation ratio* (Michelis et al., 1997); NNE – *Normalized Noise Energy* (Kasuya et al., 1996); SPI – *Soft Phonation Index* (Deliyski, 1993); VTI – *Voice Turbulence Index* (Deliyski, 1993); TNI – *Turbulent Noise Index* (Mitev & Hadjitolorov, 2000); SRA – *Sum of Rahmonics Amplitudes* (Murphy, 2006); DVB – *Degree of Voice Breaks* (Nikolov et al., 1989); DSH – *Degree of Sub-Harmonics* (Deliyski, 1993)) que, até à data, rondam a centena (Buder, 2000; Gelzinis et al., 2008; Maryn et al., 2009; Ma & Yiu, 2011; Choi et al., 2012). Nesta dissertação realizaremos a descrição e análise de medidas computacionais mais vulgarizadas que recorrem a técnicas de Processamento Digital de Sinal e que permitem a medição de parâmetros objetivos como a frequência fundamental (F_0) da voz (ou *pitch*¹¹), parâmetros de perturbação como o *jitter* ou *shimmer*, ou de qualidade como a relação harmônico-ruído (HNR), os quais são transversais aos quatro programas de *software*.

¹¹ Em rigor, o *pitch* denota o correspondente psicofísico (i.e., perceptivo) da frequência fundamental (F_0) e é condicionado por outros fatores objetivos do sinal de voz, como seja a sua intensidade.

selecionados para a análise das vozes-alvo de estudo ao longo da dissertação, e descritos no capítulo 4.

2.5.1. ESPECTROGRAFIA

Esta é uma das técnicas mais usadas nas investigações com análise do sinal acústico, para a qual os requisitos de equipamentos são acessíveis. O seu princípio básico deriva de um processo matemático, a Transformada de Fourier. A análise computorizada de Fourier ficou conhecida pelo uso de um método numérico eficiente designado FFT (*Fast Fourier Transform*). Na FFT o espectro de frequências é apresentado num número de bandas, sendo que a largura de banda é ajustada por um parâmetro designado de “tamanho” ou “número de pontos”. Em geral, esta característica é – por peculiaridades do algoritmo – uma potência de 2 (tipicamente 256, 512, 1.024, ...). Quanto maior for o número de pontos, mais estreita cada banda do espectro da FFT e melhor a capacidade de separação dos harmónicos da voz (Kent, 2004; Pinho et al., 2006; Howard et al., 2012; Ghoraani et al., 2012).

A informação disponibilizada por esta técnica na sua representação, designada por espetrograma, surge sob a forma tridimensional com: (a) o tempo no eixo horizontal; (b) a frequência no eixo vertical; e, (c) a amplitude na acentuação de cor (e.g. grau de escurecimento na gama de cinzento) das barras horizontais (*Cf. Figura 7*).

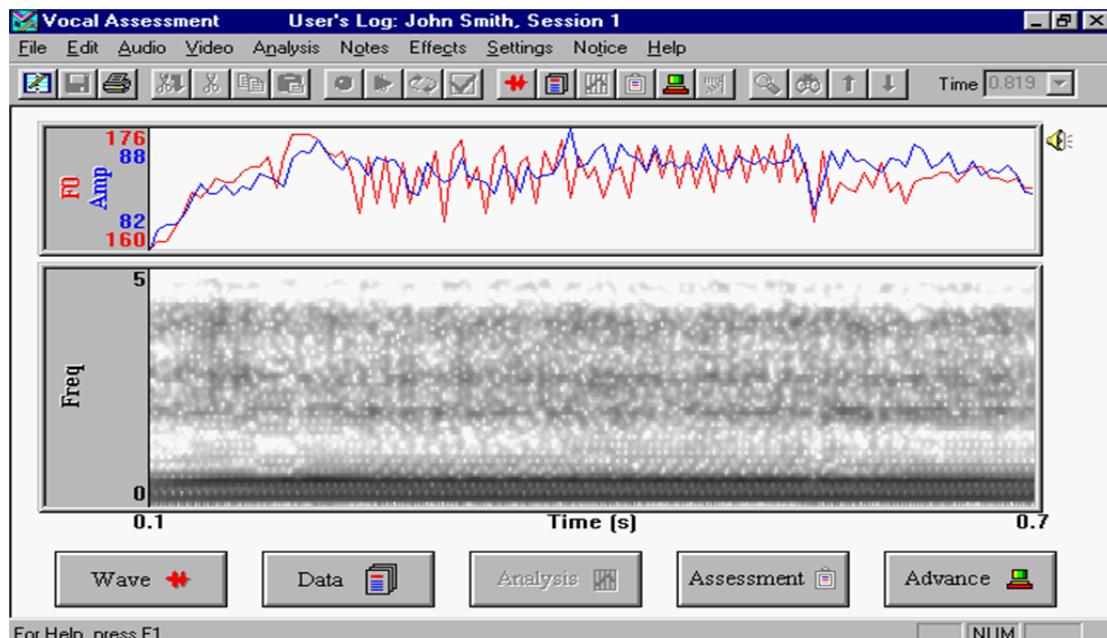


Figura 7: Espectrograma de um caso com o software Dr. Speech®.

A vantagem da espectrografia é o facto de refletir de forma expedita as características da voz, a F0 e os seus harmónicos correspondentes, a amplitude de cada um dos harmónicos e a duração do registo vocal, com os devidos ajustes na duração da janela temporal e no número de pontos da FFT. As três componentes que a compõem (enumeradas no parágrafo anterior) formam a chamada série harmónica do som periódico, onde a frequência de cada componente é um múltiplo inteiro da frequência fundamental. Isto é, a frequência mais baixa é considerada a frequência ou harmónico fundamental (F0) e as seguintes são múltiplos desta, ou seja, segundo harmónico e seguintes. Por outro lado, as componentes não periódicas da voz, como o ruído gerado pelas consoantes fricativas ou nas fendas glóticas, apresentam um espectro denso de frequências ao contrário das linhas espectrais bem definidas como nas séries harmónicas (Pinho et al., 2006; Howard et al., 2012)

Sons quase-periódicos, como os das vogais, são representados pela sobreposição de várias componentes sinusoidais, cada uma com a sua frequência, amplitude e desfasamento (atraso) (Titze, 1995; Pinho et al., 2006). O som complexo como o da voz humana pode ser analisado quanto ao seu conteúdo em frequência através do uso de filtros de banda larga (*Cf.* Figura 8; i.e., de FFT com um número reduzido de pontos) ou de banda estreita (*Cf.* Figura 9; i.e. de FFT com um número elevado de pontos). Se for usado um filtro de banda larga é obtida uma melhor resolução temporal, enquanto o de banda estreita potencia a resolução da frequência.

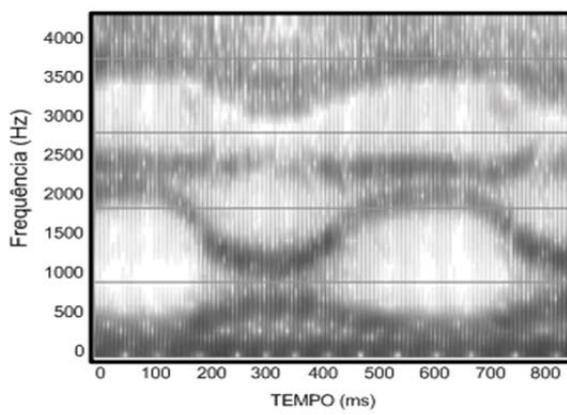


Figura 8: espectro de banda larga.



Figura 9: espectro de banda estreita.

O spectrograma de banda larga resulta da aplicação de um filtro de banda larga e, geralmente, tem uma faixa de frequências de 200-500Hz (Murdoch, 2005). É um

gráfico de resolução temporal disposto em estrias verticais, que salienta as regiões de energia do sinal, apontando a existência de atividades periódica e aperiódica do mesmo. Assim, é possível a identificação das frequências de ressonância, ou seja, os formantes, e ainda as zonas de concentração de ruídos, representadas por “chuva” no fundo do traçado. Os formantes são visíveis como barras escuras, que indicam zonas de concentração de energia. A disposição dos mesmos tem uma relação direta com a configuração do trato vocal durante a emissão do som, variável de acordo com características de postura dos articuladores – faríngeos e bocais – associados (Dejonckere, 2010; Souza, 2010; Ghoraani et al., 2012).

O espectrograma de banda estreita é retratado com estrias horizontais, cada uma representa um harmônico. São designados deste modo porque evidenciam a estreita largura de banda do sinal da fala – em torno dos 30-50Hz. Não realiza medições temporais, sendo útil na obtenção da frequência fundamental e de aspectos associados à prosódia (Murdoch, 2005). Tem como principal finalidade a representação da estrutura harmônica do sinal da fala (Dejonckere, 2010; Souza, 2010).

2.5.2. FREQUÊNCIA FUNDAMENTAL

A frequência fundamental vocal (F0) ou frequência fundamental da fala (SFF ou SF0, abreviaturas do inglês) corresponde à velocidade na qual uma forma de onda se repete por unidade de tempo, no comportamento vocal sustentado ou em fala encadeada (Buder, 2000; Behlau, 2001). Assim, reflete o número de ciclos vibratórios produzidos pelas pregas vocais, num segundo (Pinho, 2003; Seikel et al., 2010). É classificada como uma medida acústica no domínio do tempo (Titze & Liang, 1993).

A sua medição pode ser representada através da equação (Buder, 2000):

$$\frac{1}{N} \sum_{i=1}^N F0_i, \quad (= \overline{F0}). \quad (1.1)$$

Reflete a eficiência do sistema fonatório, a biomecânica laríngea (comprimento natural da prega vocal, alongamento, massa em vibração e tensão envolvida) e a sua interação com a aerodinâmica (pressão subglótica resultante da passagem do ar pulmonar) (Hess, 1983 e 1994; Titze, 1989; Torre III & Barlow, 2009).

As unidades de medida usadas podem ser os ciclos por segundo (cps), o Hertz (Hz), os semitons (ST) e/ou as oitavas. As duas primeiras unidades de medida são as mais comuns.

Os dados normativos relativos à F0 são vários, com assunção de que é enormemente afetada pelo género (Peterson & Barney, 1952; Monsen & Engebretson, 1977; Klatt & Klatt, 1990; Nittrouer et al., 1990; Baken & Orlikoff, 2000), idade (Harrington et al., 2007; Nishio & Niimi, 2008) (Cf. Figura 10), comportamento vocal avaliado, hábitos pessoais e estilo de vida (tabaco, álcool), o uso de voz profissional (Garrett & Healey, 1987; Brown et al., 1990) e o tipo de perturbação vocal apresentada (Kent & Read, 1992).

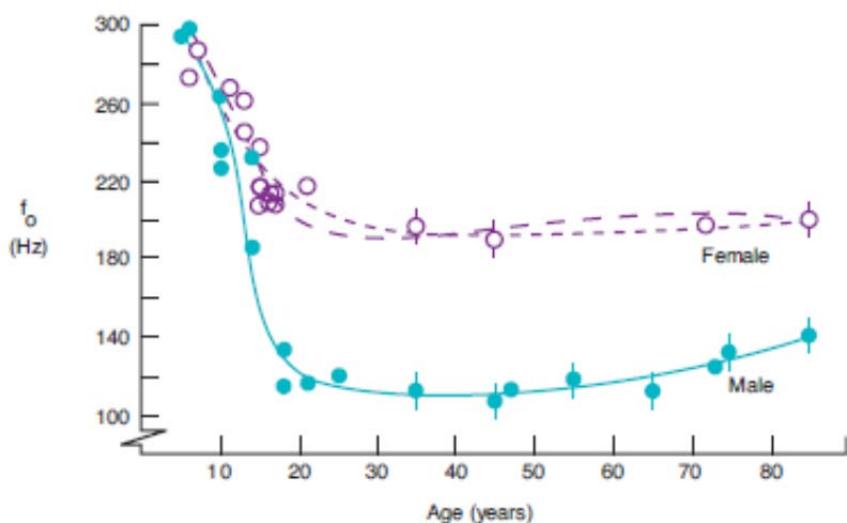


Figura 10: Representação esquemática da mudança vocal na adolescência e ao longo do ciclo de vida, para homens e mulheres (baseado em Kent, 1994 in Seikel et al., 2010).

Kelley (1977, citado por Pinho, 2003) concluiu que mulheres na faixa etária dos 20-29 anos apresentam uma F0 média que ronda os 227Hz. Segundo o mesmo estudo, esta

vai diminuindo com a idade (para o género feminino). Outros investigadores apontaram valores para homens, mulheres e crianças de 128, 225 e 265Hz, respetivamente. Wuyts et al. (1996) definiram para homens gamas de frequência entre 78-166Hz e para as mulheres 167-258Hz. A descrição e caracterização pormenorizada da variabilidade da frequência fundamental ao longo do ciclo vital está explícita no trabalho de Torre III & Barlow (2009).

Uma abordagem comum para determinar o valor de F0 de cada ciclo é pela comparação da forma de onda (*wavematching*) de dois segmentos adjacentes e com o mesmo comprimento (Sundberg, 1987). Outra alternativa é a utilização de algoritmos baseados em eventos na forma de onda (por exemplo, picos, cruzamentos por zero) (Vieira, 1997 *in* Pinho et al., 2006).

Tipicamente a estimação do valor da F0 considera um referencial médio, obtido por média aritmética, que é a soma das medidas de F0 individuais dividida pelo número de ondas captadas. Note-se que deste modo os dados são, por vezes, contaminados por sub-harmónicos ou quebras repentinas de frequência (normais/usuais) na voz humana (Manfredi et al., 2012). Assim, é crucial reter os dois viés básicos na análise e interpretação de resultados derivados da determinação da F0:

- 1) A determinação da F0 baseada em algoritmos e técnicas de Processamento Digital de Sinal é frequentemente programada para detetar e analisar ciclo-a-ciclo o sinal de voz e, neste modo, não necessariamente corresponde à sensação psicofísica de *pitch* e/ou à forma de vibração das pregas vocais.
- 2) A inspeção visual dos resultados dos algoritmos é sempre necessária, como forma de precaução, em qualquer tipo de medida derivada de cálculos que incluam a F0, especialmente para as vozes com alterações (Buder, 2000; Choi et al., 2012a; Moers et al., 2012).

Ressalte-se que a alteração do parâmetro F0 pode provocar mudanças na aferição automática de outras medidas acústicas, as quais guardam nos seus processos a etapa inicial da extração da F0 (Titze, 1995; Karnell et al., 1997; Dejonckere et al., 2001; Ma & Yiu, 2005; Maryn et al., 2009; Choi et al., 2012a). Neste sentido, as medidas

dependentes de F0 sofrem a interferência da aperiodicidade do sinal e devem ser cuidadosamente aplicadas (Titze, 1995; Vieira et al., 1996; Buder, 2000; Vieira et al., 2002; Camargo et al., 2003; Brasolotto & Rehder, 2011).

Manfredi et al. (2012) argumentam que contrariamente à análise espectral (onde os algoritmos são amplamente conhecidos e os programas, em geral, confiáveis) a extração automática da F0 é um problema delicado, potencialmente sujeito a erros que se propagam às medidas de perturbação a curto-prazo, que dela derivam (Titze & Liang, 1993; Titze, 1995; Buder, 2000; Parsa & Jamieson, 2001; Kreiman & Gerrat, 2005; Pinho et al., 2006; Maryn et al., 2009; Choi et al., 2012a).

Um outro método de extração da frequência fundamental que requer, contudo, um processador de alta velocidade, é o *cepstrum*¹². Este é um método extremamente poderoso de obtenção da F0, que pode ser aplicado em fala encadeada e em vozes acentuadamente alteradas (Heman-Ackah et al., 2002; Heman-Ackah et al., 2003; Mehta & Hilman, 2008; Awan & Roy, 2009; Moers et al., 2012). Baseia-se no facto do sinal vocal ser a convolução do sinal da fonte (laringe) e a resposta do trato vocal (sistema de ressonância). Este processo significa a interação multiplicativa entre dois conjuntos de propriedades espetrais (de Krom, 1993; Qi & Hillman, 1997; Guimarães, 2007). Awan & Roy (2005) provaram que a medida de CPP (*cepstral peak measures*) era capaz de discriminar entre vozes normais e patológicas. Estes resultados derivam do facto das medidas de *cepstrum* não serem muito perturbadas por variáveis como a técnica de gravação das amostras, volume de gravação ou aperiodicidade do sinal de voz (Carding et al., 2004; Gelzinis et al., 2008; Awan et al., 2009; Choi et al., 2012a).

Sugestões para obtenção de medidas de F0

- Utilizar, pelo menos, um segundo de emissão para a análise (mínimo de 100 ciclos) e eliminação do início e fim da produção (para limitar as irregularidades)

¹² O *cepstrum* aqui considerado (*cepstrum real*) consiste na Transformada de Fourier inversa do logaritmo do espetro (o que explica a designação de ‘ceps’ como inverso de ‘spec’). Remete portanto para um domínio do tempo que caracteriza a periodicidade existente no espetro. Em termos práticos, é útil, por exemplo, para calcular o período fundamental (em segundos) de uma estrutura harmónica (Buder, 2000).

(Madill & McCabe, 2011; Olszewski et al., 2011; Vogel, 2011; Moers et al., 2012);

- Usar vogais sustentadas, tendo em atenção que a frequência é vogal-dependente (Behlau, 2001), isto é, existe um valor intrínseco a cada vogal que deve ser tido em conta no resultado apresentado;
- De notar também que, de acordo com a tarefa de avaliação – vogal sustentada versus fala encadeada – os valores de F0 são condicionados pelo sexo do falante (efeito *white coat* (Moon et al., 2012) – os homens baixam a frequência fundamental na produção da vogal sustentada mais do que em conversação; as mulheres tendem ao inverso, isto é, aumentam a frequência fundamental na sustentação de vogais;
- As medidas de F0 obtidas com fala encadeada têm sido questionadas, pela variabilidade que introduzem no teste. Contudo, com análises *cepstrais* e métodos pautados em dinâmica não linear os resultados são mais robustos (Mehta & Hillman, 2008; Awan et al., 2010; Dejonckere et al., 2012; Moers et al., 2012).

Frequência fundamental nas perturbações vocais

- Muitas disfonias caracterizam-se por alterações significativas da F0 (por exemplo, durante a muda vocal típica da puberdade ou associadas a lesões de massa) (Buder, 2000);
- A frequência fundamental em adultos muda, de forma estatisticamente significativa, após a realização de adenoamigdalectomia (Mora et al., 2009). Akpinar et al. (2011) não encontraram este efeito no seu estudo;
- A frequência fundamental pode ser influenciada pelos fatores comprimento, alongamento, massa e tensão das pregas vocais e possui integração com a pressão sub-glótica (Torre III & Barlow, 2009; Brasolotto & Rehder, 2011);
- A F0 parece não mudar significativamente com a Terapia da Fala, salvo exceções (Hufnagle & Hufnagle, 1984; Dejonckere, 2010);
- Vozes com crepitação e rugosidade tendem a apresentar um F0 grave, enquanto vozes ásperas caracterizam-se com F0 agudo (Hirano et al., 1986);

- Situações de extrema tensão psicológica podem implicar vozes excepcionalmente agudas (Scherer, 1981 e 1986; Vogel et al., 2009).

2.5.3. DESVIO PADRÃO DA FREQUÊNCIA FUNDAMENTAL (*pitch sigma*)

Na fala normal é esperado um certo teor de variabilidade da frequência fundamental (Buder, 2000; Pinho et al., 2006; Guimarães, 2007). Todavia, as situações extremas são indesejáveis.

Esta medida é geralmente expressa em semitons ou então calculada com base num índice de variabilidade – o desvio padrão da F0 (medida de dispersão calculada pela raiz quadrada da soma dos quadrados dos desvios da média). Na literatura também pode aparecer com a designação *pitch sigma* (Ladefoged, 2006).

A sua medição pode ser representada através da equação (Buder, 2000):

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (F0_i - \bar{F0})^2} \quad (1.2)$$

Sugestões para a obtenção de medidas de variabilidade de F0

- Pode medir-se com base numa vogal sustentada ou fala encadeada (Madill & McCabe, 2011; Vogel, 2011);
- Quando medida em vogais sustentadas não ultrapassa os 2Hz em indivíduos normais. Nesta situação a variabilidade indica o desvio padrão da fundamental. Contudo, tal depende do protocolo de classificação de cada programa usado: o CSL (*Computer Speech Lab* da Kay Elemetrics) considera que para homens o F0 S.D. varia entre 0-2.9Hz, e para mulheres 0-6.6Hz (Wuyts et al., 1996). O Dr. Speech (Tiger Electronics) estabelece como limite 0-3.0Hz.
- A variabilidade da F0 na fala encadeada pode ser obtida através da leitura de um texto, conversação espontânea ou séries automáticas. A análise assume cuidados especiais consoante as tarefas (Moers et al., 2012).

Variabilidade da frequência fundamental nas perturbações vocais

- As disfonias de etiologia neurológica associam-se, frequentemente, a grande variabilidade de F0 (Behlau, 2001; Pinho et al., 2006; Vogel, 2011; Madill & McCabe, 2011);
- Valores mais elevados de *pitch sigma* estão associados ao parâmetro Instabilidade e hiperfuncionalidade/tensão (Hammarberg et al., 1986);
- Murry & Doherty (1980) consideram o aumento do desvio padrão da F0 uma medida discriminativa na análise da voz de casos com neoplasias nas pregas vocais;
- Momentos de *stress*, ansiedade ou emoções fortes positivas influenciam (aumentam) o desvio padrão da F0, salvo raras exceções, como situações de tristeza/lamúria (Scherer, 1981 e 1986);
- Estudos consideram esta medida eficaz para predizer o grau geral de alteração vocal (Wolfe & Steinfatt, 1987; Callan et al., 1999 *in* Awan & Roy, 2009; Finizia et al., 1999).

2.5.4. JITTER – PERTURBAÇÃO DA FREQUÊNCIA A CURTO PRAZO

O *jitter* é uma medida de curto termo (ciclo a ciclo), de variabilidade não voluntária na F0 ou período (Titze, 1995; Baken & Orlikoff, 2000; Pinho et al., 2006; Mora et al., 2009; Dejonckere, 2010; Manfredi et al., 2012). Não deve ser confundida com a medida do grau de oscilação de baixa frequência que forma a base do *vibrato* ou do tremor (índices de perturbação da frequência a longo prazo). Na sua forma mais simples, o *jitter* de dois períodos (ou ciclos) glóticos¹³ consecutivos de frequências $F0_{(1)}$ e $F0_{(2)}$ pode ser calculado pela equação

¹³ O ciclo glótico – também designado por vibratório ou fonatório – é um ciclo de vibrações das pregas vocais, desde que começam a abduzir até à próxima vez em que tal acontece (Pinho et al., 2008; Souza, 2010).

$$J(\%) = \frac{|F0_{(1)} - F0_{(2)}|}{\left(\frac{1}{2} [F0_{(1)} + F0_{(2)}] \right)} \times 100 . \quad (1.3)$$

O *jitter* instantâneo, definido de acordo com a função de perturbação de primeira ordem (Titze, 1995), pode ser expresso matematicamente por

$$J(k) = \frac{[P + \Delta(k)] - [P + \Delta(k-1)]}{\frac{1}{2} \{ [P + \Delta(k)] + [P + \Delta(k-1)] \}} , \quad (1.4)$$

onde $J(k)$ é o valor de *jitter* no k -ésimo ciclo, P é o período fundamental sem *jitter* e $\Delta(k)$ é a perturbação. O período fundamental é, portanto,

$$TT(k) = P + \Delta(k) . \quad (1.5)$$

Define-se \bar{J} como *jitter* médio, obtido através de

$$\bar{J} = \frac{1}{N-1} \sum_{k=2}^N \frac{|\hat{T}(k) - \hat{T}(k-1)|}{\frac{1}{2} [\hat{T}(k) + \hat{T}(k-1)]} , \quad (1.6)$$

onde $\hat{T}(k)$ é k -ésimo período fundamental de uma dada amostra e N é o número de períodos fundamentais (Buder, 2000).

A extração dos valores de *jitter* pode ser feita através de medidas absolutas ou relativas (Manfredi et al., 2012). Estas correlacionam-se com a média da F0. As primeiras ignoram a F0 do indivíduo. A partir deste parâmetro são derivadas as seguintes medidas acústicas (MEEI, 1994; Moran et al., 2006), usadas consensualmente em aplicações de *software* de processamento digital de sinal e algumas adotadas na dissertação (cf. Capítulo 4):

- *Phonatory Frequency Range* [-] (PFR): é o número máximo de semitons existente na gama de variação da frequência fundamental.

-*Jitter [%]* (Jitt): representa a variação do tom de cada par de períodos de tom consecutivos relativamente ao valor médio dos valores de frequência fundamental contidos no segmento de voz em análise.

- *Absolute Jitter [μs]* (Jita): representa a variação do tom entre cada par de períodos de tom consecutivos, medida em micro segundos.

- *Pitch Perturbation Quotient [%]* (PPQ): é calculada através do valor médio dos desvios de frequência fundamental de cada conjunto de 3, 5, 11 ou 55 períodos de tom consecutivos no segmento de voz em análise. Quando o período de suavização inclui 3 períodos de tom esta medida é usualmente chamada de *Relative Average Perturbation* (RAP) e quando inclui 55 períodos de tom tem a designação de *Smoothed Pitch Perturbation Quotient* (sPPQ).

- *Fundamental Frequency Variation [%]* (vF0): representa a variação relativa do desvio-padrão em torno da frequência fundamental, calculado período a período.

- *F0-Tremor Intensity Index [%]* (FTRI): é calculada através do cociente entre o valor da frequência da componente de baixa-frequência mais elevada e a soma dos valores de frequência contidos no segmento de voz em análise.

- *Pitch Perturbation Factor [%]* (PPF): é determinada com base no cociente entre o número de vezes que o valor de um período de tom ultrapassa um determinado limiar e o número total de períodos de tom considerados.

- *Directional Perturbation Factor [%]* (DPF): é calculada através do cociente entre o número de vezes em que a diferença dos valores de frequência de cada par de períodos de tom consecutivos muda de sinal e o número total de períodos de tom considerados.

Tal como já foi referido, a voz humana caracteriza-se por ondas quase-periódicas, pelo que é aceitável um certo grau de instabilidade vocal (Lieberman, 1961; Glaze, 1990; Ferrand, 1995; Buder, 2000; Guimarães, 2007), em especial por razões neurológicas, emocionais e biomecânicas (Pinho et al., 2006). Se o mecanismo vocal fosse perfeitamente estável, então não existiriam diferenças nos períodos fundamentais e o *jitter* seria zero a que se associaria uma sonoridade robótica. Portanto, na voz humana, o *jitter* ocorre pela diminuição dos controlos neuro-motor e aerodinâmico inerentes à produção vocal (Wertzner et al., 2005), além de alterações nos parâmetros dos tecidos

(Hemler et al., 1997; Buder, 2000) e, possivelmente, da ação do *feedback* retro-auditivo no controlo do sistema fonador (Burnett et al., 1997). O *jitter* não indica as grandes variações da sustentação da frequência, mas sim a pequena variabilidade entre os sucessivos ciclos glóticos (Buder, 2000). Para alguns autores (Lieberman, 1960; Laver, 1980) o *jitter* é uma medida de aperiodicidade que diminui a fiabilidade quando a periodicidade aumenta.

A maioria dos investigadores considera como intervalo de referência – de *jitter* saudável para as fonações sustentadas em jovens adultos – os valores entre 0,5-1,0%. As variações na sua magnitude estão associadas ao (Behlau, 2001; Pinho, 2003; Pinho et al., 2006; Guimarães, 2007; Brasolotto & Rehder, 2011; Olszewski et al., 2011; Dejonckere et al., 2012; Manfredi et al., 2012):

- a) Acoplamento entre a região glótica e supraglótica;
- b) Histologia da prega vocal e assimetria mecânica;
- c) Distribuição de muco sobre as pregas vocais durante a vibração;
- d) Pequena variação de massa ou de tensão nas pregas vocais;
- e) Sensação tátil laríngea.

Sucintamente, as fontes deste tipo de perturbação são (Baken & Orlikoff, 2000): neurológicas (por exemplo, variações no controlo da tensão muscular por lesões neuronais ou descontrolo emocional); aeromecânicas; mecânicas (por exemplo, alterações nas propriedades biomecânicas do tecido das pregas vocais) e, ainda, estilísticas (por exemplo, o vibrato).

Sugestões para a obtenção de medidas de *jitter*

- Deve ser medido com base na análise de vogais sustentadas, numa única frequência, sem variação musical ou de intensidade, em emissão habitual (Dejonckere et al., 2012), eliminando-se o início e final da produção – para ultrapassar a dificuldade de demarcação dos ciclos glóticos. Koike (1973 in de Krum, 1994) provou que o *jitter* é mais alto/alterado no *onset* e *offset* da vogal, comparativamente com a porção mais estável da amostra vocal;

- Há variações intrínsecas a cada vogal – os valores de *jitter* são mais confiáveis para a vogal /a/ do que para /i/ ou /u/ (Doherty & Shipp, 1988; Linville et al., 1990; Vieira et al., 1997 *in* Pinho et al., 2006). Horii (1982) não encontrou diferenças significativas de *jitter* na análise desta medida em vogais distintas. Deve ser indicado no protocolo de avaliação qual a vogal usada;
- Os valores de *jitter* são influenciados pelo género e idade do falante: são encontrados valores mais elevados em crianças (Linville et al., 1990; Campisi et al., 2002; Albertini et al., 2009) e mais baixos nos homens (Moon et al., 2012);
- O *jitter* pode ser aumentado artificialmente por fatores que vão desde distorções introduzidas pelo instrumento de áudio, até à fórmula usada para o seu cálculo, passando pelo tipo de vogal analisada e pelo mecanismo de extração de F0 (Buder, 2000; Pinho et al., 2006; Manfredi et al., 2012);
- A forma de captação do sinal é importante e deve ser indicada. Gravações que apresentem ruído de fundo ou interferências invalidam a análise de *jitter* (Dejonckere et al., 2012; Manfredi et al., 2012);
- Há valores mínimos de frequência de amostragem (20 kHz) e dimensão da amostra (entre 20-190 ciclos) para medidas de *jitter* mais consistentes (Karnell, 1991 *in* Zhang e Jiang, 2008; Dejonckere et al., 2012);
- Para diminuir erros na demarcação dos períodos calcula-se o *jitter* não apenas baseado em dois ciclos, mas em três ou mais (Koike, 1967; Hirano, 1981).

***Jitter* nas perturbações vocais**

- Altera-se, principalmente, com a falta de controlo da vibração das pregas vocais, como acontece nas disfonias de causa neurológica e nas lesões neoplásicas ou vozes de substituição (Takahashi & Koike, 1976; Ramig & Shipp, 1987; Buder, 2000; Behlau, 2001; Dejonckere et al., 2012);
- Existem diferenças estatisticamente significativas entre os valores de *jitter* de pacientes com alteração vocal *versus* sem alteração quando se avaliam vogais sustentadas (Buder, 2000; Ma e Yiu, 2006; Zhang e Jiang, 2008; Silva et al., 2009), e em amostras de fala espontânea (Vasilakis, M. & Stylianau, Y., 2009);

- Existem diferenças ao longo do ciclo vital – com valores esperados de $0.59\% \pm 0.54\%$ para adultos-jovens *versus* $2.10\% \pm 1.55\%$ em homens idosos (presbifonia). Para as mulheres obtiveram-se $5.34\% \pm 4.51\%$ e $2.02\% \pm 2.03\%$, respectivamente. (Xue e Deliyski, 2001);
- O *Computer Speech Lab* da Kay Elemetrics sugere o valor 1,040% como limiar máximo da normalidade e aponta como valores referência 0-2.1% para homens e 0-3.1% para mulheres (Wuyts et al., 1996; Gonzalez et al., 2002; Xue & Deliyski, 2002; Verdonck-de Leeuw & Mathieu, 2004). Contudo, estudos com outros programas de *software*, nomeadamente o Praat, encontram outros valores (Oguz et al., 2007; Brochmann et al., 2008; Stajner-Katusic et al., 2008; Amir et al., 2009; Mathieson, et al., 2009; Silva et al., 2009).
- O *jitter* correlaciona-se com a Aspereza e/ou Rouquidão (Lieberman, 1961; Koike, 1967; Yumoto et al., 1984; Wolfe & Steinfatt, 1987; Dejonckere et al., 1993; Buder, 2000; Deal e Emanuel, 1978 *in* Yiu et al., 2000; Titze, 1995; Ma & Yiu, 2006), Grau geral de disfonia (Ma & Yiu, 2006; Martens et al., 2007) e a Soprosidade (Eskenazi et al, 1990; Dejonckere, 1995).

2.5.5. SHIMMER – PERTURBAÇÃO DA AMPLITUDE A CURTO PRAZO

O *shimmer* quantifica as alterações mínimas da amplitude do sinal, a curto prazo, com base em cada ciclo fonatório (Baken & Orlikoff, 2000; Buder, 2000; Pinho et al., 2006; Dejonckere, 2010). Se o sistema fonador fosse completamente estável ele seria zero. O *shimmer* não indica as grandes variações da amplitude, ou seja, as quebras de intensidade durante a emissão, mas sim a pequena variabilidade entre os sucessivos ciclos glóticos (Buder, 2000; Handam et al., 2009).

Matematicamente o *shimmer* de dois ciclos consecutivos com amplitudes $A_{(1)}$ e $A_{(2)}$ pode ser calculado pela função de perturbação de primeira ordem como

$$S(\%) = \frac{|A_{(1)} - A_{(2)}|}{\left(\frac{1}{2}[A_{(1)} + A_{(2)}]\right)} \times 100. \quad (1.7)$$

As investigações sobre esta temática sugerem que esta medida é inversamente proporcional à intensidade¹⁴ média, ou seja, quanto maior for esta, menor o valor de *shimmer*, e vice-versa (Orlikoff & Kahane, 1991; Akpinar et al., 2012). O *shimmer* oferece-nos uma percepção indireta do ruído na produção vocal. Assim, nas perturbações da voz o *shimmer* surge mais alterado nas frequências graves e intensidade fraca. O *jitter* e o *shimmer* tendem a covariar (Wong et al., 1995) e por isso é frequente a sua medição conjunta (Buder, 2000).

As medidas relativas de *shimmer* podem ser apresentadas de diferentes formas (MEEI, 1994; Buder, 2000; Moran et al., 2006), usadas consensualmente em aplicações de *software* de processamento digital de sinal e algumas adotadas na dissertação (cf. Capítulo 4):

- *Shimmer [%]* (Shim): em percentagem (fator de perturbação direcional), cujo valor limite usado habitualmente é 3.0% (Behlau, 2001), ou então 5.0% (0.44dB) (Lindsey, 1997 *in* Pinho et al., 2006). Representa a variação da amplitude de cada par de períodos de tom consecutivos relativamente ao valor médio das amplitudes contidas no segmento de voz em análise.
- *Shimmer [dB]* (ShdB): em dB, com o coeficiente logarítmico da amplitude de ciclos consecutivos, através da fórmula

$$S(\text{dB}) = 20 \log_{10} \left[\frac{A_{(1)}}{A_{(2)}} \right], \quad (1.8)$$

em que os valores desta medida são, tipicamente, 0.4dB, diminuindo à medida que aumenta o volume vocal. Representa a variação média da amplitude entre cada par de períodos de tom consecutivos.

- *Amplitude Perturbation Quotient [%]* (APQ): em percentagem, é calculada através do valor médio dos desvios de amplitude de cada conjunto de 3, 5, 11 ou 55 períodos de tom consecutivos relativamente ao valor médio das amplitudes contidas no segmento

¹⁴ A intensidade vocal relaciona-se diretamente com a pressão sub glótica, a resistência das pregas vocais a essa pressão e a configuração do trato vocal (Boone e McFarlane, 2000; Colton e Casper, 1996).

de voz em análise. Quando o período de suavização inclui 3 períodos de tom esta medida é usualmente chamada de *Amplitude Relative average Perturbation* (ARP) e quando inclui 55 períodos de tom tem a designação de *Smoothed Amplitude Perturbation Quotient* (sAPQ).

- *Peak Amplitude Variation [%]* (vAm): em percentagem, representa a variação relativa do desvio-padrão em torno da amplitude, calculada período a período.
- *Amplitude Tremor Intensity Index [%]* (ATRI): em percentagem, é calculada através do cociente entre o valor da amplitude da componente de baixa-frequência mais elevada e a soma das amplitudes contidas no segmento de voz em análise.
- *Amplitude Perturbation Factor [%]* (APF): em percentagem, é calculada através do cociente entre o número de vezes que a amplitude de um período de tom ultrapassa um determinado limiar e o número total de períodos de tom considerado.
- *Amplitude Directional Perturbation Factor [%]* (ADPF): em percentagem, é calculada através do cociente entre o número de vezes que a diferença das amplitudes de cada par de períodos de tom consecutivos muda de sinal e o número total de períodos de tom considerado.

Sugestões para a obtenção de medidas de *shimmer*

- Deve ser medido com base na análise de vogais sustentadas, numa única frequência, sem variação musical ou de intensidade, em emissão habitual, eliminando-se o início e final da produção, para minimizar a dificuldade de demarcação dos ciclos glóticos (Vogel, 2011; Madill & McCabe, 2011; Moon et al. 2012);
- Horii (1982) não encontrou diferenças significativas de *shimmer* na análise desta medida em vogais distintas, não obstante, deve ser indicado no protocolo de avaliação qual a vogal usada;
- Moon et al. (2012) não encontraram diferenças estatisticamente significativas entre os valores de amplitude e amplitude média de sinal, pelo que ratificam a não existência de variações nos resultados obtidos em tarefas de vogal sustentada versus leitura de texto;
- Os valores de *shimmer* são influenciados pelo género e idade do falante: são obtidos valores superiores nos homens e indivíduos idosos (Ringel & Chodzko-

Zajco, 1987; Verdonck-de Leeuw & Mathieu, 2004; Albertini et al., 2009; Moon et al., 2012);

- A forma de captação do sinal é importante e deve ser indicada no protocolo de gravação;
- Há evidências práticas e científicas que indicam que o ambiente de gravação (número de interlocutores, tipo de comunicação necessária, ...) influencia a intensidade vocal usada, o que pode influenciar as medições de *shimmer* (Behlau, 2001);
- Estudos sugerem que para obter este parâmetro é necessário um mínimo de 130 ciclos glóticos (Karnell, 1991 *in* Zhang e Jiang, 2008);
- Por razões puramente empíricas (Buder, 2000) calcula-se o *shimmer* com base numa média de ciclos glóticos mais extensa, por exemplo, a medida de perturbação média relativa baseia-se em 11 ciclos (APQ – *Amplitude Perturbation Quotient*) (Takahashi & Koike, 1976; Nikolov et al., 1989).

***Shimmer* nas perturbações vocais**

- Altera-se, principalmente, nas situações de redução da resistência glótica (exemplo: paralisias/parésias¹⁵ das pregas vocais com variações na velocidade de fechamento glótico ou outras fendas glóticas) (Wong et al., 1991; Lewis et al., 1993). Note-se que as investigações de Ludlow et al. (1987) eram opostas a estes resultados;
- Também é influenciado por lesões de massa, mesmo as de tamanho reduzido, e neoplasias (Laver et al., 1986; Nikolov et al., 1989; Feijoo & Hernández, 1990);
- Existem diferenças ao longo do ciclo vital – com valores esperados de $2.52\% \pm 1.00\%$ para adultos-jovens *versus* $5.54\% \pm 3.51\%$ em homens idosos

¹⁵ As lesões nos nervos laríngeos recorrentes ou superiores podem resultar numa paralisia (ausência de movimento) ou parésia (diminuição ou lentificação do movimento) unilateral ou bilateral das pregas vocais (Boone e McFarlane, 1994; Kent, 2004). Se a adução glótica não for completa, resulta numa voz soprada (Pinho et al., 2006). Nas paralisias bilaterais os sintomas são mais exacerbados, e podem apresentar aspiração durante a deglutição (Pinho et al., 2006; Pinho e Pontes, 2008; Souza, 2010).

(presbifonia). Para as mulheres obtiveram-se $2.00\% \pm 0.79\%$ e $5.34\% \pm 4.51\%$, respetivamente (Xue e Deliyski, 2001). Wuyts et al., 1996, apontam para os homens valores de *shimmer* entre 0.7-6.4% e para mulheres entre 0.0-7.5%.

- Subramiam & Kumar (2009) encontraram diferenças estatisticamente significativas entre os valores de *shimmer* de pacientes adenoamigdalectomizados, quando avaliaram vogais sustentadas, antes e depois da cirurgia;
- Newman et al. (1989) encontraram diferenças estatisticamente significativas entre os valores de *shimmer* de pacientes com gaguez *versus* com velocidade de fala normal, quando avaliaram vogais sustentadas;
- O *Computer Speech Lab* da Kay Elemetrics sugere o valor 3.810% como limiar máximo da normalidade. Diferentes estudos com o uso deste *software* analisam esta medida (Wuyts et al., 1996; Gonzalez et al., 2002; Xue & Deliyski, 2002; Verdonck-de Leeuw & Mathieu, 2004). O PRAAT também encontra outros dados quantitativos, variáveis consoante a população em análise. (Oguz et al., 2007; Brochmann et al., 2008; Stajner-Katusic et al., 2008; Amir et al., 2009; Mathieson, et al., 2009).
- Relaciona-se com o Grau geral de disfonia (*grade*) (Martens et al., 2007) e com a Soprosidade (Takahashi & Koike, 1976; Feijoo & Hernández, 1990; Dejonckere et al., 1993), ou com a Rugosidade (Hillenbrand, 1988; Feijoo & Hernández, 1990; Tittze, 1995; Eadie & Baylor, 2006; Martens et al., 2007);
- Parece ser mais sensível à associação com as perturbações vocais que o *jitter* (Hillenbrand, 1988; Buder, 2000).

2.5.6. HNR – PROPORÇÃO HARMÓNICO-RUÍDO

Segundo Stevens (1998), em laringes normais, aquando da completa ou parcial adução glótica, tem lugar um fluxo de ar através da constrição definida pelo espaço que existe entre as pregas vocais (glote). Aquele forma um jato que pressiona as paredes do trato vocal, ligeiramente acima do nível das pregas vocais (*Cf. Figura 11*).

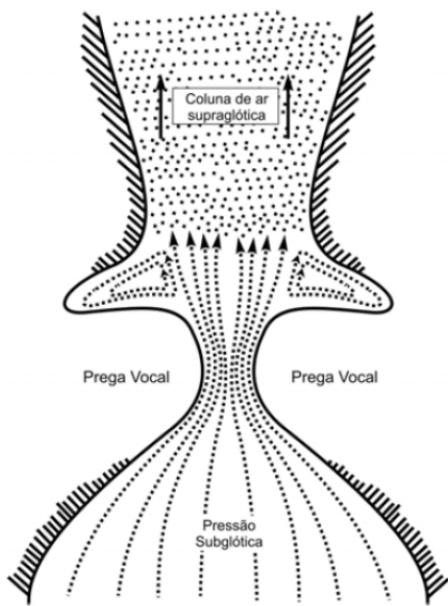


Figura 11: Desenho esquemático da direção do fluxo aéreo através da glote e em contacto com a coluna de ar supraglótica (adaptado de Titze, 1994).

O fluxo turbulento gerado nas paredes das regiões supraglóticas é considerado como a fonte dominante de ruído, admitindo que não ocorrem constricções na faringe nem na cavidade oral. Esta fonte de ruído está localizada entre 1,0 a 2,5cm acima da glote, ou no nível das pregas ventriculares (ou falsas pregas) que se encontram 0,5cm acima das pregas vocais propriamente ditas. As fontes de ruído de cada um destes locais podem ser representadas como uma fonte de pressão sonora. Além destas, podem encontrar-se fontes adicionais (por exemplo, as fendas glóticas) causadoras de flutuações aleatórias do fluxo glótico, que geram um aumento da turbulência do fluxo de ar (Hirano, 1981).

A medida HNR (*Harmonic-to-Noise Ratio*, proposta por Yumoto et al., 1982) é uma avaliação objetiva, isto é, de base matemática, que relaciona a componente periódica e aperiódica (Guimarães, 2007) que compõem um segmento de fala sustentada, estimada com base na análise de pelo menos 50 ciclos glóticos consecutivos (Yumoto, 1983; Yumoto et al. 1984) (Cf. Figura 12).

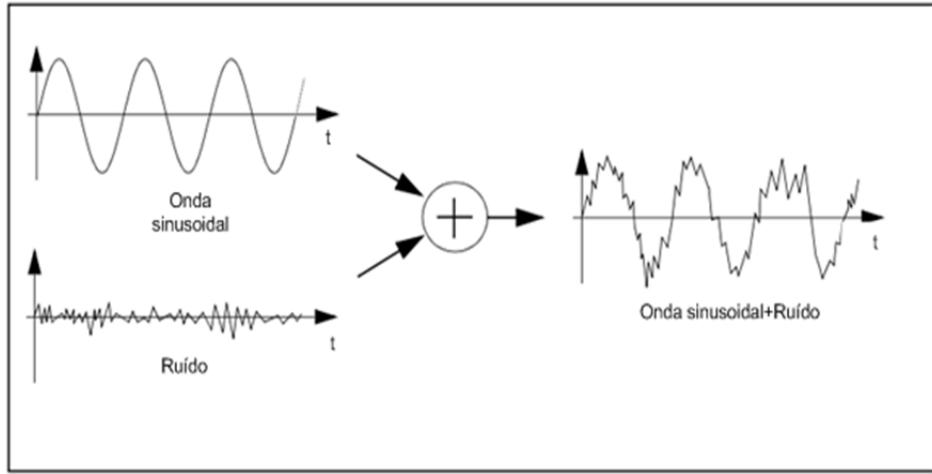


Figura 12: Sinal de onda complexa – representação isolada da onda sinusoidal e de ruído e a conjugação dos dois componentes.

A primeira componente decorre da vibração periódica das pregas (energia dos harmónicos) e a segunda do ruído glótico (sobretudo energia inter-harmónica). O valor de HNR resulta da divisão entre a média da potência do sinal e a média da potência da componente de ruído da onda (de Krom, 1993).

Yumoto e colaboradores (Yumoto et al., 1982; Yumoto, 1983; Yumoto et al. 1984) propuseram um método para estimar o ruído glótico que obedece à fórmula

$$HNR = 10 \log \left\{ \frac{50 \sum_{n=0}^{T_{\max}} [\bar{x}(n)]^2}{\sum_{i=1}^{50} \sum_{n=0}^{T_{\max}} [x_i(n) - \bar{x}(n)]^2} \right\}, \quad (1.9)$$

$$\text{em que } \bar{x}(n) = \frac{1}{50} \sum_{i=1}^{50} x_i(n).$$

Nesta fórmula $\bar{x}(n)$ representa a média dos 50 períodos de sinal após alinhamento temporal, o que tem como efeito cancelar a influência do ruído que é considerado aleatório. O valor T_{\max} é o máximo período glotal entre os 50 ciclos. Se um dado período T_i é menor que T_{\max} , o alinhamento é feito acrescentando zeros à respetiva forma de onda, isto é, $x_i = 0, T_i \leq n \leq T_{\max}, \forall T_i < T_{\max}$. A escolha do número de

ciclos é arbitrária, salientando-se que em Yumoto et al. (1982) se considera possível obter um intervalo suficientemente longo de fonação estável para proceder à análise, em vogal sustentada.

A relação entre as duas componentes – periódica e aperiódica – traduz a eficiência do processo de fonação: quanto maior for a eficiência na utilização do fluxo de ar expelido pelos pulmões em energia de vibração das pregas vocais, e quanto mais íntegro (i.e., saudável ou escorreito) for o ciclo vibratório destas, maior será a relação HNR. Inversamente, quanto menor for aquela eficiência ou quanto mais anómalo for o ciclo vibratório, maior será o ruído glótico e mais baixa resultará a relação HNR (Krom, 1993). Portanto, a relação HNR mede a quantidade relativa de ruído adicional no sinal vocal, que pode ser gerado pela turbulência do fluxo aéreo na glote em casos de fechamento incompleto durante a fonação, ou pela vibração aperiódica da prega vocal. Uma voz saudável deve, assim, caracterizar-se por uma relação HNR elevada, a que se associa a impressão de voz sonora e harmónica.

Apesar de haver outras alternativas computacionais de cálculo da relação HNR, os problemas frequentemente encontrados na aplicação do método baseado na equação (1.9) são (Cox et al., 1989): a dificuldade na demarcação dos ciclos glóticos a analisar e o aumento na estimativa de ruído devido a flutuações na duração dos ciclos (*jitter*).

Uma outra medida de ruído designa-se por NNE (*Noise Normalized Energy* – Energia Normalizada de Ruído), proposta por Kasuya et al. (1986). Assume-se como um método para a estimação do ruído no domínio do tempo, baseado em filtros pente (*comb-filter*), que corresponde a uma média da relação ruído-harmónico (e não de harmônico-ruído), correspondendo à razão entre a energia do ruído e a energia total do sinal (componentes harmónicas e de ruído), ambas medidas em dB. Como a proporção de ruído/sinal é inferior à unidade, o valor de NNE é negativo (Buder, 2000; Pinho et al., 2006).

Nesta abordagem admite-se que o sinal de voz, $x(n)$, consiste numa componente periódica corrompida pelo ruído aditivo. Assim, detetam-se os períodos do sinal $x(n)$ e

submete-se este a um determinado filtro pente. A saída do filtro, $y(n)$ é uma estimativa da componente harmónica de $x(n)$. Estabelece-se então o índice NNE, definido por

$$NNE = 10 \log \frac{\sum [x(n) - y(n)]^2}{d \sum x(n)^2}, \quad (1.10)$$

onde n representa o índice das amostras dos sinais que correspondem aos ciclos considerados, e d é uma constante de compensação, que depende do número de ciclos usado. No caso do artigo citado (Kasuya et al., 1986a.), $d = 7$. Em Kasuya et al. (1986b.) os autores propõem uma versão para a medida NNE na frequência. Neste caso analisam-se 7 ciclos glóticos e calcula-se a Transformada de Fourier do trecho. O número de ciclos é definido de modo a minimizar os efeitos de distorção do sinal. A energia de ruído é obtida com base na energia entre os harmónicos. Consequentemente, este valor é normalizado pelo valor da energia do espectro, gerando a NNE no domínio da frequência.

Tal como para as restantes medidas, existem diversas formas de cálculo matemático da proporção harmónico-ruído (Buder, 2000; Lopes et al., 2009), sendo a mais recente proposta por Krom (1993) e, subsequentemente modificado por Qi (1997). Esta abordagem baseia-se na propriedade do *cepstrum*, o qual permitir desacoplar as componentes de variação rápida do espectro (relacionadas com os harmónicos) e as de variação lenta (inerentes à envolvente espectral que retrata, razoavelmente, o perfil do ruído e, portanto, os formantes). Deste modo – identificando os picos do espectro correspondentes às componentes harmónicas e usando diversos passos de filtragem, que permitem obter uma estimativa do espectro do ruído – é possível calcular o HNR (Buder, 2000). Apesar de mais direta, esta abordagem é vulnerável à natureza dos sinais de voz e, em particular, os seus resultados dependem muito da frequência fundamental e do *jitter* (Cox et al., 1989). Estes problemas foram subsequentemente minimizados em novos resultados publicados por Murphy (2007) (Lopes et al., 2009).

Sugestões para a obtenção de medidas de HNR

- Os valores das medidas de ruído são apresentados em dB (Buder, 2000);
- A componente ruído é tão maior quanto menores forem as componentes harmónicas na emissão (Buder, 2000; Vogel, 2011; Madill & McCabe, 2011);
- O HNR é tipicamente maior nos homens e menor nas mulheres (Behlau, 2001; Kent, 2004), devido sobretudo à maior densidade espectral de harmónicos, já que a frequência fundamental é mais baixa nos homens. Provavelmente também se justifica por, nas mulheres, a frequência fundamental ser mais alta e estar associada uma fenda glótica posterior (fisiológica), inerente a fatores anatómicos do arcaboiço laríngeo, que acarreta maior turbulência glótica aquando da produção do som. Estudos para medição dos erros na estimação do HNR resultantes de diferentes taxas de amostragem demonstraram que eles eram menores nas vogais dominadas por baixas frequências como o /u/, do que no /a/ e /i/ (Cox et al., 1989).

HNR nas perturbações vocais

- Indivíduos com grandes lesões de massa, difusas, podem apresentar valores de HNR muito baixos, assim como pequenas fendas glóticas podem resultar num HNR baixo, sem possibilidade de correspondência com o grau de disfonia ou a avaliação áudio-percetiva (Behlau, 2001);
- Speyer et al. (2004; a, b, c) não encontraram diferenças significativas de HNR na análise desta medida pré e pós-terapia da fala de casos com vozes de substituição, pós cirurgia laríngea;
- Valores de HNR inferiores a 7dB são, com grande probabilidade, patológicos (Yumoto et al., 1982; Behlau, 2001). Contudo, os valores de referência são variáveis (=11,9 dB para Yumoto et al., 1982; =22 dB para Vieira, 1997 *in* Pinho et al., 2006), variando de acordo com o *software* de análise acústica em uso e os algoritmos estabelecidos para o cálculo da medida de ruído.
- Vários autores encontraram forte concordância entre o HNR ou outras medidas de ruído e o parâmetro Rugosidade (Kojima et al., 1980; Eskenazi et al., 1990; Feijoo & Hernández, 1990; Krom, 1993; Martin et al., 1995) e Soprosidade

(Fukazawa et al., 1988; Hillenbrand, 1988; Krom, 1995; Kojima, Gould, Lambinanse and Isshiki, 1980; Feijoo & Hernández, 1990; Lee & Childers, 1991; Yumoto et al., 1984 *in* Yiu et al., 2000) e outros com o Grau geral de disfonia (*Grade*) (Dejonckere et al., 1993).

Em suma, de acordo com as necessidades clínicas e possibilidades de cada instituição, tendo em atenção os devidos cuidados já explanados, durante uma avaliação vocal para análise acústica sugere-se a gravação de, pelo menos, as seguintes emissões (Dejonckere, 2010; Brasolotto & Rehder, 2011; Ma, 2011; Madill & McCabe, 2011; Vogel, 2011):

- Vogal sustentada para extração da frequência fundamental (F0), desvio padrão da F0, *jitter*, *shimmer* e medidas de ruído; é aconselhável obter 3 amostras;
- Fala encadeada para extração da frequência fundamental média, mínima e máxima;
- Emissão de fala encadeada e vogal sustentada em frequência e intensidade habitual, grave, agudo, fraco e forte, para comparação dos dados acústicos;
- Vogal sustentada e fala encadeada para visualização em espectrograma;
- Emissão em *glissando* de toda a extensão vocal ou emissão de notas da extensão vocal em intensidades forte e fraca, para determinar o perfil de extensão vocal;
- Fala encadeada, para caracterizar o perfil de extensão da fala/fonetograma ou *voice range profile* (IALP, 1992);
- Emissões repetidas o mais rápido possível de /a/ e/ou /i/, como prova de diadococinésia láríngea¹⁶.

¹⁶ Tarefa de coordenação, alternância ou dissociação de movimentos glóticos (Kent, 2004; Brasolotto & Rehder, 2011).

2.6. REPRODUTIBILIDADE DOS RESULTADOS DA AVALIAÇÃO ACÚSTICA

O uso clínico de um determinado teste ou escala depende da reprodutibilidade dos seus parâmetros. Este aspeto, remetendo para o uso clínico das medidas do método acústico, sujeita-se a três fatores:

- A variabilidade interobservador, isto é, a variação de julgamentos entre diferentes programas de *software* de avaliação com recurso ao método acústico;
- A variabilidade intraobservador, ou seja, a variação de julgamentos de um mesmo avaliador ao longo do tempo. Este aspeto não se coloca no uso de aplicações de *software* de avaliação acústica, uma vez que o algoritmo de análise é constante num mesmo programa;
- A variabilidade intrassujeitos, entendida como a variação da qualidade vocal do falante em diferentes momentos temporais da recolha e análise da mesma. Este aspeto pode estar associado a fatores intrínsecos (emocionais, cansaço) ou extrínsecos que, se não forem controlados, condicionam os resultados da avaliação através do método acústico (Buder, 2000).

Esta secção foi construída em simultâneo com a descrita anteriormente para a reprodutibilidade dos resultados da avaliação áudio-percetiva. Consistiu numa pesquisa na Pubmed que possibilitou a revisão da literatura publicada (1950-2/julho/2012) sobre os resultados da avaliação acústica – formas de uso e respetiva validade dos resultados.

Foram estabelecidos os seguintes critérios de inclusão:

- Escritos em inglês e publicados até 2 de julho/2012
- Relativos a estudos com a qualidade vocal de humanos (objetiva e subjetiva)
- Artigos com referência a teste-reteste¹⁷.

Número total de artigos encontrados: 108

Total de artigos sobre avaliação acústica: 14

¹⁷ Este tipo de estudo é definido com a aplicação múltipla (duas ou mais vezes) de um mesmo teste em dois momentos temporais distintos, a uma dada amostra (Maroco, 2010).

Os catorze artigos que se enquadram nos critérios de inclusão foram publicados num período de 23 anos (entre 1989 e 2012). Aparentemente as medidas objetivas de quantificação das perturbações da voz são usadas há cerca de 25 anos, porém, são poucas as publicações que analisam a validade e consistência interna destas medidas (14 estudos).

Ao tentar esboçar uma comparação entre os resultados dos estudos deparamo-nos com as seguintes dificuldades: as medidas/parâmetros analisados são muito dispares, embora com um claro predomínio das medidas de perturbação; os programas de *software* usados nas publicações são também diferentes e com algoritmos de quantificação das medidas nem sempre conhecidos; a conceção metodológica das investigações é, igualmente, muito variável (Maryn et al., 2009).

A maioria das investigações analisadas centra os resultados na interpretação da consistência da medida/parâmetro em estudo (Higgins & Saxman, 1989; Gelfer, 1989; Gramming et al., 1991; Stassen, 1991; Stone & Rainey, 1991; Dwire & MaCauley, 1995; Mendoza & Munoz, 1996; Lee et al., 1999; Carding et al., 2004; Kania et al., 2004; Manfredi et al., 2012). Dois estudos tiram conclusões acerca do equipamento usado ou método de recolha e análise dos dados (Bough et al., 1996; Lee et al., 1999). Num dos estudos obtêm-se conclusões sobre a aplicabilidade clínica e interpretação fisiológica das medidas (Higgins et al., 1994). A influência de diferentes observadores/investigadores sobre os resultados foi analisada por dois dos estudos (Gramming et al., 1991; Dejonckere et al., 2012).

Uma quantidade marcante de investigações nesta área foi suportada por amostras de vozes sem patologia, e apenas num estudo são feitas análises com vozes patológicas (Bough et al., 1996). A investigação de um artigo analisa vozes normais, dos casos e controlos, mas em momentos temporais distintos (Carding et al., 2004). Dois artigos são referentes a amostras vocais sintéticas (Dejonckere et al., 2012; Manfredi et al., 2012).

Em síntese, o reduzido número de estudos publicados sobre esta temática pode dever-se à falta de consenso acerca das medidas mais adequadas para a definição da qualidade vocal (Titze, 1995; Buder, 2000; Maryn et al., 2009). Esta falta de uniformidade pode ser ainda a razão que justifica a multiplicidade de medidas/parâmetros estudados. As conclusões acerca da validade do teste-reteste são muito díspares; algumas medidas são consideradas válidas, mas outras apenas apresentam um fator pobre-moderado (Gelzinis et al., 2008; Dejonckere et al., 2012). Nenhum dos estudos analisados centra muita atenção nas diferenças interobservadores, talvez devido à reduzida intervenção do avaliador na análise acústica. Porém, na prática clínica a recolha de duas amostras de voz do mesmo paciente, em dois momentos temporais distintos, pode ser realizada por diferentes profissionais, pelo que também é válido questionar a existência de diferenças.

Deve ser então feito um esforço para uniformizar medidas e criar algum tipo de consenso no estudo e investigação das mesmas, para que deste modo seja mais viável o entendimento teórico-prático – entre a componente clínica e o trabalho de engenharia de processamento de sinal – que suporta esta área de conhecimento (Eskenazi et al., 1993; Wolfe & Martin, 1997; Buder, 2000). Há, inevitavelmente, falhas na aproximação entre a vertente acústica, áudio-percetiva e laríngea da caraterização de uma dada voz, as quais têm que ser reconhecidas, respeitadas e compreendidas, para melhor atuar com toda a informação pertinente que cada uma destas vertentes de caraterização vocal oferece (Dejonckere, 2010; Kreiman & Gerratt, 2011; Shrivastav, 2011; Choi et al., 2012).

Capítulo 3 – CONCORDÂNCIA ENTRE MEDIDAS DE AVALIAÇÃO ACÚSTICA E PARÂMETROS ÁUDIO-PERCETIVOS NA CARACTERIZAÇÃO DA VOZ HUMANA

3.1. ESTUDOS COM REFERÊNCIA À RELAÇÃO ENTRE AS MEDIDAS ACÚSTICAS E A AVALIAÇÃO PERCETIVA

A voz falada envolve, provavelmente, o sistema mais elaborado da comunicação humana, sendo compreensível que seja difícil, ou mesmo impossível, usar um método único que avalie de forma abrangente e precisa a qualidade vocal ou a sua deterioração (Behrman, 2005; Guimarães, 2007; Choi et al., 2012).

Neste sentido torna-se importante realizar análises multifatoriais que permitam um conhecimento amplo, adequado e eficaz da função laríngea e da qualidade vocal. Behlau, Madazio, Feijó e Pontes (2001) referem que a avaliação do paciente disfônico deve ser multiprofissional. A multidimensionalidade da voz humana apenas pode ser compreendida com a complementaridade de distintas formas de caracterização vocal (Takahashi & Koike, 1976; Orlikoff et al., 1999; Welhman et al., 2003; Carding, 2004; Hsiung et al., 2004; Ma & Yiu, 2006; Martens et al., 2007; Pinho et al., 2008; Sweeney e Sell, 2008; Behlau, 2010; Eadie et al., 2010a).

Saliente-se, então, que nenhuma forma de avaliação substitui outra – todas são complementares e construtivas no processo terapêutico (Colton e Casper, 1996; Andrade e Silva & Duprat, 2004; Bonatto et al., 2004; Bhuta, 2004; Oliveira, 2004; Sader & Hannayama, 2004; Eadie et al., 2005; Ma & Yiu 2006; Martens et al., 2007). Apesar de objetivos, os parâmetros acústicos não invalidam a avaliação áudio-percetiva por duas razões fundamentais: as dimensões de apreciação da qualidade de uma voz (quando caracterizada áudio percetivamente) são em maior número e, portanto, mais ricas do que os parâmetros acústicos relevantes e que reúnem maior consenso e aceitação na comunidade científica (Awan & Lawson, 2009). As tarefas de fala encadeada possibilitam uma avaliação mais realista, contudo traduzem maior

variabilidade e complexidade de parâmetros acústicos (Pabon & Plomp, 1988; Pabon, 1991; Bunton et al., 2007; Zhang e Jiang, 2008; Maryn et al., 2009; Moon et al., 2012). Aliás, quando a qualidade vocal é medida com base em escalas áudio-percetivas que contemplam parâmetros específicos, tal implica que a impressão global que o ouvinte capta de uma determinada voz pode ser decomposta em diferentes aspectos áudio percetivos distintos (Kent & Ball, 2000; Choi et al., 2012). Por outro lado, a correspondência entre os resultados destas duas dimensões não é nem óbvia nem direta, sendo ainda matéria de investigação e debate, o que denota a dificuldade clássica que existe em exprimir a acuidade auditiva humana através de modelos matemáticos (Kreiman & Gerratt, 1990; Rabinov et al., 1995; Shrivastav & Sapienza, 2003; Bhuta et al., 2004; Gregio et al., 2006; Maryn et al., 2009; Kreiman & Gerratt, 2011; Choi et al., 2012).

A análise dos parâmetros vocais alterados torna-se importante no processo de diagnóstico de uma perturbação da voz e condiciona toda a intervenção terapêutica (Awan & Roy, 2009; Camargo & Madureira, 2010; Ma & Yiu, 2011). Aquela pode ser avaliada de forma subjetiva (avaliação áudio-percetiva, Cf. Capítulo 2, Secção 2.1. e 2.2.) e/ou objetivamente com o auxílio de equipamentos de análise do método acústico (Cf. Capítulo 2, Secção 2.3. e 2.4.). Note-se que a qualidade vocal resulta da interação entre o sinal acústico da voz e o ouvinte; o sinal acústico *per se* não possui qualidade, mas evoca-a no avaliador/ouvinte (Kent & Ball, 2000). Por esta razão, as medidas acústicas são significativas a partir da sua correspondência ao que o ouvinte escuta (Gerratt & Kreiman, 1995; Kreiman & Gerratt, 1996; Kreiman & Gerratt, 2011; Shrivastav, 2011). Nos diferentes estudos de correspondência entre as medidas subjetivas e instrumentais, a percentagem de variância pode encontrar-se entre o ligeiro a moderado (Eskenazi et al., 1990; Kempster et al., 1991; Wolfe et al., 1991), com valores objectivos de concordância entre 49,9% (Wuyts et al. 2000) e 86,0% (Yu et al. 2001).

Vários autores investigaram a relação entre as medidas acústicas isoladas e a avaliação áudio percetiva. Tanto quanto nos foi dado a inferir, os seus resultados foram essencialmente inconclusivos pois revelaram correlações com um reduzido grau de

significado estatístico (Dejonckere, 1996; Giovanni, 1999; Heman-Ackah et al., 2002, 2003; Morsomme, 2001; Carding et al., 2004; Rabinov, 1995; Ma e Yiu, 2006; Kreiman & Gerrat, 2007; Patel & Shrivastav, 2007), o que indica que a análise dos parâmetros isolados não descreve o comportamento vocal e deve ser evitada (Michaelis et al., 1998; Buder, 2000; Behlau, 2001; Parsa & Jamieson, 2001; Batalla et al., 2004; Carmargo & Madureira, 2010; Kreiman & Gerratt, 2011). Destaca-se mais uma vez a importância da abordagem integrada, no sentido de considerar múltiplos recursos de avaliação disponíveis (Sader & Hannayama, 2004; Dejonckere, 2010; Ma & Yiu, 2011; Choi et al., 2012).

Alguns autores assumem que as alterações da qualidade vocal condicionam modificações nas características particulares do sinal acústico (Hiroto, 1967), embora existam razões empíricas para discordar desta afirmação já que a produção sonora do trato vocal não é nem linear, nem passiva (Buder, 2000; Dajer, 2010; Akpinar et al., 2011) e existem evidências consideráveis de que o processo áudio-percetivo utiliza múltiplas pistas acústicas para formar uma percepção (Klatt & Klatt, 1990; Repp, 1988; Shrivastav & Sapienza et al., 2003; Eadie & Doyle, 2005; Kreiman & Gerratt, 2011). Deste modo, uma modificação da qualidade vocal pode ser o resultado de múltiplos traços acústicos que interagem de forma complexa e, na maioria das vezes, não linear (Patel & Shrivastav, 2007). O desenvolvimento dos índices de perturbação ou das medidas de ruído deriva dos modelos de produção de fala (Fant, 1970) e não da sua percepção, pelo que relacionam indiretamente o modelo acústico com o auditivo. As medidas ou parâmetros de avaliação propostos pelo método acústico acabam, então, por contemplar apenas uma fração da totalidade das apreciações ou julgamentos passíveis ao ouvido humano (Eskenazi & Childers, 1990; Buder, 2000; Kent & Ball, 2000), o qual assume um caráter multidimensional e consegue aglutinar características áudio-percetivas (Parsa & Jamieson, 2001; Patel & Shrivastav, 2007; Harnisch et al., 2008; Kreiman & Gerratt, 2011).

Correlacionar as medidas acústicas com os parâmetros áudio-percetivos e, ainda, com os mecanismos fisiológicos da produção vocal é uma tarefa complexa (Kreiman et al., 1990; Eskenazi et al., 1990; Bhuta et al., 1991; Kreiman et al., 1993; Dejonckere, 1995;

Rabinov et al., 1995; Dejonckere & Lebacq, 1996; Frolick et al., 2000; Kent & Ball, 2000; Shrivastav & Sapienza, 2003; Batalla et al., 2004; Ma & Yiu, 2011; Choi et al., 2012). Isto porque se crê que uma dimensão ou parâmetro específico da qualidade vocal possui uma ou mais pistas acústicas para a sua percepção, as quais são variáveis de voz para voz, e ainda consoante o ouvinte que a caracteriza, tal como explanado no capítulo anterior desta dissertação.

Atualmente tenta-se que a conjugação entre medidas acústicas possibilite a quantificação de aspetos da produção vocal. É pela combinação de vários parâmetros objetivos que se obtém uma concordância mais forte com a análise percepitiva (Wolfe & Martin, 1997; Michaelis, 1997, 1998; Piccirillo, 1998; Klein, 2000; Wuyts, 2000; Yiu et al., 2000; Parsa & Jamieson, 2001; Yu et al., 2001, 2002; Hartl, 2003; Kent et al., 2003; Batalla et al., 2004; Murdoch, 2005; Patel & Shrivastav, 2007; Hakkesteegt et al., 2008; Dejonckere, 2010; Uloza et al., 2011).

Recentemente alguns autores começaram a propor o estudo de três medidas derivadas de processos de modelação periódica linear (Qi et al., 1999):

- a) SNR – *signal-to-noise ratio*;
- b) PA – *pitch amplitude*; e,
- c) SFR – *spectral flatness ratio*.

Tanto o PA como o SFR mostraram a possibilidade de diferenciação entre vozes normais e alteradas, em tarefa de vogal sustentada e conversação espontânea (Buder, 2000). Hillenbrand et al. (1996) e seguidores notaram que as medidas de *cepstrum* se correlacionam com as medidas percepitivas de Grau geral de alteração vocal e Soprosidade, de forma mais forte face às restantes medidas.

Tal como previamente mostrado, muitos estudos não permitem uma comparação entre resultados, devido à disparidade entre os casos analisados e os métodos usados para recolher as avaliações, tanto em termos de medidas acústicas, como de escalas percepitivas (Finizia et al., 1999; Bunton et al., 2007; Kreiman et al., 2007; Patel & Shrivastav, 2007; Uloza et al., 2011).

São 20 os estudos recentes mais significativos nesta temática:

Wolfe et al., 1995, caracterizaram quatro medidas acústicas (frequência fundamental, *jitter*, *shimmer* e HNR), num estudo com 20 falantes com vozes “normais” e 60 casos de pacientes disfónicos (lesões nodulares, paralissias de prega unilaterais e disfonias funcionais). Através de uma análise de regressão linear múltipla, a correlação entre os parâmetros acústicos e a análise áudio-percetiva foi de $r=0.56$. A análise fatorial da correlação apresentou $r=0.54$ entre o *shimmer* isolado e a avaliação percetiva.

Giovanni et al., 1996, associaram em 239 vozes disfónicas duas medidas de perturbação acústica (*jitter* e o *signal-to-noise ratio*) com duas aerodinâmicas (*voice onset time* e *glottal leakage*) que foram captadas e classificadas pelo sistema EVA®, o qual apresenta uma escala de severidade percetiva, em cinco níveis, desde o 0=normal a 4=severo. O tratamento estatístico com a análise de uma função de entrada direta discriminante revelou que a combinação das quatro medidas conseguia um grau de concordância com a classificação percetiva de 66,1%. Salientaram que esta concordância foi estatisticamente significativa para todos os níveis, exceto para o grau 1=ligeiro/intermitente de alteração vocal.

Michaelis et al., 1998, construíram um diagrama de rugosidade com base na combinação de fatores acústicos, analisado a partir de uma base de dados de 88 amostras eufónicas ou normais e 447 patológicas. Concluíram que a média de perturbação relativa (RAP) e o cociente de perturbação da amplitude (APQ) eram as duas variáveis acústicas que continham mais informação acerca da aperiodicidade de uma voz e, como tal, capaz de discriminar entre uma qualidade vocal normal de uma alterada.

Piccirillo et al., 1998, desenvolveram dois estudos na tentativa de construírem um índice multiparamétrico que correspondesse à severidade da disfonia. Assim, através de uma análise de regressão multivariada identificaram 14 medidas básicas capazes de distinguir entre vozes eufónicas/normais e disfónicas. Concluíram que a combinação entre quatro delas – estimativa de pressão sub-glótica, a gama de variação da

frequência, o cociente de ar medido nos lábios e o tempo máximo de fonação – se correlacionava moderadamente com o grau geral de alteração vocal (r de Pearson=0.58).

Wuyts et al., 2000, realizaram a análise multivariada de um *corpus* de 68 falantes com vozes “normais” e 319 indivíduos disfónicos, classificados percetivamente através da escala GRBAS. Deste estudo resultou a proposta de uma escala que faz a descrição da qualidade vocal através de provas de avaliação aerodinâmica e parâmetros da análise acústica – o DSI (*Dysphonia Severity Index*)¹⁸.

O *Dysphonia Severity Index* é definido como uma combinação ponderada de 4 medidas acústicas selecionadas por uma análise multivariada de 13 parâmetros. A análise discriminativa da combinação dos quatro parâmetros finais – *jitter*, tempo máximo de fonação da vogal /a/, pico de frequência e mínimo valor de intensidade – revelou uma concordância de 49,9% (193 de 387 sujeitos) com a avaliação percetiva. Embora tendo mostrado uma elevada utilidade clínica, as medidas subjacentes não medem necessariamente a qualidade vocal, mas sim focam-se num conjunto de medidas que covariam com a severidade global da alteração vocal.

Yu et al., 2001, fizeram a análise multiparamétrica de uma amostra de 63 vozes masculinas disfónicas e 21 falantes sem alterações vocais, avaliados percetivamente através do parâmetro G (Grau) da escala GRBAS, por um painel de juízes apoiado num *software* de apoio à decisão. Caracterizaram com o EVA® 10 parâmetros: frequência fundamental, *jitter*, intensidade, HNR, HNR ($f>1\text{kHz}$), coeficientes de Lyapounov, débito de ar oral, extensão vocal e tempo máximo de fonação, medidos com base na vogal /a/ sustentada e pressão sub-glótica pela repetição da sílaba /pa/. A análise discriminativa dos dados possibilitou a concordância significativa entre os julgamentos percetivos e seis dos diferentes parâmetros. Especificamente, a combinação frequência fundamental, HNR, coeficientes de Lyapounov, pressão sub-glótica,

¹⁸ A fórmula de cálculo do DSI = $0.13 \times \text{TMF} + 0.0053 \times \text{F0-High} - 0.26 \times \text{l-low} - 1.18 \times \text{Jitter}(\%) + 12.4$

extensão vocal e tempo máximo de fonação, obteve 86% (72 dos 84 sujeitos) de concordância com o painel de juízes.

Yu et al., 2002, realizaram um estudo semelhante ao anterior, exceto no corpus de análise – um total de 74 vozes femininas, 6 eufónicas/normais e 68 disfónicas. Os resultados revelam uma concordância entre avaliação percepção e acústica da seguinte magnitude:

- 64% para a classificação do parâmetros G(rau) com uma escala analógica;
- 88% para a mesma tarefa, mas com recurso a uma escala visual analógica discreta, com uma segmentação não linear.

Heman-Ackah et al., 2002, analisaram 38 amostras de voz de casos com paralisia unilateral de prega vocal, em fala encadeada (leitura de uma passagem (2^a) do “Rainbow Passage”) ou vogal sustentada. Usaram dois juízes com experiência que classificaram as amostras com a escala GRBAS. A análise de consistência interavalíadores revelou valores fortes para o G(Grau), R(Rugosidade) e S(Soprosidade) – parâmetros perceptivos que se mantiveram em análise e foram associados com medidas acústicas (MDVP da Kay Elemetrics): CPPS (*cepstral peak prominence smoothed*), sPPQ (*smoothed pitch perturbation quotient*), NHR (*noise-to-harmonic ratio*), APQ (*amplitude perturbation quotient*), RAP (*relative average perturbation*). Concluíram que o CPPS se correlaciona de forma forte mas inversa com o G (Grau) ($r=-0.86$, $r^2=0.74$, na fala; $r=-0.80$, $r^2=0.64$, na vogal) e S(Soprosidade) ($r=-0.71$, $r^2=0.51$, na fala; $r=-0.70$, $r^2=0.54$, na vogal). O CPPS correlaciona-se inversamente ($r=-0.50$, $r^2=0.25$) e o sPPQ positivamente ($r=0.40$, $r^2=0.16$) com a percepção de R (Rugosidade), embora a variância para cada uma destas correlações seja <25%. As medidas NHR, APQ, RAP e sPPQ correlacionam-se positivamente com o G (Grau) ($r=0.53$, $r^2=0.28$; $r=0.54$, $r^2=0.29$; $r=0.60$, $r^2=0.36$; $r=0.57$, $r^2=0.32$, respetivamente) e S (Soprosidade) ($r=0.54$, $r^2=0.30$; $r=0.52$, $r^2=0.27$; $r=0.54$, $r^2=0.30$; $r=0.53$, $r^2=0.28$, respetivamente). NHR, APQ e RAP não se correlacionaram com a percepção de R (Rugosidade).

Batalla et al., 2004, analisaram através de um modelo estatístico de regressão logística as amostras de 107 vozes de casos com Edema de Reinke (29 homens/78 mulheres),

classificadas percetivamente através da escala GRBAS, por dois terapeutas da fala com experiência em avaliação de vozes patológicas, em tarefa de produção sustentada da vogal /e/ (microfone a 10 cm de distância da boca, cabine insonorizada). Comparando-as com cinco medidas acústicas obtidas através do programa Dr. Speech da Tiger Electronics (frequência fundamental, *jitter*, *shimmer*, NNE, HNR) e a análise espectrográfica para especificar a classificação de Yanagihara, a presença de sub-harmónicos e a ausência de harmónicos no espectro. Os autores obtiveram equações que relacionaram cada um dos parâmetros percetivos da GRBAS com as medidas acústicas enumeradas. Ainda apontaram conclusões acerca da necessidade de recorrer às escalas de classificação percetiva nos casos de vozes extremamente perturbadas.

Butha et al., 2004, estudaram através de um modelo estatístico de regressão multivariada amostras de vozes de 37 pacientes (12 homens/25 mulheres) que classificaram percetivamente através da escala GRBAS, por um terapeuta da fala com pelo menos um ano de experiência, em tarefas de conversação espontânea preferencialmente (microfone a 20cm distância da boca e ligeiramente à direita), e 19 medidas acústicas do MDVP da Kay Elemetrics. O coeficiente de correlação (R^2) para a análise do parâmetro G (*Grade*) foi 0.43, para o parâmetro R (*Roughness*) foi 0.14, para o parâmetro B (*Breathiness*) foi 0.32, para o parâmetro A (*Astheny*) foi 0.35. O S (*Strain*) não se correlacionou com nenhuma medida da avaliação acústica. Concluíram ainda que:

- VTI (*Voice Turbulence Index*) correlacionou-se com o G (*Grade*), $p=0.001$;
- NHR (*Noise-Harmonic Index*) correlacionou-se com G (*Grade*), $p=0.007$ e R(*roughness*), $p=0.02$;
- SPI (*Soft Phonation Index*) correlacionou-se com G (*Grade*), $p=0.04$, B (*Breathiness*), $p=0.01$ e A (*Astheny*), $p=0.04$.

Eadie & Doyle, 2005, estudaram a concordância entre seis medidas espectrais de longo-tempo – LTAS (1. *overall spectral tilt*; 2. *voiced frames of spectral tilt* e medida de ruído glótico; 3. HNR no domínio do tempo (FHNR) e medidas modeladas de LP (*linear prediction*); 4. LP-SNR; 5. PA – *pitch amplitude*; 6. SFR – *spectral flatness ratio*) e a classificação áudio-percetiva do grau geral de perturbação vocal e “agradabilidade”

(através de uma escala visual analógica – VAS – de 200mm) realizada por 12 alunos de Terapia da Fala, de uma amostra de leitura de uma passagem do texto “Rainbow Passage” realizada por 30 adultos disfónicos e 6 eufónicos/normais. Concluíram que as medidas acústicas (PA e HNR) se correlacionavam em 48% com o grau geral de perturbação vocal e em 40,2% com a “agradabilidade”. A medida acústica isolada com a associação mais forte à avaliação percetiva foi a PA – *pitch amplitude*, correspondente a 33% da variância ($r=0.573$, $F(1,28)=13,847$, $p<0.001$) face ao Grau geral de disfonia.

Eadie & Baylor, 2006, analisaram os resultados da classificação percetiva através da escala visual analógica (VAS) – parâmetros Grau, Rugosidade e Soprosidade – de 36 vozes (21 disfónicas e 6 normais) a produzir uma vogal sustentada e em leitura (fala encadeada), levada a cabo por 16 avaliadores não treinados, a quem foi proporcionado um período de 2 horas de treino com 15 vozes e 18 estímulos-âncora ao longo do teste. A análise acústica foi realizada em três aplicações de *software* (MDVP, da Kay Elemetrics, o IVANS da Avaaz Innovations (1998 *in* Eadie & Baylor, 2006) e um programa obtido de Hillenbrand et al., 1996, para determinar o *cepstral peak prominence* (CPP) e o *CPP smoothed* (CPPS)). Concluíram que o CPPS era o melhor preditor do Grau geral de alteração vocal (67 a 71% em vogais sustentadas; 74 a 77% em fala encadeada) e de Soprosidade em amostras de fala encadeada (58 a 68%); o *shimmer* foi o parâmetro que mais se associou à Rugosidade (entre 39 a 44%), em amostras de fala encadeada.

Ma & Yiu, 2006, caracterizaram a correspondência (especificamente, a capacidade de predição) entre o Grau geral de alteração vocal e um conjunto de provas de avaliação aerodinâmica (no programa Aerophone II da Kay Elemetrics), o *voice range profile* (fonetograma do programa Swell) e as medidas de perturbação acústica (medidas no MDVP da Kay Elemetrics: frequência fundamental média, *jitter*, *shimmer* e HNR) em provas de leitura (5 repetições) de uma frase em língua Cantonesa constituída apenas por plosivas bilabiais e vogais e do tempo máximo de fonação para as vogais /a/, /i/ e /u/ (5 tentativas/cada). Quatro alunos finalistas de Terapia da Fala caracterizaram o parâmetro G (Grau geral da escala GRBAS) de 110 vozes patológicas e 41 amostras de

eufonia (acrescido de uma taxa de repetição de 25%), numa escala de EAI (*Equal-Appearing Interval*) com onze níveis, após terem um treino prévio com 25 estímulos vocais não pertencentes à amostra de teste. Beneficiavam ainda de “âncoras” externas de vozes sintéticas, com diferentes graus de alteração da Soprosidade e Rugosidade. Concluíram que o tempo máximo de fonação, o pico de pressão intraoral na consoante-vocal /pi/, a área do fonograma e o *jitter* prediziam corretamente 67,3% (103 de 153 sujeitos) dos níveis de Grau geral de alteração vocal. As amostras classificadas como normais e severamente alteradas foram associadas às medidas acústicas de forma mais precisa (82,5 e 71,9%, respetivamente) em comparação com as vozes ligeira a moderadamente disfónicas (67,9 e 36,0%, respetivamente).

Martens et al., 2007, estudaram a classificação áudio-percetiva de seis diferentes avaliadores (com 9 a 20 anos de experiência), de um *corpus* de 70 vozes moderadamente alteradas durante a produção de vogal sustentada e leitura de uma frase em alemão, através da escala GRBAS, na sua forma original de classificação em 4 níveis (de 0 a 3 pontos). Na experiência usaram dois momentos de classificação, em que no segundo era acrescida uma pista visual (fonograma) de apoio à avaliação áudio-percetiva. A análise acústica foi realizada pelo MDVP (Kay Elemetrics), da porção estável da vogal sustentada, com a caracterização das medidas de *jitter*, *shimmer* e HNR. Concluíram que o uso da pista visual não trouxe diferenças significativas à correlação entre os parâmetros percetivos e as medidas acústicas. Encontraram valores da correlação de Spearman entre -0.4 e 0.7. O valor mais alto foi estabelecido para o *Jitter* com o parâmetro G-Grau geral de alteração vocal (0.7); e para o *Shimmer* relacionaram-se o G-Grau geral (0.58) e o R-Rugosidade (0.57). Neste estudo o HNR não se relacionou com nenhum dos parâmetros percetivos estudados.

Awan & Roy, 2009, analisaram 176 vozes (porção central de vogal sustentada /ɑ/) de mulheres com disfonia por tensão muscular (pré e pós-terapia), classificando-as através de medidas acústicas de base temporal (*shimmer*, desvio padrão da F0) e espectral (ratio de energia espectral-DFTR-discrete Fourier transformation ratio); ratio da amplitude atual da CPP (*cepstral peak proeminence*) face à amplitude esperada

(CPP/EXP) – que faziam parte da equação do *Predicted Dysphonia Severity*¹⁹ (PDS, Awan & Roy, 2006). Com um programa computorizado específico de classificação perceptiva da severidade da alteração vocal, 10 alunas de mestrado em Terapia da Fala quantificavam áudio-perceptivamente as amostras. Classificaram-nas numa escala com sete níveis de EAI, divididas por duas sessões (88 vozes cada). Estas eram separadas temporalmente por 48 horas. Antes do teste tinham 20 minutos de explicação dos procedimentos e definição de conceitos. Ao longo da prova podiam recorrer a um estímulo-“âncora” com uma voz “normal”. Concluíram que: a percepção da severidade da alteração vocal se correlacionava de forma forte com o PDS ($r=0.906$, $R^2=0.82$, $p<0.0001$). Isoladamente, cada medida da PDS também apresentou uma concordância entre a percepção de severidade de alteração da voz forte e inversa com INVSQRTSIG ($r=-0.819$, $R^2=0.67$) e CPP/EXT ($r=-0.754$, $R^2=0.80$), e forte e positiva com o LOGSHIM ($r=0.814$, $R^2=0.77$). A correlação com o DFTR foi moderada ($r=-0.511$, $R^2=0.82$). Note-se que a medida apresentada por estes investigadores revelou-se robusta na distinção entre vozes ligeira e moderadamente alteradas (níveis intermédios da escala perceptiva usada), assim como a analisar as amostras vocais extremamente alteradas.

Maryn et al., 2009b, levaram a cabo uma meta-análise com a revisão da literatura publicada acerca da relação entre as medidas acústicas e a classificação áudio-perceptiva da qualidade vocal global/geral (*overall voice quality ou grade ou severity of voice disorder ou severity of dysphonia ou overall abnormality*). Assim, analisaram um *corpus* de 25 estudos (cerca de 21 com medidas de vogal sustentada e 7 de fala encadeada; 3 apresentavam os dois tipos de amostra vocalica) os quais expunham um total de 87 medidas ou marcadores acústicos (os estudos com vogal sustentada identificavam, individualmente, 69; os artigos com fala encadeada associavam-se a 26 parâmetros acústicos). A medida de corte para a determinação do grau forte de correlação foi o valor de coeficiente de correlação $\bar{r_w} \geq 0.60$. Deste modo, a percepção de alteração vocal associava-se:

¹⁹ A fórmula de cálculo do PDS= $6.923+1.394 (\text{LOGSHIM}) - 1.002 (\text{INVSQRTSIG}) - 0.030 (\text{DFTR}) - 0.19 (\text{CPP/EXP})$

- a) Em amostras de vogal sustentada com – r de Pearson na *autocorrelation peak* $\overline{r_w} \geq 0.74$, amplitude do *pitch* $\overline{r_w} \geq 0.75$, *spectral flatness* do sinal residual após análise *LPC* $\overline{r_w} \geq 0.69$ e *SCPP (smoothed cepstral peak prominence)* $\overline{r_w} \geq 0.62$;
- b) Em tarefas de avaliação de fala encadeada com – *signal-to-noise ratio* de Qi $\overline{r_w} \geq 0.69$, *cepstral peak prominence* $\overline{r_w} \geq 0.88$ e *SCPP (smoothed cepstral peak prominence)* $\overline{r_w} \geq 0.88$.

Concluem que estas seis medidas acústicas são as mais promissoras para o estabelecimento de correlações entre o parâmetro áudio-percetivo de qualidade vocal geral e o método acústico.

Em 2011, Uloza e colaboradores caracterizaram 148 vozes (69 “saudáveis” e 79 “patológicas”) através de um conjunto de 10 diferentes grupos de medidas acústicas e analisaram a sua efetividade na discriminação entre voz saudável *versus* voz patológica (lesões nodulares ou lesões difusas), também classificada por quatro juízes de avaliação áudio-percetiva (com mais de 15 anos de experiência). Os juízes tinham – antes do estudo – um treino com 30 vozes (10 de cada classe). Concluíram que o *software* automático de categorização das vozes obteve uma taxa de classificações corretas de 92% (discriminação voz saudável *versus* patológica) e de 90% quando a classificação se referia às classes “saudável”, “patológica-nodular” e “patológica-difusa”. Estas taxas para os juízes áudio-percetivos foram de 74% e 60%, respetivamente.

Em 2012a, Choi et al. estudaram 23 vozes de casos com sulcus vocais de diferentes graus e compararam-nas com a mesma quantidade de controlos, emparelhadas²⁰ por sexo e idade. As medidas acústicas usadas foram: a dimensão de correlação (D_2), *plots* de fase e medidas de perturbação do domínio do tempo (*jitter*, *shimmer* e *SNR-Signal-to-Noise Ratio*). A escala de avaliação áudio-percetiva reportava-se ao parâmetro G-

²⁰ As amostras dizem-se emparelhadas quando para a sua constituição são usados os mesmos sujeitos experimentais ou os mesmos materiais a analisar em distintas situações, tendo como base algum critério unificador dos elementos (por exemplo, amostras em que a mesma variável é medida antes e depois de determinado tratamento, nos mesmos sujeitos) (Maroco, 2010).

Grau Geral de qualidade vocal, classificado por três alunos de um programa doutoral de terapia da fala, através de uma EAI (*Equal Appearing Scale*) de 7 pontos. A amostra estudada correspondia à edição de 1 segundo de um segmento estável da vogal / a /, após extração do início e final da produção. A correlação mais forte entre as duas formas de avaliação foi encontrada para a medida SNR ($r=-0.88$) e a dimensão de correlação (D2), com $r=0.68$. Após remoção dos sinais de tipo 2 e 3, e com um tamanho amostral muito inferior, a correlação foi para o *jitter* $r=0.52$, para o *shimmer* $r=-0.25$, para SNR $r=-0.44$ e, por fim, para D2 $r=0.37$.

Moers et al. (2012) compararam a avaliação áudio-percetiva realizada por 5 juízes através da escala RBH Alemã e o parâmetro “grau geral”, a 73 casos com alterações vocais, e as medidas *cepstral peak prominence* (CPP) e a *smoothed cepstral peak prominence* (CPPS). Concluíram que os resultados mais fortes de correlação entre os parâmetros áudio-percetivos e acústicos tiveram lugar para a experiência com texto, em vez da com vogal sustentada. Acrescentam também que os parâmetros “soprosidade”, “rouquidão” e “grau geral” são melhor mapeados pelas medidas do *software*, e que o parâmetro rugosidade se assume como uma “super-classe” aglutinadora dos aspectos anteriores.

De seguida apresentam-se os quadros-síntese das características metodológicas dos estudos analisados:

Quadro 1: Características metodológicas da avaliação áudio-percetiva nos 21 estudos com vogais sustentadas usados na meta-análise de Maryn et al., (2009b), um de Uloza et al., (2011), um de Choi et al. (2012a) e um de Moers et al., (2012).

Autor/Fonte bibliográfica	Sujeitos			Amostra de Voz		Avaliação Áudio-Percetiva				
	N	Patolog	Total	Vogal	Duração	Nº Juízes	Escala	Parâmetro	Consistência Intra-Avaliador	Consistência Inter-Avaliador
Kojima et al., 1980	28	30	58	/ a /	NA	5	EAI(4)	Rugosidade	NA	NA
Yumoto et al., 1984	0	87	87	/ a /	3s	8	EAI(4)	Rugosidade	NA	0.51-0.79 Sp
Hirano et al., 1986	0	68	68	/ e /	NA	NA	EAI(4)	G, Grau	NA	NA
Prosek et al, 1987	0	90	90	/ a /	2s	9	EAI(7)	Severidade da Alteração Vocal; Rugosidade	0.90 Pe	0.82 Cr
	16	44	60			14			NA	NA
Wolfe & Steinfatt, 1987	0	51	51	/ a / e /i/	1s	8	EAI(7)	Severidade da Disfonia; Rugosidade	89% Ag	0.95 Cr
Feijoo & Hernández, 1990	64	57	121	/ e /	NA	4	EAI(4)	G, Grau	77,48% Ag	98.35% Ag
Kreiman et al., 1990	0	18	18	/ a /	1.67s	10	EAI(7)	Anormalida de geral	NA	NA
Wolfe et al., 1995	20	60	80	/ a /	1s	22	EAI(7)	Severidade geral	0.99 Cr	0.98 Cr
Dejonckere et al., 1996	0	943	943	/ a /	2s	2	EAI(4)	G, Grau	0.51 Co	0.87 Sp
Dejonckere & Wieneke, 1996	0	28	28	/ a /	0.1s	2	EAI(5)	Severidade Geral da Rugosidade	NA	NA
De Bodt, 1997	98	634	732	/ a /	3s	1	EAI(4)	G, Grau	NA	NA
Plant et al., 1997	0	26	26	/ i /	2s	3	EAI(5)	Qualidade Vocal Geral	0.86 NA	NA

Wolfe & Martin, 1997	0	51	51	/ a / e / i /	1s	11	EAI(7)	Rugosidade	76% Ag	0.95 Cr
Wolfe et al., 1997	0	51	51	/ a / e / i /	1s	18	VAS	Rugosidade	0.80 Pe	0.94 Cr
Wolfe et al., 2000	0	20	20	/ a /	1s	11	EAI(7)	Anormalida de	0.81 Pe	0.98 Cr
Yu et al., 2001	21	63	84	/ a /	2s	6	EAI(4)	G, Grau	Cons	Cons
Heman-Ackah et al., 2002	0	14	14	/ a /	1s	2	EAI(4)	G, Grau	NA	0.83 Pe
Halberstam, 2004	0	60	60	/ a /	1s	2	EAI(7)	Rugosidade	0.89 NA	0.91 Cr
Eadie & Baylor, 2006	3	9	12	/ a /	1s	16	VAS	Severidade geral	0.82-0.95 Pe	0.72-0.83 Pe
Gorham-Rowan & Laures-Gore, 2006	0	28 ^{ym}	28 ^{ym}	/ a /	1s	10	FMMEP	Rugosidade	-0.32-0.86 Pe	0.80 Cr
	0	28 ^{ew}	28 ^{ew}							
	0	28 ^{em}	28 ^{em}							
Yu et al., 2007	38 ^w	308 ^w	308 ^w	/ a /	2s	4	VAS	G, Grau	NA	NA
	20 ^m	141 ^m	141 ^m							
Uloza et al., 2011	69	79	148	/ a /	2.4s	4	EAI(3)	Normal/Pat ológico	NA	NA
Choi et al., 2012 ^a	23	23	46	/ a /	1s	3	EAI(7)	G, Grau	0.94; 0.93; 0.93 (ICC) 79%; 80%; 80% Ag	0.84-0.92 Pe 65% Ag
Moers et al., 2012	0	73	47	/ e /	0.5-1s	5	EAI(4) e VAS	RBH Alemã, Grau	NA	0.89-0.93 Cr

N= número de sujeitos normais; Patolog= número de sujeitos patológicos ou com disfonia; Total= número total de sujeitos; ^{ym}=homens jovens; ^{ew}=mulheres idosas; ^{em}=homens idosos; ^m=homens; ^w=mulheres

NA = a informação não está disponível no manuscrito original

EAI=*equal-appearing interval scale*, e entre parêntesis o número de pontos da escala; VAS=*visual analogue scale*; FMMEP=*free modulus magnitude estimation paradigm*
 Sp=Coeficiente de correlação de Spearman; Pe=Coeficiente de correlação de Pearson; Co=Coeficiente de Correlação kappa de Cohen; Cr=Coeficiente de correlação alfa de Cronbach; Ke=Coeficiente de Concordância de Kendall; Ag=Percentagem de concordância/consistência entre os juízes; Cons=Consenso entre os juízes, sem medida quantitativa de consistência; ICC=*Intraclass Correlation Coefficients*.

Quadro 2: Características metodológicas da avaliação áudio-percetiva dos 7 estudos com fala encadeada usados na meta-análise de Maryn et al., (2009b) e um de Moers et al., (2012).

Autor/Fonte bibliográfica	Sujeitos			Amostra de Voz	Avaliação Áudio-Percetiva				
	N	Patolog	Total		Nº Peritos	Escala	Parâmetro	Consistência Intra-Avaliador	Consistência Inter-Avaliador
Askenfelt & Hammarberg, 1986	0	41	41	Segmentos de voz, 40s de leitura de história	6	EAI(6)	Qualidade Vocal Geral	0.86-0.98 Pe	NA
Qi et al., 1999	0	87	87	1ª e 2ª frase do “Rainbow Passage”	5	VAS	Qualidade Vocal Geral	0.93-0.96 Pe	0.97 Cr
Heman-Achah et al., 2002	0	18	18	2ª frase do “Rainbow Passage”	2	EAI(4)	G, Grau	NA	0.83 Pe
Halberstam, 2004	0	60	60	12 s do “Rainbow Passage”	2	EAI(7)	Rugosidade	0.93 NA	0.97 Cr
Eadie & Doyle, 2005	6	24	30	2ª frase do “Rainbow Passage”	12	DME	Qualidade Vocal Geral	0.69 Pe	0.97 Cr
Eadie & Baylor, 2006	3	9	12	2ª frase do “Rainbow Passage”	16	VAS	Qualidade Vocal Geral	0.80-0.97 Pe	0.84-0.91 Pe
Ma & Yiu, 2006	41	112	153	/ ba ba da bo /	4	EAI(11)	G, Grau	≥ 0.90 Pe	0.84-0.91 Pe
Moers et al., 2012	0	73	47	1ª frase de um texto em Alemão	5	EAI(4) e VAS	RBH Alemã, Grau	NA	0.89-0.93 Cr

N= número de sujeitos normais; Patolog= número de sujeitos patológicos ou com disfonia; Total= número total de sujeitos

NA = a informação não está disponível no manuscrito original

EAI= *equal-appearing interval scale*, e entre parêntesis o número de pontos da escala; VAS=*visual analogue scale*; DME= *direct magnitude estimation*

Pe=Coeficiente de correlação de Pearson; Cr=Coeficiente de correlação alfa de Cronbach.

Em suma, atualmente foram encontrados mais de 300 termos diferentes para descrever a qualidade vocal dos falantes, contudo apenas uma pequeniníssima porção é amplamente aceite e conhecida por parte dos clínicos e investigadores da área. As metodologias subjetivas são, provavelmente, as mais antigas na classificação da qualidade vocal, e talvez o peso do seu uso ao longo dos séculos – sem grandes modificações – tenha feito com que se torne aceite e representativa da avaliação da voz humana (Kreiman & Gerratt, 2000 e 2011). Porém, a ubiquidade e consolidação da era informática e a presença fácil de computadores nas clínicas onde são acompanhados os casos com alterações da voz, fez com que o método acústico também se tornasse popular, especialmente quando as medidas acústicas são obtidas de forma rápida e imediata, com base no sinal de voz (Maryn et al., 2009b; Olszewski et al., 2011; Uloza et al., 2011). São várias as medidas propostas e testadas que quantificam a qualidade da voz com base em medições da mudança de alguns parâmetros relacionados com a sua produção, porém, poucas são encaradas como universais na descrição da sua qualidade (Maryn et al., 2009a,b; Ma & Yiu, 2011; Moers et al., 2012).

3.2. ESTUDOS COM RECURSO A MEDIDAS ACÚSTICAS OU A AVALIAÇÃO ÁUDIO-PERCETIVA, ASSOCIADOS À AVALIAÇÃO LARÍNGEA E/OU COM ALTERAÇÕES VOCais ESPECÍFICAS

Pretende-se, nesta secção, rever as principais investigações que servem de referência para os clínicos e pesquisadores dedicados à análise da voz humana, as quais se debruçam sobre estudos que recorrem a medidas obtidas através do método acústico ou a parâmetros áudio-percetivos. Porém, não as comparam entre si, mas correlacionam-nas com a avaliação visual da anatomo-fisiologia laríngea, ou ainda que estabelecem correspondências de acordo com alterações vocais específicas.

Yumoto et al., 1982, avaliou o parâmetro HNR (*Harmonic-to-Noise Ratio*) em amostras de vozes normais (n=18), pré e pós-cirurgia a lesões de massa (n=18), tendo concluído que:

- a) Os valores da medida acústica em análise nas vozes normais se encontra entre 7.0 e 17.7dB;
- b) 15 das 18 amostras de qualidade vocal alterada no pré-operatório – 84.6% - estavam fora deste intervalo;
- c) 95.7% das vozes no pós-operatório incluíam-se na gama de normalidade prevista, o que indicava a robustez da medida HNR como forma de medir a eficácia do tratamento e classificação distintiva entre voz normal e com alterações.

Finizia et al., 1999, analisaram/compararam 3 grupos – 12 sujeitos a laringectomias; 12 acompanhados em Radioterapia; e 10 sem alterações vocais – em termos de medidas acústicas (F0 e perturbação absoluta da F0); medidas temporais (TMF, velocidade de fala) e avaliação percetiva por 15 juízes, relativamente à “inteligibilidade da fala”, “qualidade vocal” e “aceitabilidade da fala”. Encontraram diferenças entre as várias avaliações implementadas e os grupos de sujeitos estudados. Houve uma correlação fraca ($=0.59$), embora estatisticamente significativa ($p<0.01$), entre a perturbação absoluta da F0 e a percepção de “qualidade vocal”.

Dogan et al., 2007, estudaram 40 casos de pacientes com asma a tomar corticosteroides e fizeram o seu emparelhamento quanto ao género com a mesma quantidade de não asmáticos. Analisaram percetivamente a voz através da escala GRB classificada por 4 juízes e 4 medidas acústicas (frequência fundamental, *jitter*, *shimmer* e HNR). Encontraram diferenças estatisticamente significativas entre os grupos para o HNR (nas mulheres), *Jitter* (nas mulheres) e *Shimmer* (nos homens e mulheres), o que correlacionaram com a percepção de vozes mais soprosas e rugosas dos asmáticos. 62,5% dos casos foram classificados no grau moderado a severo de disfonia.

Dedivitis et al., 2008, avaliaram 20 casos de pacientes sujeitos a laringectomias fronto-laterais, após pelo menos 12 meses da cirurgia, através da escala GIRBAS e a análise acústica com o MDVP (*Multi-Dimensional Voice Program* da Kay Elemetrics). Concluíram que todos os pacientes apresentavam algum Grau de disfonia (G=Grade), com alteração mais marcada da Rugosidade e Tensão. A Soprosidade foi identificada em 50% da amostra. O parâmetro “Astenia” foi o único não classificado pelo avaliador. As medidas acústicas elencadas foram a frequência fundamental (F0) (212,43Hz), *Jitter*(%) (4,42%), *Shimmer*(%) (12,08%), PPQ (2,95), APQ (9,38) e NHR (dB) (0,36dB), estando todas com médias acima do esperado para o software usado.

Boucher, 2008, estudou o efeito de uma tarefa de esforço vocal (leitura durante 3-5 minutos, em forte intensidade (>74 dBA), num total de 50 repetições) realizada ao longo de 12 a 14 horas, por 5 homens e 2 mulheres, sobre vinte medidas acústicas de frequência fundamental, perturbação da frequência, da amplitude, tremor e ruído no MDVP (*Multi-Dimensional Voice Program* da Kay Elemetrics). As análises estatísticas não revelaram qualquer associação estatisticamente significativa entre os parâmetros acústicos e as estimativas de fadiga muscular (análise do cricoaritenoideu lateral com Eletromiografia).

Após a revisão e análise da bibliografia descrita neste Capítulo 3 reforça-se a noção de que há muita dificuldade em associar diretamente as medidas acústicas às classificações áudio-percetivas. Por definição, a qualidade de voz é um fenómeno percetivo (Kreiman & Gerratt, 2011; Ma & Yiu, 2011; Shrivastav, 2011). Assim,

qualquer tentativa para classificar a voz humana deve, por um lado, usar diretamente os julgamentos dos ouvintes e, por outro, usar medidas do método acústico. Apesar de ser cada vez mais fácil aceder a programas de *software* de análise acústica, nem todas as medidas oferecidas têm uma representatividade clínica eficiente e robusta (Karnell et al., 1995; Buder, 2000; Maryn et al., 2009a,b; Schindler et al., 2012). Note-se, ainda, que a capacidade de avaliação áudio-percetiva incorre em muitos viés, os quais também condicionam negativamente o resultado final apresentado (Kreiman et al., 1993; Awan & Lawson, 2009; Kreiman & Gerratt, 2011).

Capítulo 4 – METODOLOGIA

4.1 MATERIAIS E MÉTODOS DE RECOLHA DOS DADOS

4.1.1 JUÍZES

Dez terapeutas da fala (três Brasileiros e sete Portugueses) especialistas/peritos em avaliação e intervenção terapêutica com voz foram recrutados por conveniência como juízes para a escuta, análise e classificação áudio-percetiva (*Cf.* Tabela 6). O contacto foi estabelecido pela autora e confirmado através de e-mail e/ou pessoalmente.

<i>Perito</i>	<i>País de Atividade</i>	<i>Género</i>	<i>Atividade Principal</i>
1	Portugal	Masculino	Docência/Clínica VP
2	Portugal	Masculino	Clínica
3	Portugal	Feminino	Docência/Clínica VP
4	Portugal	Masculino	Clínica
5	Portugal	Masculino	Docência
6	Portugal	Masculino	Docência/Clínica VP
7	Portugal	Feminino	Docência/Clínica
8	Brasil	Feminino	Docência/Clínica
9	Brasil	Feminino	Docência/Clínica
10	Brasil	Feminino	Docência/Clínica

Tabela 6: Caracterização dos peritos na avaliação áudio-percetiva através da escala GRBAS

A seleção foi baseada nos seguintes critérios: anos de experiência profissional (>6 anos) (Sofranko & Prosek, 2012) a nível hospitalar ou com voz profissional (VP) e/ou com responsabilidades Docentes na área da Voz, com conhecimento e uso da escala GRBAS (Webb et al., 2004). A distribuição por géneros é equitativa (50%). A média dos anos de experiência é 12,6, variando entre 6 e 26 anos. 7 peritos estavam na classe entre 6-15 anos; os restantes 3 possuíam mais de 16 anos de prática clínica. Foram critérios de inclusão a não referência a antecedentes de alterações da linguagem e/ou fala, assim como historial de hipoacusia ou perturbação da audição à data da realização da avaliação.

4.1.2 AMOSTRAS VOCais

Foram selecionadas 90 vozes de um banco de dados resultante dos últimos 10 anos de experiência profissional da autora da dissertação (2000-2010; após concedida a autorização hospitalar). Desta amostra, 20 vozes foram consideradas normais pelo facto de – numa avaliação preliminar²¹ – todas assumirem o grau 0 na escala GRBAS (*Cf.* Capítulo 2, Secção 1.1.); e, 70 possuíam algum grau de alteração (em distintos graus de severidade, de ligeiro a extremo), para a ilustração do tipo e severidade de cada um dos cinco parâmetros da escala GRBAS, classificados numa escala analógica visual (VAS – *Visual Analogue Scale*) de 100mm. A distribuição por sexo foi realizada da seguinte forma: 28% de vozes masculinas (n=25) e 72% de vozes femininas (n= 65). Esta distribuição percentual corresponde à representatividade de cada um destes géneros no banco de dados da prática clínica da autora da dissertação. Todos os casos eram adultos, com idade superior a 18 anos.

A ordem de apresentação dos estímulos vocais foi determinada de forma aleatória (Vertigan et al., 2008), para evitar os efeitos de familiaridade. Às 90 vozes diferentes acrescentaram-se 10 repetidas (10% do total).

As gravações das vozes seguiram o mesmo protocolo de recolha de amostras, o que implicou: uma frequência de amostragem de 44100Hz, e uma resolução por amostra

²¹ Realizada pela autora da dissertação.

de 16bits, utilizando um microfone de mesa da marca Philips SBC ME 400, unidirecional (cardióide), numa sala com um nível de ruído inferior a 40dB SPL, embora não acusticamente tratada. A distância do microfone à boca foi fixa a 10 cm, tendo sido pedido ao paciente – após modelo – para produzir de forma sustentada e confortável a vogal [α] (Heman-Ackah et al., 2002; Vertigan et al., 2008; Awan & Lawson, 2009; Awan & Roy, 2009; Hamdam et al., 2009; Olszewski et al., 2011), durante pelo menos 5 segundos (Smits et al., 2005; Awan & Lawson, 2009; Awan & Roy, 2009), em duas tentativas. A última, com o orador em pé, foi aquela usada para estudo. Estas gravações fazem parte da rotina de avaliação vocal da consulta hospitalar a que pertence a autora da dissertação, recolhidas através do software *Dr. Speech®*, versão 4.0 (Tiger Electronics). Foi analisada a porção de sinal a partir do 2º segundo da amostra (Heman-Ackah et al., 2002; Smits et al., 2005; Vertigan et al., 2008; Awan & Roy, 2009; Moers et al., 2012), segmentada no software Praat e considerada a zona de espectro mais estável para este estudo.

4.1.3 ESCALA DE AVALIAÇÃO ÁUDIO-PERCEITIVA

A escala de avaliação áudio-perceitiva usada foi a GRBAS (Hirano, 1981), adotada após análise das escalas apresentadas no Capítulo 2, secção 2.1. desta dissertação, tendo em consideração a reprodutibilidade mais forte de parâmetros descritos nos estudos enumerados (Dejonckere et al., 2001; Webb et al., 2004). A descrição de cada parâmetro é apresentada na Tabela 7:

Parâmetros	Definição
<i>G – Grade</i> (Grau Geral)	Grau de alteração vocal – impressão global da voz, identificação percetiva do grau de alteração vocal, como um todo.
<i>R – Roughness</i> (Rugosidade)	Irregularidade (aperiodicidade) na vibração das pregas vocais, sensação de “rugosidade” (<i>raucité</i> em Francês) que corresponde a flutuações irregulares da F0 e/ou da amplitude do som glótico. A voz é percecionada com ruídos imprevistos que mascaram os harmónicos, ou surgem sub-harmónicos. Típico em lesões de massa nas pregas vocais.
<i>B – Breathiness</i> (Soprosidade)	Relativo a turbulência audível, escape de ar importante através da glote, sensação de ar na voz, que corresponde fisiologicamente e de forma proporcional à fenda glótica, com voz de <i>loudness</i> reduzido.
<i>A – Astheny</i> (Astenia)	Astenia, fraqueza vocal, falta/perda de potência na voz relacionada com intensidade fraca, energia vocal reduzida, harmónicos pouco definidos e pouco elevados. Característico de alterações neuro-motoras ou

	deterioração vocal típica do avanço da idade.
S – Strain (Tensão)	Tensão, impressão de estado hiperfuncional – esforço vocal por aumento da adução glótica, geralmente inerente ao aumento da atividade da musculatura extrínseca da laringe, com elevação da posição desta, frequência anormalmente aguda (ruído nas frequências altas do espectro e/ou riqueza de harmónicos agudos marcados).

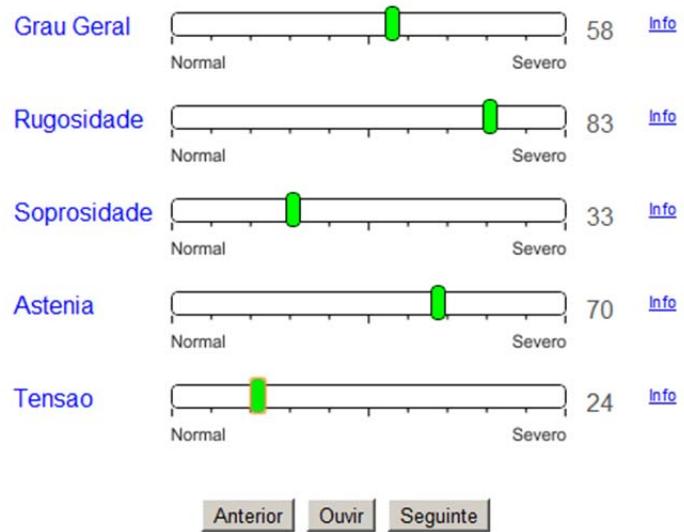
Tabela 7: Definição dos parâmetros áudio-percetivos da escala GRBAS (com base em Hirano, 1981; Dejonckere et al., 1996; Dejonckere et al., 2001; Webb et al., 2004)

Estas definições foram facultadas aos juízes e são, então, consideradas “âncoras” ou pistas escritas à classificação áudio-percetiva.

A avaliação áudio-percetiva das 100 vozes foi apresentada numa página Web, segundo uma ordem aleatória, escutada e cotada seguindo a mesma sequência por um painel de dez juízes. A página Web consiste num formulário por cada voz a ser avaliada, com um sistema de pré alocação do ficheiro de áudio a ser ouvido e um conjunto de cinco cursores interativos, reguláveis – um para cada parâmetro GRBAS – numa escala analógica visual (VAS – *Visual Analogue Scale*), representada por uma régua com 10cm de extensão (sendo que quanto mais à direita, mais alterada será a qualidade vocal), de forma a que a experiência de avaliação de cada parâmetro seja próxima da analógica (*Cf. Figura 13*). Cada estímulo podia ser escutado repetidamente (aspeto que também ficou registado), de acordo com a necessidade de cada juiz para realizar a classificação de cada um dos cinco parâmetros percetivos. Quando todos os campos tivessem sido classificados, o avaliador poderia corrigir qualquer das respostas dadas, avançar ou retroceder na página.

A resposta final ficou registada automaticamente numa base de dados SQL construída para o efeito, e com mecanismos que asseguraram que no caso da avaliação das vozes não ser feita em uma só sessão, não se perderiam as avaliações já efetuadas nem a informação de quais as vozes que já tinham sido avaliadas e as que ainda não tinham. Estes mecanismos consistiram também numa salvaguarda no caso de existirem

problemas com o serviço de Internet, a fim de não se perderem as avaliações já feitas no caso de perca de ligação no decorrer da avaliação.



É necessário clicar em "Ouvir" para poder avaliar a voz.

Voz: 40/100

Figura 13: Página Web de avaliação áudio-percetiva

Durante o desenvolvimento da página Web deparamo-nos com problemas de compatibilidade entre tecnologias e os *browsers* existentes. Assim, analisamos várias tecnologias²² e tomamos a decisão de utilizarmos para o efeito HTML, PHP, Javascript, SQL e tomamos também a decisão do sistema ser desenvolvido de forma a ser compatível com o Internet Explorer 6.

Esta maior adequabilidade de sistemas operativos foi informada aos juízes, por contacto de e-mail, e reforçada na primeira janela da página web de avaliação áudio-percetiva (*Cf. Figura 14*). Os resultados foram obtidos entre fevereiro e maio de 2011.

²² Em colaboração com a equipa de engenharia, no âmbito do projeto ARTTS (FCT PTDC/SAU-BEB/104995/2008).

AVALIAÇÃO AUDIO-PERCEPTIVA DA VOZ
<p>No âmbito da realização da Dissertação de Doutoramento solicito a V. colaboração no preenchimento da avaliação perceptiva das 100 amostras de vozes que se seguem, com diferentes graus de alteração da qualidade vocal. Os parâmetros a analisar serão o G (Grau Geral), R(Rugosidade), B(Soprosidade), A(Astenia), e S(Tensão), cujo significado estará descrito em cada página do programa, classificado numa escala analógico-visual representada por uma régua de 100mm." Numa primeira fase é necessário que submeta o acesso no campo abaixo,introduzindo o email através do qual foi contactado. Posteriormente, é apresentado um interface com 5 régua de 100mm para cada parâmetro e um conjunto de botões que permitem conduzir a avaliação. A avaliação seguirá os seguintes passos:</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> 1- Carregar em "Ouvir" e ouvir a voz. 2- Seleccionar os valores para cada parâmetro, deslizando o cursor. 3- Carregar em "Seguinte" para submeter a avaliação e passar para a próxima voz. </div> <p style="color: blue; font-size: small;">Em cada momento da avaliação é possível voltar a avaliar as vozes anteriores, carregando em "Anterior".</p>

Email:

Algumas informações:

- 1-Usar auscultadores tipo head phones (evitar auscultadores de ouvidos)
- 2-Usar de preferência o Internet Explorer ou o Chrome (versões mais recentes).
- 3-No caso de não funcionar,instalar o Java Version 6 Update 23 a partir de
http://www.java.com/pt_BR/download/index.jsp
- Depois de instalar, reiniciar o browser.

Figura 14: Primeira janela da página Web de avaliação áudio-percetiva

4.1.4 MEDIDAS DE AVALIAÇÃO ACÚSTICA

As 90 amostras de vozes segmentadas foram avaliadas, descritas e classificadas de acordo com os seguintes parâmetros acústicos – *MeanFO* (Hz), *S.D. FO* (Hz), *Jitter(local)* (%) *Jitter(PPQ5)* (%), *Shimmer(local)* (%), *Shimmer(APQ5)* (%), *HNR* – *Harmonic-to-Noise Ratio* (dB) – por quatro aplicações de *software* de análise acústica (*opensource* (PRAAT) e comerciais (*Dr. Speech* da Tiger Electronics; MDVP – *Multi Dimensional Voice Program* da Kay Elemetrics; *VoiceStudio* da Seegnal)).

Atendendo a que a obtenção das amostras foi realizada com uma frequência de amostragem de 44100Hz, esta foi mantida para as aplicações *Dr. Speech*® da Tiger Electronics e MDVP – *Multi Dimensional Voice Program* da Kay Elemetrics; e foi subamostrada para 22050Hz com o intuito de facilitar a análise pelo *software Praat* e *VoiceStudio*, utilizando um *script* construído para o efeito em MatLab. A versão final das amostras para obtenção de medidas acústicas implicou que estas fossem pré-processadas no *software Praat*, sem recurso a qualquer suavização, com a frequência de amostragem ajustada como indicado previamente, sobre as partes mais estáveis, o que culminou em amostras de voz com uma duração média de 2 segundos (Vogel, 2011; Madill & McCabe, 2011).

4.1.5 ANÁLISE ESTATÍSTICA

A análise estatística dos dados foi realizada através do programa informático *Statistical Package for the Social Sciences – IBM® SPSS® for Windows*, versão 19.0.

A análise dos dados foi elaborada em duas fases. Numa primeira fase, em função da natureza das variáveis em estudo, calcularam-se as estatísticas descritivas de tendência central²³ (média e mediana); de dispersão²⁴ (desvio padrão); e ainda, os valores extremos (mínimo e máximo). Numa segunda fase, após avaliação da assimetria e do achatamento ou kurtose²⁵ (através dos respetivos coeficientes), e da normalidade das distribuições (por aplicação do teste não paramétrico de Kolmogorov-Smirnov-K-S (*Cf.* Anexo II), com correção de Lilliefors²⁶). Sempre que os pressupostos de normalidade eram verificados, procedeu-se à aplicação de testes paramétricos, caso contrário, recorreu-se à aplicação de testes não paramétricos (Maroco, 2010).

Relativamente à análise das medidas acústicas de acordo com o programa de *software* e para a análise das medidas acústicas de acordo com o tipo de voz (completa e segmentada), uma vez que não se verificou a normalidade das distribuições das medidas para cada aplicação de *software* nem para cada tipo de voz (*Cf.* Anexo III e Anexo IV), para a comparação dos grupos independentes, ou não relacionados (neste caso, as quatro aplicações de *software*) utilizou-se o teste não paramétrico de *Kruskal-Wallis*²⁷, uma vez que tínhamos mais do que dois grupos em comparação,

²³ Medidas de tendência central – que procuram caracterizar, de forma sumária, o valor ou característica de uma dada variável sob estudo que ocorre com mais frequência (Maroco, 2010).

²⁴ Medidas de dispersão – que refletem a dispersão das observações em torno das estatísticas de tendência central (Maroco, 2010).

²⁵ Medidas de assimetria e achatamento ou kurtose – caracterizam a forma de distribuição dos elementos da população amostrada em torno da média. São geralmente usados para comparar a forma de distribuição em estudo com uma distribuição teórica de uso ubíquo em inferência estatística (a distribuição normal). Para que uma distribuição se possa assumir como normal os valores destes coeficientes devem ser próximo de zero, i.e. dentro de um intervalo de]-0.5; 0.5[(Maroco, 2010).

²⁶ O teste de Kolmogorov-Smirnov é usado para decidir se a distribuição de uma variável sob estudo numa determinada amostra provém de uma população com uma distribuição específica. Atendendo a que, geralmente, são realizadas apenas estimativas com base em amostras, é necessário usar a aproximação analítica da estatística do teste de Lilliefors (i.e. Correção de Lilliefors), por forma a não violar o pressuposto de normalidade da distribuição em estudo (Maroco, 2010).

²⁷ O teste de Kruskal-Wallis, também chamado de ANOVA em ordens de Kruskal-Wallis, é um teste não paramétrico para testar duas ou mais amostra que provêm de uma mesma população ou de populações diferentes ou se, de igual modo, as amostras provêm de populações com a mesma distribuição (Maroco, 2010).

simultaneamente. Por sua vez, sempre que tínhamos apenas dois grupos – por exemplo, a comparação de vozes completas e segmentadas – utilizou-se o teste não paramétrico de *Mann-Whitney*²⁸. Havendo a indicação de existência de diferenças significativas, procedeu-se à sua identificação por comparação múltipla das médias das ordens recorrendo ao teste LSD (*least significant difference*) de Fisher²⁹ (Maroco, 2010). De igual modo, para grupos dependentes ou relacionados (amostras emparelhadas) repetiu-se o procedimento através da aplicação de um teste não paramétrico de *Wilcoxon*³⁰ (Cf. Anexo V). Adicionalmente, realizou-se ainda a representação gráfica, e o coeficiente de correlação não paramétrico Ró de *Spearman*³¹, para analisar a correlação entre as medidas acústicas para as diferentes aplicações de *software*. A correlação de *Spearman* baseia-se na ordenação de duas variáveis sem qualquer restrição ou pressuposto quanto à distribuição de valores das variáveis, ou seja, é mais usada para dados não paramétricos (Maroco, 2010).

Atendendo a que houve necessidade de amplificação (20dB) de um conjunto de vozes (5 amostras no *Dr. Speech*; 20 amostras no *VoiceStudio*; 58 amostras no MDVP – *Multi Dimensional Voice Program*) fez-se a análise das medidas acústicas (pares de vozes amplificados, sem *clipping*, e não amplificados) para o programa MDVP (n=13 amostras). Foram utilizados testes *t* de *Student*³² para 2 amostras independentes, uma vez que, após terem sido retirados os valores extremos severos (*outliers*) correspondentes à voz "abilird1", foram verificados os pressupostos da normalidade das distribuições das medidas acústicas para cada grupo independente à comparação (voz não amplificada e amplificada).

²⁸ O teste de Wilcoxon-Mann-Whitney, ou simplesmente teste de Mann-Withney é o teste não paramétrico adequado para comparar as funções de distribuições de uma variável pelo mesmo ordinal, medida em duas amostras independentes (Maroco, 2010).

²⁹ O teste LSD de Fisher é feito *post-hoc* para comparações múltiplas de médias, especialmente indicado para quando se compara um número reduzido de grupos ou amostras independentes (Maroco, 2010).

³⁰ O teste de Wilcoxon é o teste não paramétrico a usar quando se pretende comparar a medida de tendência central da população sob estudo com um determinado valor teórico, desde que as amostras sejam aleatórias e que a distribuição da variável sob estudo, apesar de não ser normal, seja simétrica (Maroco, 2010).

³¹ O coeficiente de correlação de Spearman (-1≤ρ≤1) é uma medida de associação não paramétrica entre duas variáveis pelo menos ordinais (Maroco, 2010).

³² O teste *t*- Student serve para testar se as médias de duas populações são ou não significativamente diferentes, requerendo que as duas amostras tenham sido obtidas aleatoriamente de duas populações e que as variáveis dependentes possuam distribuição normal e variâncias homogéneas (Maroco, 2010).

Por fim foi utilizada a Regressão Linear³³ Múltipla (*Multiple Regression Analysis*) com seleção de variáveis *stepwise*³⁴, de forma a obter um modelo parcimonioso que permitisse predizer os parâmetros da GRBAS (variáveis dependentes) em função das medidas acústicas (variáveis independentes), em cada uma das quatro aplicações de *software* em estudo (Maroco, 2010). A matriz de regressão linear múltipla obtida possibilitou a determinação de sinais de correlação (positivo e negativo) – pela análise dos resíduos³⁵ – e a qualidade da mesma (qualidade <20% péssima; entre 21 e 40% = má; entre 41 e 60% = regular; entre 61 e 80% = boa; acima de 81% = ótima). Foram consideradas e analisadas as correlações com qualidade regular a ótima, isto é, acima de 41%.

Todos os testes foram aplicados com um grau de confiança de 95%, exceto quando devidamente assinalado.

Na secção seguinte serão apresentados os resultados e, posteriormente, discutidos.

³³ Regressão Linear – conjunto vasto de técnicas estatísticas usadas para modelar relações entre variáveis e predizer o valor de uma variável dependente (ou de resposta) a partir de um conjunto de variáveis independentes (ou preditoras), através de uma reta (ou no caso específico, de uma extensão de retas para múltiplas dimensões – plano hipergeométrico – correspondentes a duas ou mais medidas acústicas estudadas) (Maroco, 2010).

³⁴ O procedimento de seleção de variáveis *stepwise* é um método híbrido que indica, com base num critério exato, quais as variáveis independentes que representam relações mais fortes com as variáveis dependentes, e por isso são melhores candidatas ao modelo definitivo (Maroco, 2010).

³⁵ A análise dos erros ou resíduos do modelo serve quer para estimar os coeficientes de regressão, quer para validar os pressupostos de aplicação do modelo de regressão linear. Quando esta análise indica relações do tipo não linear, então o modelo de regressão linear não é válido. A análise gráfica dos resíduos serve também para identificar eventuais *outliers* (Maroco, 2010).

4.2. RESULTADOS

4.2.1. DESCRIÇÃO DAS MEDIDAS ACÚSTICAS DE ACORDO COM O TIPO DE VOZ (COMPLETA OU SEGMENTADA) RELATIVAMENTE AOS TIPOS DE SOFTWARE.

Com vista a verificar se existiam diferenças estatisticamente significativas entre as vozes – sem e com pré-processamento³⁶ – comparou-se em cada software (*Dr. Speech*, *Praat* e *Voice Studio*) a duração total e a versão pré-processada com seleção da componente mais estável de cada amostra (Sousa et al., 2011), analisando-se os valores médios e respetivos desvios padrão das medidas acústicas. Do ponto de vista prático, a distinção entre as duas versões está na duração e estabilidade, ou seja, a versão pré-processada é mais curta e mais estável, respetivamente.

4.2.1.1. Dr. Speech

Para o software *Dr. Speech*, apenas a medida acústica *SDF0* apresentou valores significativamente diferentes consoante o tipo de voz ($p = 0,008$), sendo que as vozes “completas” (média = 2,37) apresentaram valores significativamente superiores às vozes “segmentadas” (média = 1,79) (Cf. Quadro 3).

Quadro 3: Comparação da média (desvio padrão) das medidas acústicas de acordo com o tipo de voz (*Dr. Speech*)

Medidas acústicas	Software - Dr. Speech						
	Voz Completa			Voz Segmentada			valor-p*
	n	Média	Desvio padrão	n	Média	Desvio padrão	
MeanF0(Hz)	85	199.96	53.16	86	197.50	55.00	0.774
SDF0(Hz)	85	2.37	1.71	86	1.79	1.08	0.008
jitter(local)%	85	0.33	0.33	86	0.30	0.24	0.700
Shimmer(local)%	85	2.86	2.03	86	2.75	1.94	0.727
HNR(dB)	85	22.61	5.24	86	22.85	5.31	0.671

* Resultados de acordo com o teste não paramétrico de *Mann-Whitney*, a 95 % de confiança.

³⁶ O pré-processamento aqui considerado consiste na pré-selecção, com base na análise visual do espectrograma, da componente mais estável de cada uma das amostras (Buder, 2000; Sousa et al., 2011).

4.2.1.2. Praat

De igual forma, para o *Praat*, apenas a medida acústica *SDF0* apresentou valores significativamente diferentes consoante o tipo de voz ($p = 0,012$), sendo que as vozes “completas” (média = 7,97) apresentaram valores significativamente superiores às vozes “segmentadas” (média = 6,45) (Cf. Quadro 4).

Quadro 4: Comparação da média (desvio padrão) das medidas acústicas de acordo com o tipo de voz (*Praat*)

Medidas acústicas	Software - Praat						
	Voz Completa			Voz Segmentada			valor-p*
	n	Média	Desvio padrão	n	Média	Desvio padrão	
MeanF0(Hz)	90	196.59	56.51	90	196.59	56.03	0.961
SDF0(Hz)	90	7.97	18.50	90	6.45	16.76	0.012
Jitter(PPQ5)%	90	0.31	0.32	90	0.31	0.32	0.874
Jitter(local)%	90	0.53	0.59	90	0.52	0.60	0.689
Shimmer(APQ5)%	90	2.52	1.95	90	2.57	2.07	0.908
Shimmer(local)%	90	4.18	3.33	90	4.16	3.39	0.918
HNR(dB)	90	19.38	5.75	90	19.48	5.84	0.882

* Resultados de acordo com o teste não paramétrico de *Mann-Whitney*, a 95 % de confiança.

4.2.1.3. Voice Studio

Por fim, para o *Voice Studio*, verifica-se também que apenas a medida acústica *SDF0* apresentou valores significativamente diferentes consoante o tipo de voz ($p = 0,041$), sendo que as vozes “completas” (média = 22,37) apresentaram valores significativamente superiores às vozes “segmentadas” (média = 13,30) (Cf. Quadro 5).

Em suma, conclui-se que independentemente do programa de software, a medida acústica *SDF0* apresentou valores significativamente mais elevados nas vozes não processadas, designadas como “completas”. Para as restantes medidas, a análise dos dados não revelou diferenças estatisticamente significativas (valores-p > 0,05).

Quadro 5: Comparação da média (desvio padrão) das medidas acústicas de acordo com o tipo de voz (*Voice Studio*)

Medidas acústicas	Software - Voice Studio						
	Voz Completa			Voz Segmentada			valor-p*
	n	Média	Desvio padrão	n	Média	Desvio padrão	
MeanF0(Hz)	90	197.82	56.91	90	196.51	56.05	0.899
SDF0(Hz)	90	22.37	49.47	90	13.30	26.31	0.041
Jitter(PPQ5)%	90	2.18	2.36	90	1.99	2.44	0.367
Jitter(local)%	90	3.51	3.85	90	3.17	3.68	0.416
Shimmer(APQ5)%	90	3.84	3.18	90	3.68	3.18	0.592
Shimmer(local)%	90	6.05	4.97	90	5.80	4.94	0.593
HNR(dB)	90	13.80	8.33	90	15.82	7.02	0.169

* Resultados de acordo com o teste não paramétrico de *Mann-Whitney*, a 95 % de confiança.

4.2.2. DESCRIÇÃO DAS MEDIDAS ACÚSTICAS (PARES DE VOZES - MDVP) DE ACORDO COM A AMPLIFICAÇÃO DAS AMOSTRAS DE VOZ.

Atendendo à elevada quantidade de vozes que não possuía amplitude suficiente para análise pelas diferentes aplicações de *software*, recorreu-se à sua amplificação (20dB) até atingir entre 50-75% de magnitude espectral, sem incorrer em fenómenos de *clipping*. Assim, foram amplificadas 5 amostras no *Dr. Speech*; 20 amostras no *VoiceStudio* e 58 amostras no MDVP – *Multi Dimensional Voice Program*. Realizou-se a análise das medidas acústicas (pares de vozes amplificadas – sem *clipping* – e não amplificadas) para o programa MDVP (n=13 amostras).

No quadro 6 apresentam-se os valores da média e respetivos desvios padrão das medidas acústicas, segundo a amplificação da voz. Os resultados evidenciam que não foram encontradas diferenças estatisticamente significativas entre os valores das medidas acústicas para as vozes não amplificadas e amplificadas (valores-p > 0,05). Analisando as diferenças das médias (praticamente nulas) e a magnitude dos níveis de significância dos testes realizados (valores-p ≈ 0,000), conclui-se que, de uma forma

geral, a amplificação da voz não provoca alterações significativas nos valores das medidas acústicas estudadas.

Quadro 6: Comparação da média (desvio padrão) das medidas acústicas de acordo com a amplificação da voz

Medidas acústicas	Amplificação da voz						Diferença de Médias	valor-p		
	Voz não amplificada			Voz amplificada (20db)						
	n *	Média	Desvio padrão	n *	Média	Desvio padrão				
Mean F0 (Hz)	12	219.1	36.8	12	219.1	36.8	0.0	1.000		
SDF0 (Hz)	12	2.8	1.5	12	2.8	1.5	0.0	0.994		
Jitter (PPQ5) %	12	0.8	0.5	12	0.8	0.5	0.0	0.999		
Jitter (Jitt) %	12	1.3	0.9	12	1.4	0.8	-0.1	0.703		
Shimmer (APQ5) %	12	2.4	1.0	12	2.4	1.0	0.0	1.000		
Shimmer (Shim) %	12	3.4	1.5	12	3.4	1.5	0.0	1.000		
HNR (dB)	12	0.1	0.0	12	0.1	0.0	0.0	1.000		

Resultados de acordo com o teste *t* de *Student* para duas amostras independentes, a 95 % de confiança.

* Foi retirada da análise a voz "abilird1" por se ter verificado que os seus valores eram *outliers* severos em todas as medidas acústicas.

4.2.3. DESCRIÇÃO DAS MEDIDAS ACÚSTICAS DE ACORDO COM O PROGRAMA DE SOFTWARE.

No sentido de averiguar possíveis relações entre as medidas acústicas e o programa de software analisaram-se os valores médios e respetivos desvios padrão das medidas acústicas de acordo com os tipos de software em estudo (*Cf. Quadro 7*).

Quadro 7: Comparação da média (desvio padrão) das medidas acústicas de acordo com o programa de *software*

Medidas acústicas / Software	Estatísticas descritivas						valor-p
	n*	Média	Mediana	Desvio padrão	Mínimo	Máximo	
MeanF0(Hz)							ns
Dr. Speech	86	197.50	205.90	55.00	82.52	316.56	
MDVP	90	199.15	205.88	56.88	82.52	368.91	
Praat	90	196.59	205.18	56.03	82.52	309.01	
Voice Studio	90	196.51	205.91	56.05	82.52	308.64	
SDF0(Hz)							p<0.05
Dr. Speech	86	1.79 ^c	1.54	1.08	0.48	7.77	
MDVP	90	5.80 ^b	2.70	14.31	0.52	112.47	
Praat	90	6.45 ^c	1.51	16.76	0.45	89.07	
Voice Studio	90	13.30 ^a	4.40	26.31	0.61	201.04	
Jitter(local)%							p<0.05
Dr. Speech	86	0.30 ^d	0.22	0.24	0.10	1.39	
MDVP	90	1.77 ^b	1.15	1.76	0.23	11.29	
Praat	90	0.52 ^c	0.35	0.60	0.15	4.28	
Voice Studio	90	3.17 ^a	2.03	3.68	0.38	26.26	
Jitter(PPQ5)%							p<0.05
Dr. Speech	0	
MDVP	90	1.04 ^b	0.67	1.10	0.14	7.67	
Praat	90	0.31 ^c	0.21	0.32	0.09	2.20	
Voice Studio	90	1.99 ^a	1.27	2.44	0.26	16.82	
Shimmer(local)%							p<0.05
Dr. Speech	86	2.75 ^c	2.26	1.94	0.73	13.84	
MDVP	90	5.27 ^a	4.11	4.27	1.35	24.99	
Praat	90	4.16 ^b	3.17	3.39	1.07	18.31	
Voice Studio	90	5.80 ^a	4.61	4.94	1.47	32.36	
Shimmer(APQ5)%							p<0.05
Dr. Speech	0	
MDVP	90	3.94 ^a	3.00	3.51	1.08	19.06	
Praat	90	2.57 ^b	2.05	2.07	0.64	12.24	
Voice Studio	90	3.68 ^a	2.80	3.18	0.92	19.67	
HNR(dB)							p<0.05
Dr. Speech	86	22.85 ^a	23.21	5.31	9.76	30.71	
MDVP	90	17.07 ^c	17.62	3.06	-0.64	23.74	
Praat	90	19.48 ^b	19.56	5.84	3.17	28.52	
Voice Studio	90	15.82 ^c	17.37	7.02	-4.98	25.60	

* Número de casos válidos (sem *missing values*).

Resultados de acordo com o teste não paramétrico de *Kruskal-Wallis*, a 95 % de confiança.

a, b, c, d - grupos homogêneos de acordo com a comparação múltipla de médias das ordens (LSD de *Fisher*), a 95% de confiança.

Os resultados evidenciam que, relativamente à medida acústica “SDF0(Hz)”, verifica-se que o *software Voice Studio* apresenta valores significativamente superiores aos

restantes, tendo as aplicações de *software Dr. Speech* e *Praat* apresentado os valores mais baixos. Para “HNR(dB)”, constata-se que o *Dr. Speech* tem valores significativamente superiores às restantes aplicações, sendo que as aplicações de *software MDVP* e *Voice Studio* apresentam os valores mais baixos. Em relação a “Jitter(local)%”, apura-se que o *Voice Studio* apresenta valores significativamente superiores, tendo o *Dr. Speech* apresentado os valores mais baixos. Quanto a “Jitter (PPQ5)%”, apura-se que o *Voice Studio* apresenta valores significativamente superiores aos restantes, tendo o *Praat* apresentado os valores mais baixos. Para “Shimmer(local)%”, as aplicações *MDVP* e *Voice Studio* apresentam valores significativamente superiores aos restantes, tendo o *Dr. Speech* os mais baixos. Por fim, para “Shimmer(APQ5)%”, *MDVP* e *Voice Studio* apresentam valores significativamente superiores aos do *Praat*.

Em suma, de uma forma geral, conclui-se que o *software Voice Studio* apresentou valores significativamente superiores aos restantes para todas as medidas acústicas, exceto para *HNR*. Esta foi alta para o *software Dr. Speech*. Em relação a *MeanF0(Hz)* não foram encontradas diferenças significativas segundo o programa de *software* ($p > 0,05$).

4.2.4. DESCRIÇÃO DAS MEDIDAS ACÚSTICAS DE ACORDO COM O PROGRAMA DE SOFTWARE – REPRESENTAÇÃO GRÁFICA E COEFICIENTE DE CORRELAÇÃO.

A análise da relação entre as medidas acústicas e os quatro tipos de *software* recorreu a representações gráficas (gráficos de linhas para as 90 vozes em estudo) e à análise do grau de associação entre os valores das medidas acústicas em cada programa de *software*, através do coeficiente de correlação não paramétrico Ró de Spearman.

4.2.4.1. Mean F0Hz

A análise da figura seguinte (*Cf. Gráfico 1*) revela que há uma sobreposição quase total dos valores obtidos para a medida *Mean F0*, para as quatro aplicações de *software* em estudo. Saliente-se que o *Dr. Speech* não forneceu valores em 4 amostras de vozes. Foi tentada a sua amplificação, a qual resultou em *clipping*.

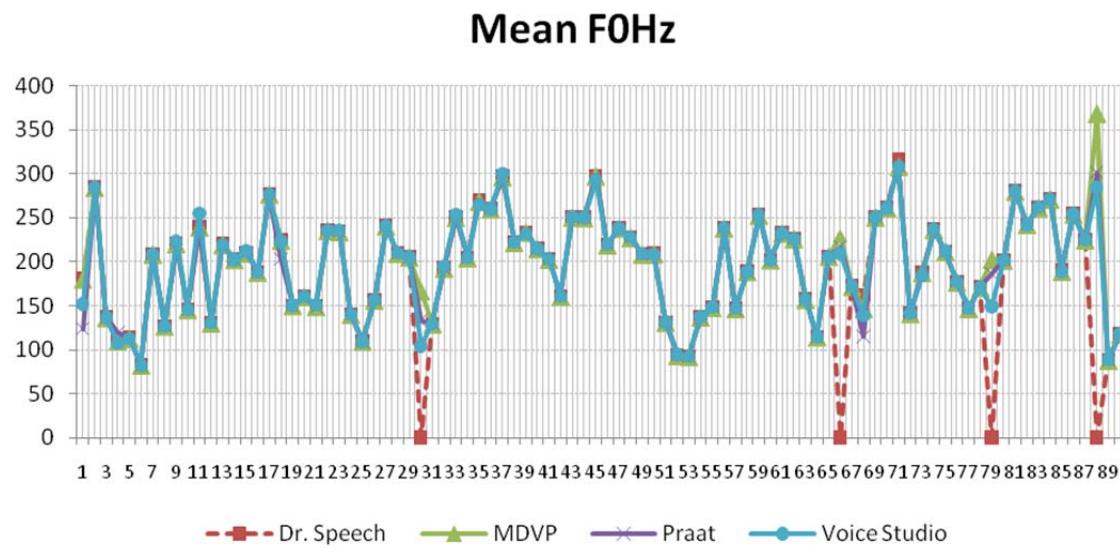


Gráfico 1: Representação gráfica do comportamento da medida “Mean F0” para os quatro tipos de *software* em estudo

O quadro 8 mostra a existência de uma associação linear muito elevada, positiva e estatisticamente significativa ($R_\text{o} > 0.99$; $p < 0.01$), entre todos os tipos de *software*, para a medida *Mean F0*.

Quadro 8: Coeficientes de correlação de *Spearman* entre os valores da medida acústica “Mean F0Hz” para os tipos de *software*

Software	Mean F0Hz			
	Dr. Speech	MDVP	Praat	Voice Studio
Dr. Speech	1	.999**	.989**	.997**
MDVP		1	.989**	.989**
Praat			1	.992**
Voice Studio				1

Resultados de acordo com o coef. de correlação não paramétrico R_o de Spearman
(* $p < 0.05$; ** $p < 0.01$).

4.2.4.2. SDFOHz

Da análise da figura seguinte (*Cf. Gráfico 2*) destaca-se a discrepância entre os valores da medida “SDFO” entre os vários tipos de *software*, facto este mais notado no *Voice Studio*.

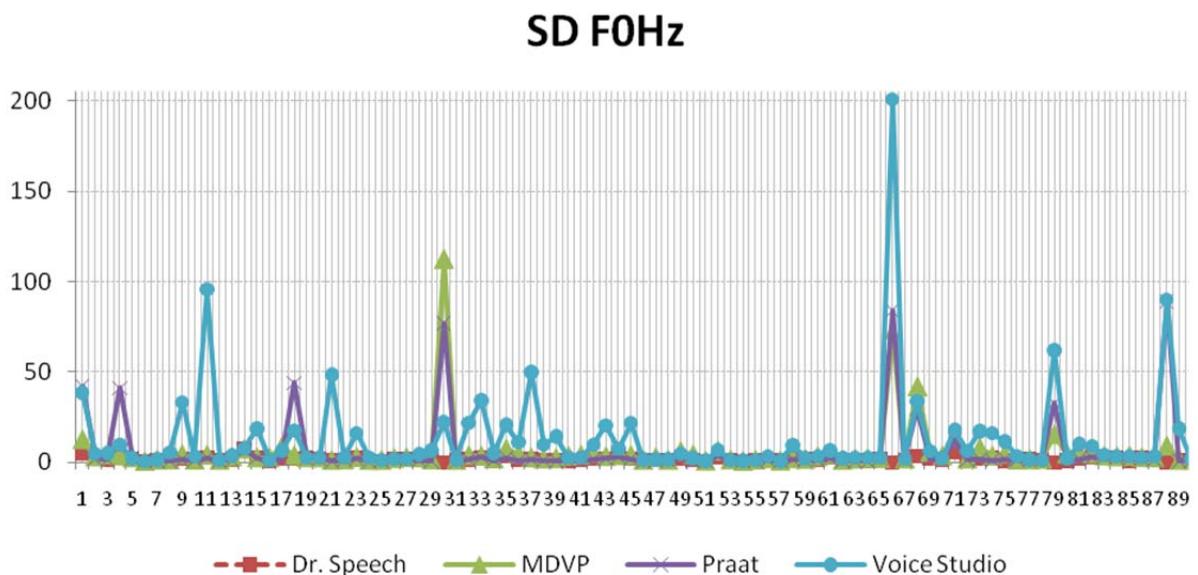


Gráfico 2: Representação gráfica do comportamento da medida “SD F0” para os quatro tipos de *software* em estudo

Este dado fica a dever-se às diferentes regras de suavização usadas pelo algoritmo de estimativa de F0, específico para cada uma das aplicações de *software*. Posto isso, é de esperar que a elevada variação entre valores instantâneos, já descrita anteriormente (*Cf. Capítulo 4, secção 2.3.*) condicione a variância encontrada, inerente à regra de suavização usada.

Quadro 9: Coeficientes de correlação de Spearman entre os valores da medida acústica “SD F0Hz” e os tipos de *software*

Software	SD F0Hz			
	Dr. Speech	MDVP	Praat	Voice Studio
Dr. Speech	1	.690**	.907**	.573**
MDVP		1	.726**	.647**
Praat			1	.580**
Voice Studio				1

Resultados de acordo com o coef. de correlação não paramétrico Ró de Spearman
(*p<0.05; **p<0.01).

O quadro 9 mostra a existência de uma associação linear muito elevada entre as aplicações de software *Dr. Speech* e *Praat* ($R_o = 0.91$) e moderada com os restantes tipos ($0.57 < R_o < 0.73$). Todas as correlações são positivas e altamente significativas (valores de $p < 0.01$).

4.2.4.3. HNRdB

Da análise da figura seguinte (Cf. Gráfico 3) destaca-se a discrepancia entre os valores da medida “HNRdB” entre os vários tipos de software, facto este mais notado no *Voice Studio*. Saliente-se que o *Dr. Speech* não forneceu valores em 4 amostras de vozes. Foi tentada a sua amplificação, a qual resultou em *clipping*. Neste software também se observam os valores mais elevados da medida “HNRdB”.

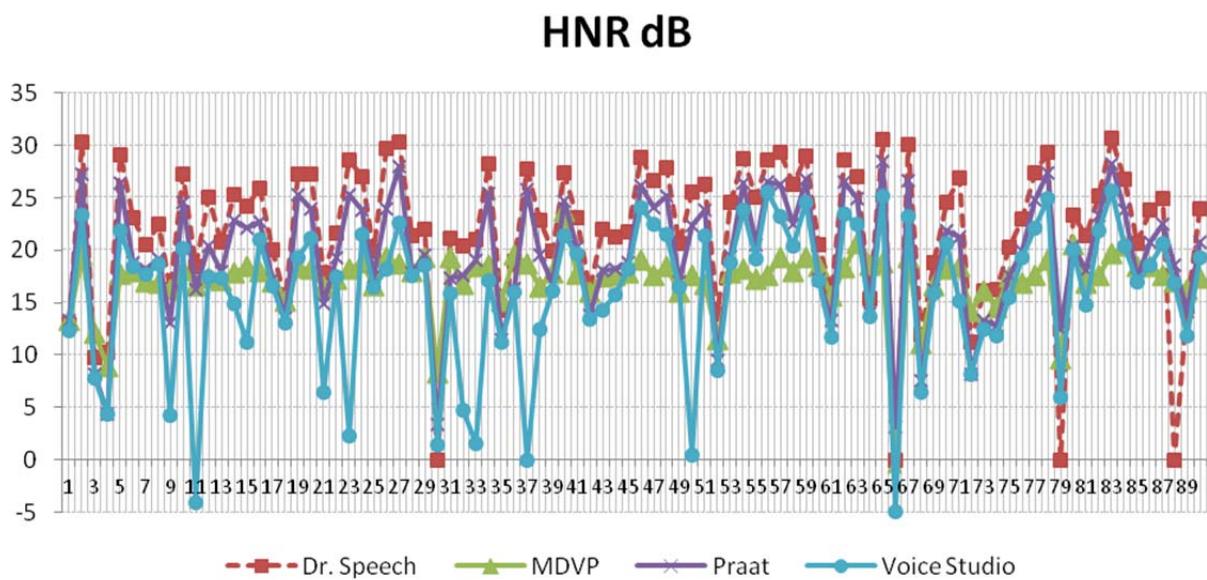


Gráfico 3: Representação gráfica do comportamento da medida “HNRdB” para os quatro tipos de software em estudo

O quadro 10 mostra a existência de uma associação linear muito elevada entre o software *Dr. Speech* e *Praat* ($R_o = 0.99$) e moderada/elevada com os restantes tipos ($0.61 < R_o < 0.79$). Todas as correlações são positivas e altamente significativas (valores de $p < 0.01$).

Quadro 10: Coeficientes de correlação de *Spearman* entre os valores da medida acústica “*HNRdB*” e os tipos de *software*

Software	<i>HNRdB</i>			
	Dr. Speech	MDVP	Praat	Voice Studio
Dr. Speech	1	.718**	.985**	.765**
MDVP		1	.738**	.614**
Praat			1	.789**
Voice Studio				1

Resultados de acordo com o coef. de correlação não paramétrico Ró de Spearman

(*p<0.05; **p<0.01).

4.2.4.4. *Jitterlocal*

A análise da figura seguinte (*Cf.* Gráfico 4) evidencia a discrepância entre os valores da medida “*Jitter local*” entre os vários tipos de *software*, facto este mais notado no *Voice Studio*.

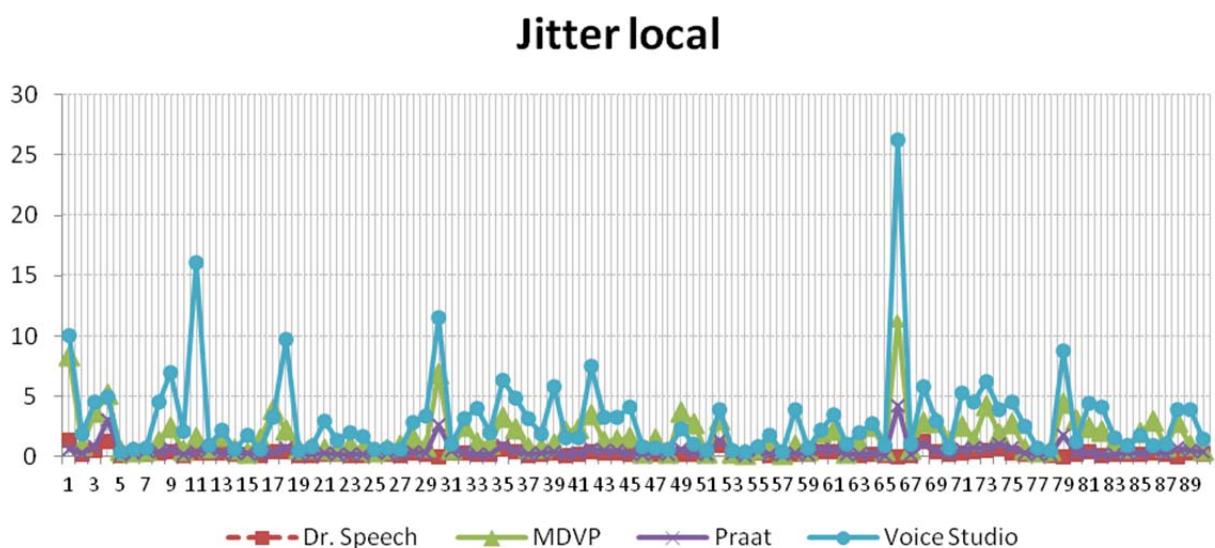


Gráfico 4: Representação gráfica do comportamento da medida “*Jitter local*” para os quatro tipos de *software* em estudo

O quadro 11 mostra a existência de uma associação linear muito elevada entre o *software Dr. Speech* e *Praat* ($R_\text{o} = 0.95$) e moderada com os restantes tipos ($0.67 < R_\text{o} < 0.73$). Todas as correlações são positivas e altamente significativas (valores de $p < 0.01$).

Quadro 11: Coeficientes de correlação de *Spearman* entre os valores da medida acústica “*Jitterlocal*” e os tipos de *software*

Software	<i>Jitterlocal</i>			
	Dr. Speech	MDVP	Praat	Voice Studio
Dr. Speech	1	.674**	.947**	.697**
MDVP		1	.697**	.726**
Praat			1	.676**
Voice Studio				1

Resultados de acordo com o coef. de correlação não paramétrico Ró de Spearman
(*p<0.05; **p<0.01).

4.2.4.5. *JitterPPQ5*

Da observação da figura seguinte (*Cf.* Gráfico 5) destaca-se a discrepância entre os valores da medida “*Jitter PPQ5*” entre os vários tipos de *software*, facto este mais notado no *Voice Studio*.

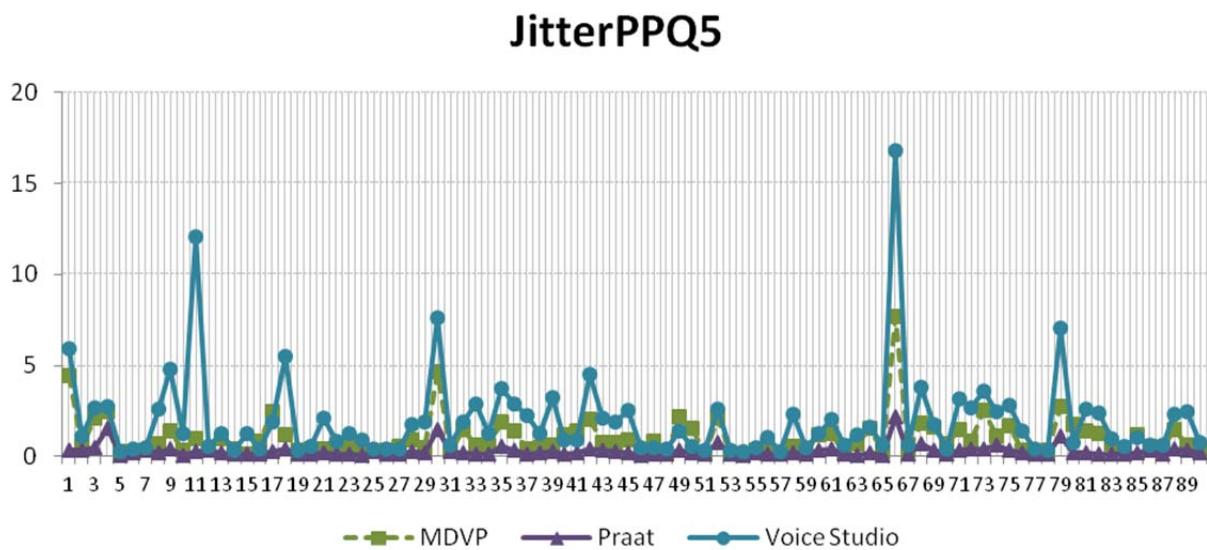


Gráfico 5: Representação gráfica do comportamento da medida “*Jitter PPQ5*” para os quatro tipos de *software* em estudo

O quadro 12 mostra a existência de uma associação linear moderada entre os três tipos de *software* ($0.68 < \text{Ró} < 0.74$). Todas as correlações são positivas e altamente significativas (valores de $p < 0.01$). O *software Dr. Speech* não possui esta medida, pelo que não foi contemplado.

Quadro 12: Coeficientes de correlação de *Spearman* entre os valores da medida acústica “*JitterPPQ5*” e os tipos de software

Software	<i>JitterPPQ5</i>		
	MDVP	Praat	Voice Studio
MDVP	1	.691**	.738**
Praat		1	.680**
Voice Studio			1

Resultados de acordo com o coef. de correlação não paramétrico Ró de Spearman (*p<0.05; **p<0.01).

4.2.4.6. *Shimmerlocal*

Na figura seguinte (*Cf.* Gráfico 6) evidencia-se a discrepância reduzida entre os valores da medida “*Shimmer local*” entre as várias aplicações. Para muitas amostras o comportamento dos diferentes tipos de *software* é quase coincidente.

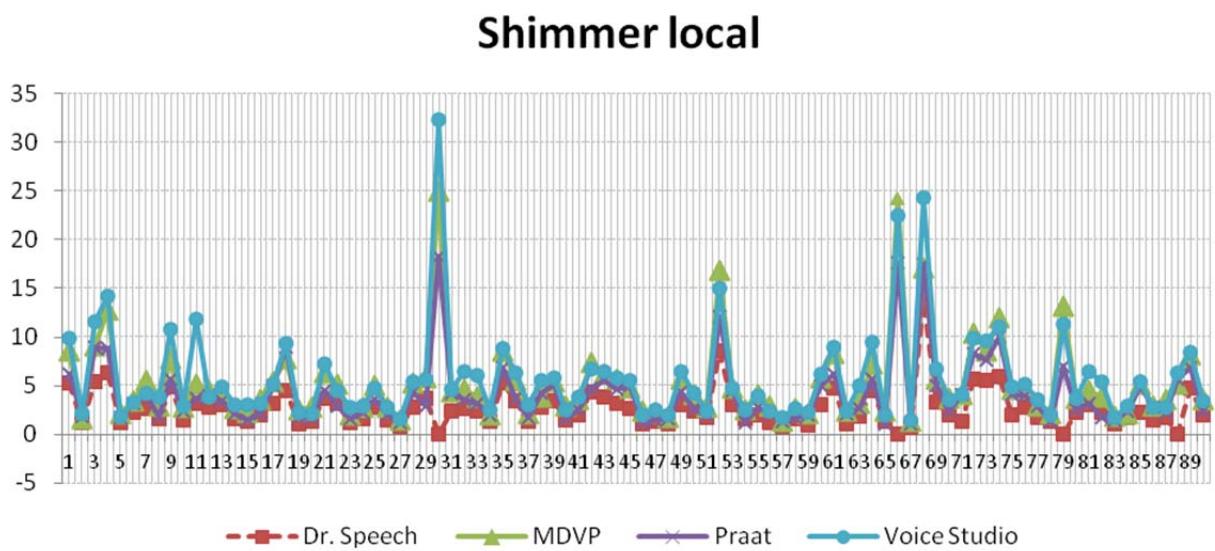


Gráfico 6: Representação gráfica do comportamento da medida “*Shimmer local*” para os quatro tipos de *software* em estudo

O quadro 13 mostra a existência de uma associação linear muito elevada, positiva e estatisticamente significativa ($R\hat{o} > 0.91$; $p < 0.01$), entre todos os tipos de software.

Quadro 13: Coeficientes de correlação de *Spearman* entre os valores da medida acústica “*Shimmerlocal*” e os tipos de software

Software	<i>Shimmerlocal</i>			
	Dr. Speech	MDVP	Praat	Voice Studio
Dr. Speech	1	.959**	.908**	.947**
MDVP		1	.942**	.944**
Praat			1	.922**
Voice Studio				1

Resultados de acordo com o coef. de correlação não paramétrico Ró de Spearman
(*p<0.05; **p<0.01).

4.2.4.7. *ShimmerAPQ5*

A análise da figura seguinte (*Cf.* Gráfico 7) revela a discrepância reduzida entre os valores da medida “*Shimmer APQ5*” entre os vários tipos de *software*.

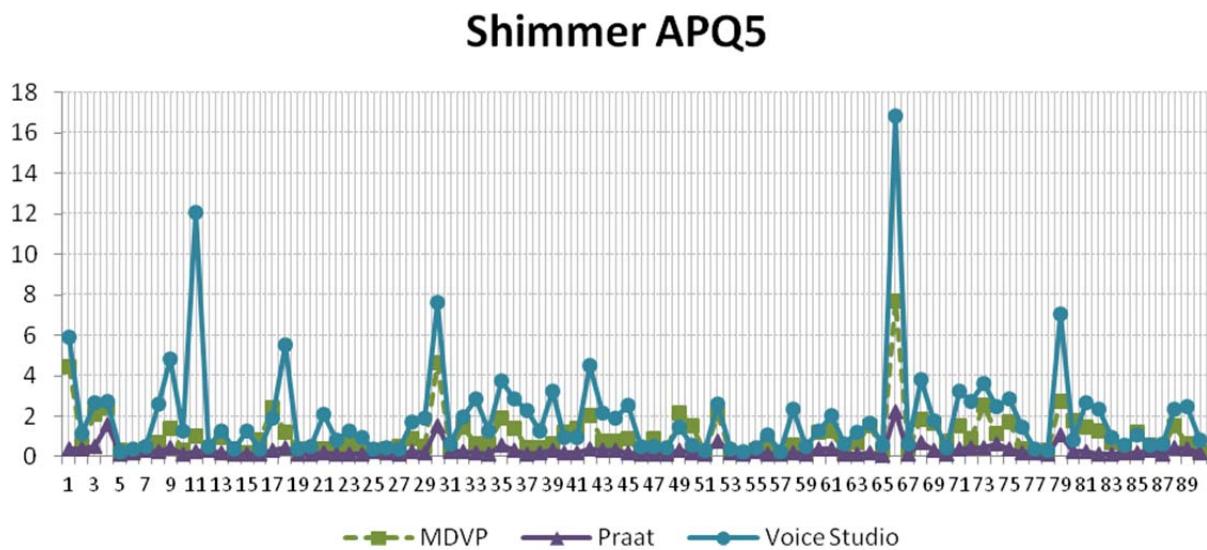


Gráfico 7: Representação gráfica do comportamento da medida “*ShimmerAPQ5*” para os quatro tipos de *software* em estudo

O quadro 14 mostra a existência de uma associação linear elevada/muito elevada entre os três tipos de *software* ($0.88 < R\bar{o} < 0.91$), sendo a mais elevada entre as aplicações *MDVP* e *Voice Studio* ($R\bar{o} = 0.91$). Todas as correlações são positivas e altamente significativas (valores de $p < 0.01$). O *software Dr. Speech* não possui esta medida, pelo que não foi contemplado.

Quadro 14: Coeficientes de correlação de *Spearman* entre os valores da medida acústica “ShimmerAPQ5” e os tipos de *software*

Software	ShimmerAPQ5		
	MDVP	Praat	Voice Studio
MDVP	1	.883**	.908**
Praat		1	.884**
Voice Studio			1

Resultados de acordo com o coef. de correlação não paramétrico Ró de Spearman (*p<0.05; **p<0.01).

A análise do grau de associação entre os valores das medidas acústicas em cada programa de *software* permite concluir da existência de uma associação moderada a forte, positiva e estatisticamente significativa (valores de p < 0.01) entre os quatro tipos de *software*. Verificam-se situações, pontuais, de discrepância.

4.2.5. DESCRIÇÃO DAS CLASSIFICAÇÕES DOS PARÂMETROS ÁUDIO-PERCETIVOS DE ACORDO COM OS PERITOS.

4.2.5.1. Descrição da avaliação áudio-percetiva (GRBAS)

No quadro 15 estão representadas as principais estatísticas descritivas da GRBAS, de acordo com os dez peritos estudados.

Quadro 15: Estatísticas descritivas para a GRBAS de acordo com os peritos

Perito	Escala									
	G(Grau Geral)		R(Rugosidade)		B(Soprosidade)		A(Astenia)		S(Tensão)	
	Média	Dp	Média	Dp	Média	Dp	Média	Dp	Média	Dp
PERITO1	48.19	25.00	33.63	29.00	36.32	25.00	9.98	18.00	45.51	27.00
PERITO2	28.40	30.00	13.61	26.00	5.61	15.00	13.56	24.00	13.23	23.00
PERITO3	29.46	35.00	14.82	28.00	17.76	25.00	9.60	18.00	12.03	26.00
PERITO4	36.59	26.00	34.99	26.00	51.94	27.00	21.34	26.00	51.11	27.00
PERITO5	56.27	18.00	37.51	20.00	37.12	27.00	30.00	28.00	29.26	30.00
PERITO6	39.63	31.00	16.08	24.00	20.13	26.00	17.13	25.00	17.57	22.00
PERITO7	47.52	26.00	42.74	25.00	43.66	27.00	28.00	31.00	16.43	25.00
PERITO8	51.73	25.00	26.92	29.00	40.74	29.00	13.63	21.00	14.73	26.00
PERITO9	53.21	23.00	22.01	26.00	40.26	30.00	9.17	18.00	10.71	22.00
PERITO10	50.48	17.00	32.87	24.00	32.00	22.00	4.96	12.00	23.08	20.00

Dp = Desvio Padrão.

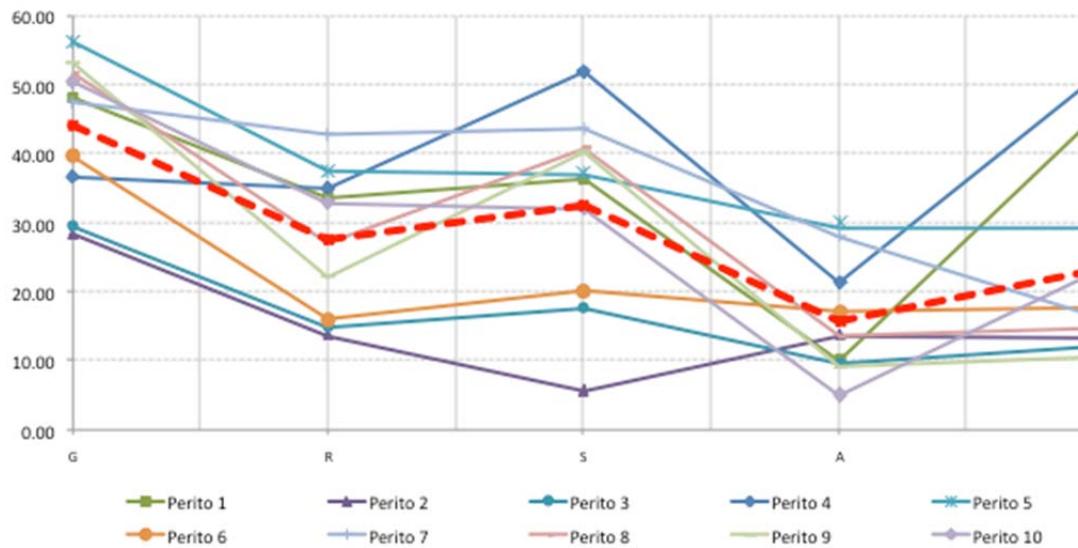
Em média, os valores mais elevados para cada um dos parâmetros da escala verificaram-se em diferentes peritos, nomeadamente:

- G (PERITO 5; média = 56.3);
- R (PERITO 7; média = 42.7);
- B (PERITO 4; média = 51.9);
- A (PERITO 5; média = 30.0);
- S (PERITO 4; média = 51.1).

Note-se que o PERITO 4 cotou como mais severos, em média, os parâmetros B(Soprosidade) e S(Tensão), e o PERITO 5 o G(Grau Geral) e A(Astenia).

É de realçar a elevada dispersão observada para todos os parâmetros em estudo (Desvio Padrão entre 12.00 e 35.00), compatível com a variabilidade de graus de severidade das amostras vocais usadas. Observa-se que o parâmetro com classificações mais altas foi o G(Grau Geral), e aquele em que a pontuação obtida tendeu a ser menor foi o A(Astenia).

Gráfico 8: Valores médios da GRBAS para cada perito (valor total da média assinalado a tracejado)



A análise do gráfico 8 mostra-nos que, de uma forma geral, o PERITO 6, PERITO 3 e PERITO 2, apresentam valores em média inferiores ao valor total médio para a GRBAS. No sentido oposto encontramos os avaliadores PERITO 4, PERITO 5 e PERITO 7. A análise e discussão destes dados está no Capítulo 4, Secção 3.4..

4.2.5.2. Descrição da consistência Inter-Avaliador (CCIC)

Uma vez que foram obtidas as avaliações das componentes da GRBAS por dez peritos, para as 90 vozes da amostra, utilizou-se o Coeficiente de Correlação Intra-Classe (CCIC) de forma a avaliar o seu grau de concordância, sendo assim possível apurar se todos interpretaram de forma idêntica os parâmetros percetivos de classificação da escala GRBAS, garantindo assim a sua consistência (*Cf.* Quadro 16).

Quadro 16: Estatísticas descritivas e Coeficiente de Correlação Intra-Classe para as componentes da GRBAS, de acordo com os peritos

Escala	Nº Itens (peritos)	Média ($\pm D_p$)	Minimo	Máximo	CCIC*	IC 95%	p ^a
G	10	44.1 (± 10.0)	28.4	56.3	0.952	[0.936; 0.966]	p<0.05
R	10	27.5 (± 10.4)	13.6	42.7	0.938	[0.916; 0.955]	p<0.05
B	10	32.6 (± 14.0)	5.6	51.9	0.938	[0.924; 0.959]	p<0.05
A	10	15.7 (± 8.3)	5.0	30.0	0.887	[0.849; 0.919]	p<0.05
S	10	23.4 (± 14.3)	10.7	51.1	0.879	[0.838; 0.913]	p<0.05

* CCIC = Coeficiente de Correlação Intra-Classe.

^a Resultados de acordo com o teste não paramétrico de Kruskal-Wallis, a 95% de confiança.

Concluiu-se que a classificação em tendência central dada pelos 10 peritos a cada componente da GRBAS difere de modo estatisticamente significativo (valores de p < 0.05). Os valores dos intervalos de confiança a 95% das correlações entre as classificações médias dadas pelos peritos para todas as componentes da GRBAS situa-se, de forma geral, entre 0.838 e 0.966 – valores superiores a 0.8 – pelo que se conclui existir uma boa consistência dos parâmetros da escala usada. Note-se ainda que os três primeiros parâmetros da escala (G, R e B) têm uma consistência superior aos restantes (A e S). Assim, os dez peritos manifestam ter critérios similares de avaliação para os cinco parâmetros da GRBAS em análise, porém não fazem a sua quantificação de forma estatisticamente semelhante.

4.2.5.3. Descrição da diferença entre as vozes “originais” e respetivas

“repetições”, de acordo com os peritos (Consistência Intra-Avaliador)

Atendendo aos resultados do CCIC supradescrito procedeu-se a uma análise pormenorizada do grau de facilidade de avaliação e a vocação das vozes para conduzirem a resultados mais consensuais entre os dez peritos e, assim, selecionar

aqueles cujas classificações áudio-percetivas serão usadas para estabelecer correspondências entre as medidas acústicas e os parâmetros percetivos.

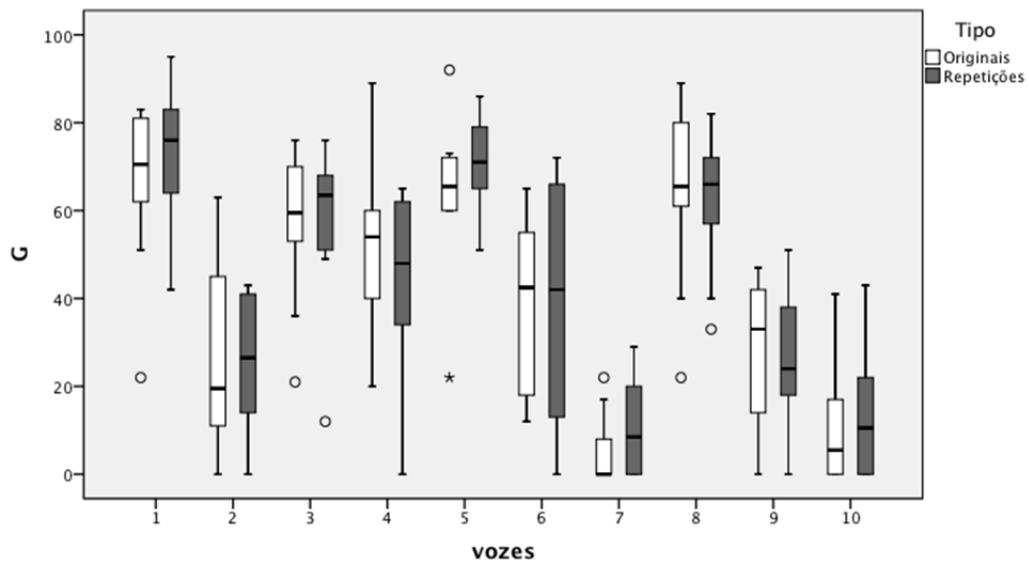
Empreenderam-se três estudos parcelares, com base nas avaliações das dez vozes “originais” e respetivas “repetições” do banco de 100 amostras vocais usadas na dissertação. O primeiro teve como objetivo caracterizar o poder dos parâmetros áudio-percetivos selecionados para gerarem avaliações consensuais. De seguida procedeu-se à seleção – de entre o conjunto das “repetições” – das vozes com maior vocação para o consenso inter-avaliadores. Por fim, e de acordo com os resultados prévios, encetou-se um terceiro estudo parcelar no qual se selecionaram os peritos com mais tendência para classificações áudio-percetivas concordantes.

4.2.5.3.1. Estudo Parcelar 1 – Análise e caracterização do consenso na avaliação áudio-percetiva em cada um dos parâmetros classificados da GRBAS

Primariamente foi feita a representação e análise gráfica da classificação áudio-percetiva pelos 10 peritos em formato de *box-plot*³⁷ ou diagrama de extremos e quartis (*Cf.* Gráfico 9 a 13), para cada um dos parâmetros da GRBAS para as vozes “originais” e “repetições” de 10% (equivalente a 10 vozes) do total da amostra. A sua análise evidencia as discrepâncias e uniformidades de classificação dos peritos.

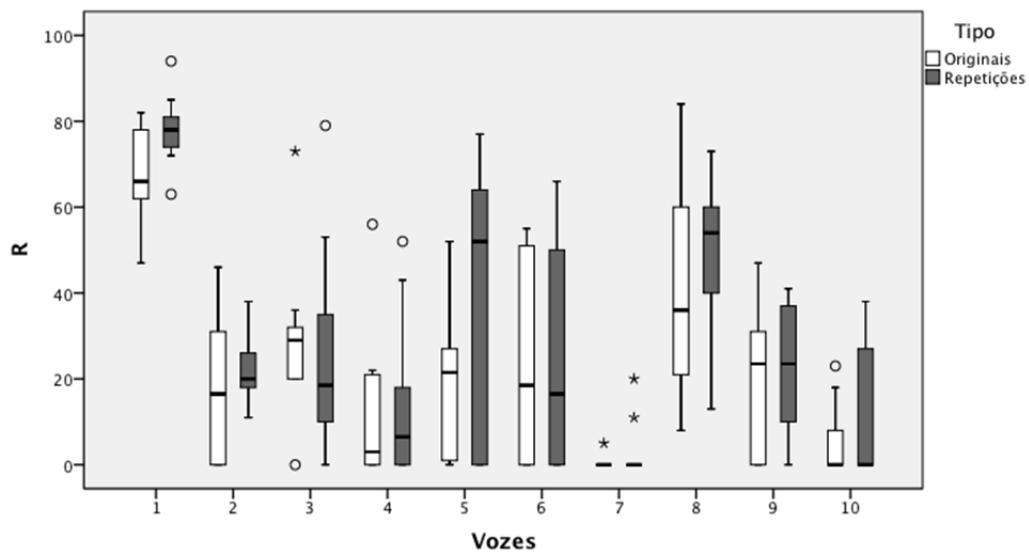
³⁷ O *Box-Plot*, também designado por *box-and-whisker plot*, corresponde a uma opção de representação gráfica que tem como objetivo principal a visualização de características da variável em estudo, os seja, das características amostrais. Cada coluna do gráfico é composta por uma caixa delimitada inferiormente pelo 1º quartil (percentil 25) e superiormente pelo 3º quartil (percentil 75). A linha média corresponde ao 2º quartil ou a mediana. Dessa caixa deriva uma linha que culmina, acima, no maior valor que não é um *outlier*, e o mesmo acontece para o extremo inferior da linha (ou bigode). A existência de círculos fora do diagrama representa os *outliers* (i.e. valores 1.5 vezes superiores ao 3ºQuartil+Amplitude Inter-Quartílica ou 1.5 vezes inferiores ao 1ºQuartil-Amplitude Inter-Quartílica). Quando surge o símbolo estrela, tal corresponde a um *extremo* da amostra, ou seja, valores três vezes superiores ao 3ºQuartil+Amplitude Inter-Quartílica ou três vezes superior ao 1ºquartil-Amplitude Inter-Quartílica (Maroco, 2010).

Gráfico 9: Box-Plot das 10 vozes “originais” e respetivas “repetições” do parâmetro G-Grau da GRBAS, para os 10 peritos



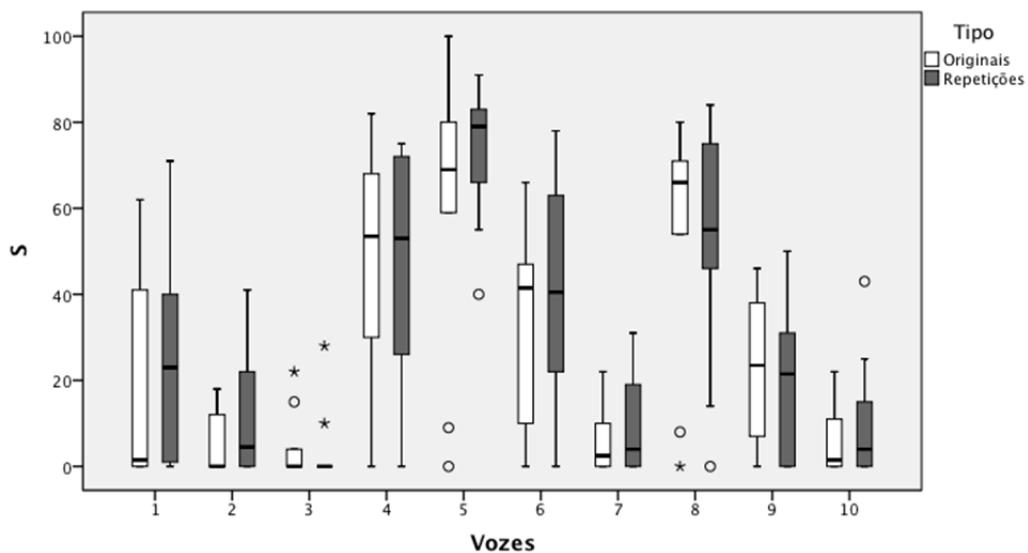
Para o parâmetro G(Grau Geral) (Cf. Gráfico 9), observa-se uma variabilidade de classificações em torno da média maior para as vozes 2 e 6. Nas restantes, a tendência de classificação dos peritos foi mais concordante.

Gráfico 10: Box-Plot das 10 vozes “originais” e respetivas “repetições” do parâmetro R-Rugosidade da GRBAS, para os 10 peritos



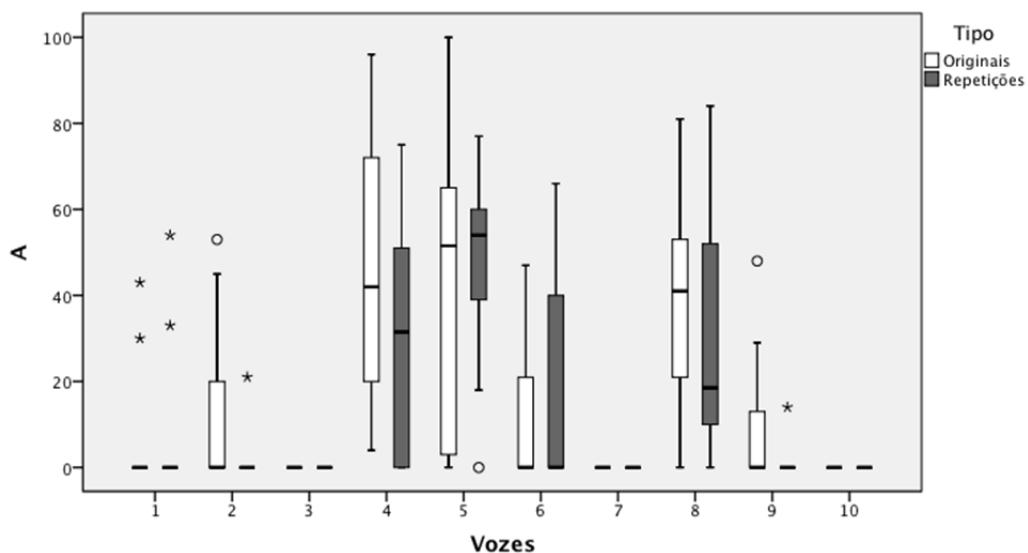
Para o parâmetro R(Rugosidade) (Cf. Gráfico 10), observa-se uma variabilidade de classificações em torno da média maior para as vozes 5 e 6. Nas restantes a tendência de classificação dos peritos foi mais concordante.

Gráfico 11: Box-Plot das 10 vozes “originais” e respetivas “repetições” do parâmetro B-Soprosidade da GRBAS, para os 10 peritos



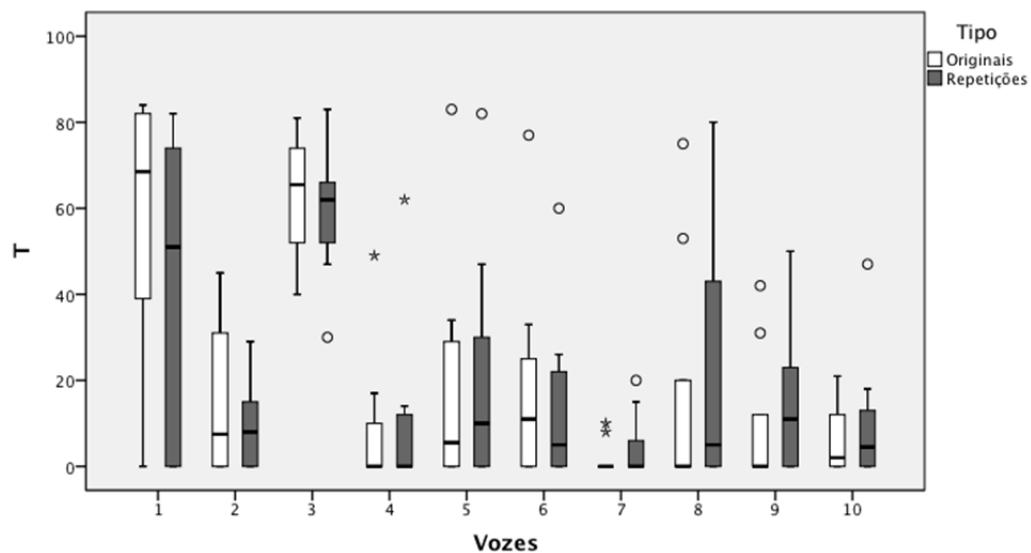
Para o parâmetro B(Soprosidade) (Cf. Gráfico 11), observa-se uma variabilidade de classificações em torno da média maior para as vozes 1, 4 e 6. Nas restantes a tendência de classificação dos peritos foi mais concordante.

Gráfico 12: Box-Plot das 10 vozes “originais” e respetivas “repetições” do parâmetro A-Astenia da GRBAS, para os 10 peritos



Para o parâmetro A(Astenia) (Cf. Gráfico 12), observa-se uma variabilidade de classificações em torno da média maior para as vozes 4, 5, 6 e 8. Nas restantes a tendência de classificação dos peritos foi mais concordante. Saliente-se a concordância quase total para as vozes 1, 3, 7 e 10.

Gráfico 13: Box-Plot das 10 vozes “originais” e respetivas “repetições” do parâmetro S-Tensão da GRBAS, para os 10 peritos



Para o parâmetro S(Tensão) (Cf. Gráfico 13), observa-se uma variabilidade de classificações em torno da média maior para as vozes 1 e 8. Nas restantes a tendência de classificação dos peritos foi mais concordante.

4.2.5.3.2. Estudo Parcelar 2 – Análise e caracterização da vocação das 10 vozes para gerarem consenso nas avaliações dos peritos

Pelo facto do CCIC ser mais baixo para os parâmetros A(Astenia) e S(Tensão) supradescritos (dado discutido posteriormente, no Capítulo 4, secção 3.4. e 3.5.), então estes não serão tidos em conta para determinar o grau de facilidade de avaliação e a vocação das vozes para conduzirem a resultados mais consensuais entre os peritos, para as 10 vozes (Cf. Quadro 17). Deste modo, e atendendo ao facto da classificação em tendência central dada pelos 10 peritos a cada componente da GRB diferir de modo estatisticamente significativo (valores de $p < 0.05$), verificada com base nos resultados do teste não paramétrico de Wilcoxon, foi realizada a análise da consistência inter-avaliador, para as 10 vozes repetidas no total do conjunto de vozes classificadas, de modo a apurar as que se revelam vocacionadas para gerarem avaliações mais consensuais entre os diferentes avaliadores.

Determinou-se, então, o coeficiente $\rho_G[i]$ que representa a correlação linear para o parâmetro G(Grau Geral) entre as avaliações realizadas pelos 10 avaliadores da

primeira (i.e. “originais”) para a segunda audição (i.e. “repetições”) do trecho de voz com índice i . Os coeficientes $\rho_R[i]$ e $\rho_S[i]$ têm o mesmo significado para os parâmetros R(Rugosidade) e S(Soprosidade). Cada um destes coeficientes representa uma aproximação prática do coeficiente de Pearson, que consiste na razão entre a covariância entre dois vetores e o produto dos respetivos desvios-padrão. Têm portanto uma gama normalizada entre o valor -1 e a unidade. Neste sentido, assumindo independência estatística entre as avaliações realizadas para as diferentes vozes, define-se a correlação global para o parâmetro G(Grau Geral) como o produtório das correlações obtidas para o conjunto das dez vozes de teste:

$$\rho_G = \left| \prod_{i=1}^{10} \rho_G[i] \right|, \quad i = 1, 2, \dots, 10. \quad (2.1)$$

Quanto mais este valor se aproximar da unidade, maior é a consistência entre a primeira e a segunda avaliação realizada para as 10 vozes, pelos 10 avaliadores. Toma-se o valor absoluto para assegurar que um número ímpar de correlações negativas não produza um resultado global negativo. Similarmente, definem-se ρ_R e ρ_S . A importância relativa de cada um destes parâmetros é traduzida num peso dado por:

$$\begin{aligned} p_G &= \frac{\rho_G}{\rho_G + \rho_R + \rho_S}, \\ p_R &= \frac{\rho_R}{\rho_G + \rho_R + \rho_S}, \\ p_S &= \frac{\rho_S}{\rho_G + \rho_R + \rho_S}. \end{aligned} \quad (2.2)$$

Estes pesos exprimem uma possível medida de fiabilidade relativa de cada parâmetro no apuramento de avaliações consistentes inter-avaliador para as diferentes vozes. Com base nos dados aproximaram-se os valores realmente obtidos ($p_G = 0.15$, $p_R = 0.7$, $p_S = 0.15$) para $p_G = 0.25$, $p_R = 0.5$, $p_S = 0.25$ de modo a evitar uma excessiva polarização imposta pelo peso devido ao parâmetro R(Rugosidade), o que poderia, no limite, tornar insignificantes os restantes dois. Em todo o caso, estes

resultados configuram alguma surpresa porque a partir dos Gráficos 9, 10, e 11 seria de esperar um maior peso para o parâmetro G(Grau Geral). Os resultados revelam, contudo, que o parâmetro R(Rugosidade) traduz uma referência interna de avaliação mais coerente para diferentes avaliadores. Os valores adotados, julgados mais equilibrados são, deste modo, úteis para atuarem como fator de ponderação em critérios posteriores, tal como se referirá em relação à equação (2.4).

Portanto, considerando que a classificação em tendência central dada pelos 10 peritos a cada componente da GRB difere de modo estatisticamente significativo (valores de $p < 0.05$) mas com uma polarização diferencial sobre os parâmetros G(Grau Geral), R(Rugosidade), e B(Soprosidade) então, realizou-se o estudo das classificações áudio-perceptivas, centrando a atenção na classificação destas dez vozes com um ajuste do peso da classificação do $G=0,25$; $R=0,50$ e $S=0,25$. Concluiu-se, como também se detalha a seguir, que a classificação dos peritos nas vozes 1, 5 e 6 diferia de forma mais acentuada, pelo que foram excluídas da análise subsequente.

De facto, uma avaliação estatística pertinente, como ilustrado nos Gráficos 9, 10 e 11, é o desvio padrão das avaliações efetuadas pelos 10 avaliadores para cada uma das dez vozes (índice i), calculado quer na primeira apresentação ($\sigma_{G1}[i], \sigma_{R1}[i], \sigma_{S1}[i]$), quer na segunda apresentação ($\sigma_{G2}[i], \sigma_{R2}[i], \sigma_{S2}[i]$). Quanto menor for cada um destes valores, maior é a consistência inter-avaliador para a voz em causa.

De modo a realizar uma seleção das vozes geradoras de apreciações mais consensuais, adotou-se um critério que combina para cada trecho de voz com índice i , a correlação linear entre as avaliações realizadas pelos 10 avaliadores entre a primeira e a segunda audição, e o produto dos desvios-padrão das avaliações realizadas pelos 10 avaliadores, quer para primeira audição, quer, independentemente, para a segunda audição. Por exemplo, para o parâmetro G(grau Geral) será:

$$f_G[i] = \frac{\rho_G[i]}{\sigma_{G1}[i] \times \sigma_{G2}[i]}. \quad (2.3)$$

Finalmente, para cada trecho de voz com índice i , obtém-se uma pontuação de desempenho calculada através de:

$$d[i] = p_G f_G[i] + p_R f_R[i] + p_S f_S[i]. \quad (2.4)$$

Aplicada a equação anterior a cada uma das dez vozes, chegou-se a uma seriação que conduziu à exclusão das vozes 1, 5 e 6, devido a obterem valores de $d[i]$ comparáveis e inferiores aos das restantes vozes. Curiosamente, a voz com número 6 é a que possui maior variabilidade de avaliação de acordo como parâmetro G(Grau Geral), a voz com número 5 é que possui maior variabilidade de avaliação de acordo como o parâmetro R(Rugosidade), e a voz com o número 1 é uma das que possui maior variabilidade de avaliação de acordo com o parâmetro B(Soprosidade).

4.2.5.3.3. Estudo Parcelar 3 – Análise e caracterização do consenso na avaliação áudio-percetiva por parte dos peritos (Consistência Intra-Avaliador)

Por último foi feito o estudo das classificações áudio-percetivas, centrando ainda mais a atenção no comportamento dos peritos para as 7 vozes selecionadas. No quadro 17 apresentam-se os valores da média e respetivos desvios padrão das componentes da GRB para as vozes “originais” (G1, R1, B1) e respetivas “repetições” (G2, R2, B2) de 7% do total da amostra (incluídas as repetições), de acordo com os dez peritos em estudo. Pretendeu-se, com base nesta análise de dispersão, selecionar os avaliadores que revelaram maior consistência na avaliação intra-avaliador, de modo a integrarem os estudos subsequentes de correspondência entre parâmetros acústicos e áudio-percetivos.

Os resultados evidenciam que foram encontradas diferenças estatisticamente significativas ($p < 0.05$) relativamente aos PERITOS 1, 6, 9 e 10. Feita a análise da consistência intra-avaliador foram portanto, numa primeira fase, excluídos os peritos 1, 6, 9 e 10 – aqueles que avaliaram de forma significativamente diferente o conjunto

das 7 vozes selecionadas, nos dois momentos de avaliação previstos ($p < 0.05$) – o que se traduz numa pobre consistência intra-avaliador.

Quanto ao PERITO 1, as classificações atribuídas às componentes G(Grau Geral) ($p = 0.028$) e B(Soprosidade) ($p = 0.043$) foram significativamente superiores nas “repetições” do que nas vozes “originais”. De igual forma, o PERITO 6, classificou as componentes G(Grau Geral) ($p = 0.047$) e R(Rugosidade) ($p = 0.009$) de forma significativamente superior nas “repetições” do que nas vozes “originais”. O PERITO 9 classificou a componente G(Grau Geral) ($p = 0.028$) de forma significativamente superior nas “repetições” do que nas vozes “originais”. O PERITO 10 classificou a componente B(Soprosidade) ($p = 0.005$) de forma significativamente superior nas “repetições” do que nas vozes “originais”.

Constata-se que para os restantes PERITOS (2, 3, 4, 5, 7, 8), não foram encontradas diferenças estatisticamente significativas entre os valores atribuídos às componentes da GRB, para as vozes “originais” e respetivas “repetições” ($p > 0.05$). Estes serão, assim, os seis peritos em avaliação áudio-percetiva que contribuirão com as respetivas quantificações das 90 vozes para os cálculos da regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas.

Quadro 17: Estatísticas descritivas e Coeficiente de Correlação Intra-Classe para as componentes da GRB de acordo com os peritos

Escala GRBAS	Perito																			
	1		2		3		4		5		6		7		8		9		10	
	M	DP	M	DP	M	DP	M	DP	M	DP	M	DP	M	DP	M	DP	M	DP	M	DP
<i>G*</i>	p < 0.05		ns		ns		ns		ns		p < 0.05		ns		ns		p < 0.05		ns	
G1	39.7	32.8	45.7	30.5	40.4	18.8	35.3	29.7	19.6	17.8	45.1	20.9	20.0	26.5	28.9	30.3	25.3	29.9	53.0	31.1
G2	46.0	25.4	29.9	25.3	39.1	13.4	36.0	29.6	18.6	18.6	49.3	23.1	24.3	21.5	28.6	23.6	16.9	29.2	33.4	23.5
<i>R*</i>	ns		ns		ns		ns		ns		p < 0.05		ns		ns		ns		ns	
R1	15.4	20.0	17.9	18.2	24.0	23.4	23.6	28.3	14.3	10.3	22.4	17.4	8.6	15.7	6.9	9.4	5.3	9.1	42.1	26.0
R2	20.0	22.0	19.3	22.5	24.1	21.6	21.7	25.2	18.3	19.1	31.9	19.3	10.0	15.3	15.4	17.4	5.0	8.9	42.4	24.7
<i>B*</i>	p < 0.05		ns		ns		ns		p < 0.05		ns									
S1	23.4	33.0	29.4	38.4	22.9	23.7	27.0	26.7	32.9	33.2	30.4	26.7	0.0	0.0	14.4	20.4	21.6	24.4	5.0	2.7
S2	26.9	31.5	28.0	25.9	27.4	17.0	20.9	19.2	29.7	29.0	34.0	33.1	0.0	0.0	5.7	10.4	14.0	29.2	37.1	23.0

M = Média; DP = Desvio Padrão.

* Resultados de acordo com o teste não paramétrico de Wilcoxon para duas amostras emparelhadas, a 95% de confiança.

4.2.6. Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas.

Para cada programa de software (*Dr. Speech*, *MDVP*, *Praat* e *Voice Studio*) foi utilizada a Regressão Linear Múltipla (RLM) com seleção de variáveis *Stepwise*, para obter um modelo que permitisse predizer as componentes da escala GRBAS (variáveis dependentes) em função das medidas acústicas (variáveis independentes) (Cf. Anexo VI). Analisaram-se os pressupostos do modelo, nomeadamente o da distribuição normal, homogeneidade e independência dos erros. Os dois primeiros pressupostos foram validados graficamente e através do teste de Kolmogorov-Smirnov (valor $p > 0.05$) e o pressuposto da independência foi avaliado com a estatística de Durbin-Watson (todos os $d \approx 2$). Utilizou-se o “Fator de Inflação da Variância” (do anglo-saxónico *Variance Inflation Factor*) - VIF, para diagnosticar a multicolinearidade (todos os VIF’s < 5). Procedeu-se também à eliminação das observações *outliers*, isto é, observações com um resíduo *studentizado* (*studentized residual*), em valor absoluto, superior a 1,96. Adicionalmente, para testar as hipóteses de que cada resíduo não é um *outlier* (H_0) versus é um *outlier* (H_1), calculou-se o valor-p associado a cada um dos valores da variável *sdr* (*Studentized Deleted Residual*), uma vez que cada um destes resíduos possui distribuição *t-student* com $(n-p-1)$ graus de liberdade (Moroco, 2010).

4.2.6.1. Regressão Linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o software *Dr. Speech*.

No quadro 18 apresentam-se os valores dos coeficientes de regressão normalizados (*standardized coefficients of predictors* – β - Beta values) e dos coeficientes de determinação ajustados (*Adjusted R Square* - R^2_a) relativos às regressões lineares múltiplas realizadas de forma a avaliar o poder preditivo das medidas acústicas em estudo nas diferentes componentes da escala GRBAS. Verifica-se que os preditores significativos ($p < 0.05$) das componentes da escala GRBAS, são:

- “G” – HNRdB ($\beta = -0.687$; $t(80) = -9.97$; $p < 0.001$) e SDF0Hz ($\beta = 0.329$; $t(80) = 4.77$; $p < 0.001$);
- “R” – Shimmerlocal ($\beta = 0.650$; $t(76) = 9.25$; $p < 0.001$), MeanF0Hz ($\beta = -0.344$; $t(76) = -4.67$; $p < 0.001$) e SDF0Hz ($\beta = 0.230$; $t(76) = 3.11$; $p < 0.01$);

- “B” – HNRdB ($\beta = -0.784$; $t(77) = -6.62$; $p < 0.001$), MeanF0Hz ($\beta = 0.357$; $t(77) = 5.44$; $p < 0.001$); jitterlocal ($\beta = -0.455$; $t(77) = -3.93$; $p < 0.001$) e Shimmerlocal ($\beta = 0.372$; $t(77) = 2.62$; $p < 0.01$);
- “A” – HNRdB ($\beta = -0.817$; $t(79) = -5.96$; $p < 0.001$) e jitterlocal ($\beta = -0.336$; $t(79) = -2.45$; $p < 0.05$);
- “S” – SDF0Hz ($\beta = 0.492$; $t(82) = 5.12$; $p < 0.01$).

Quadro 18: Resultados das regressões lineares múltiplas com seleção de variáveis *Stepwise*, das medidas acústicas sob as componentes da escala GRBAS para o software *Dr. Speech*: coeficientes normalizados dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$) (apenas são mostrados os resultados significativos)

	G	R	B	A	S
MeanF0Hz		-0.344***	0.357***		
SDF0Hz	0.329***	0.230***			0.492***
HNRdB	-0.687***		-0.784***	-0.817***	
Jitterlocal			-0.455**	-0.336*	
JitterPPQ5					
Shimmerlocal		0.650***	0.372***		
ShimmerAPQ5					
Modelo	***	***	***	***	***
R^2_a	0.61	0.64	0.66	0.35	0.23

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

R^2_a = Coeficiente de determinação ajustado.

Vozes *Outliers*: G=1,5,40; R=47,71,75,90,91,99; B=14,56,78,93; A=55,79,94,99; S=66,99.

Os modelos finais ajustados são, então:

- $G = 100.829 - 2.587HNRdB + 6.261SDF0Hz$
- $R = 30.436 + 8.466ShimmerLocal - 0.127MeanF0Hz + 4.217SDF0Hz$
- $B = 88.010 - 3.464HNRdB + 0.158MeanF0Hz - 43.353Jitterlocal + 4.480Shimmerlocal$
- $A = 105.482 - 3.200HNRdB - 27.360Jitterlocal$
- $S = 2.593 + 9.067SDF0Hz$

Em suma, as medidas acústicas que se associam de forma significativa a valores mais baixos das componentes da escala GRBAS (sinal negativo nos coeficientes de regressão) são: HNRdB (G); MeanF0Hz (R); HNRdB e jitterlocal (B); HNRdB e jitterlocal (A). Por sua vez, as variáveis que se associam de forma significativa a valores mais elevados das componentes da escala GRBAS (sinal positivo nos coeficientes de regressão) são: SDF0Hz (G); Shimmerlocal e SDF0Hz (R); Shimmerlocal e MeanF0Hz (B) e SDF0Hz (S).

Todos os modelos de regressão múltipla foram altamente significativos ($p < 0.001$) e, de uma forma geral, explicam uma proporção moderada das componentes da GRBAS ($0.35 < R^2_a < 0.66$), exceto para S(Tensão) ($R^2_a = 0.23$). Isto é, os modelos ajustados explicam uma proporção moderada (35% a 66%; $p < 0.001$) da variação observada nas componentes da GRBAS.

4.2.6.2. Regressão Linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o software MDVP.

No quadro 19 apresentam-se os valores dos coeficientes de regressão normalizados (*standardized coefficients of predictors – β - Beta values*) e dos coeficientes de determinação ajustados (*Adjusted R Square - R²_a*) relativos às regressões lineares múltiplas realizadas de forma a avaliar o poder preditivo das medidas acústicas em estudo nas diferentes componentes da escala GRBAS, de acordo com o programa *MDVP*. Verifica-se para este *software* que os preditores significativos ($p < 0.05$) das componentes da escala GRBAS, são:

- “G” – Shimmerlocal ($\beta = 0.926$; $t(85) = 7.94$; $p < 0.001$) e SDF0Hz ($\beta = -0.266$; $t(85) = -2.28$; $p < 0.05$);
- “R” – Shimmerlocal ($\beta = 0.742$; $t(83) = 10.08$; $p < 0.001$);
- “B” – JitterPPQ5 ($\beta = 0.692$; $t(81) = 6.60$; $p < 0.001$), SDF0Hz ($\beta = -0.522$; $t(81) = -5.61$; $p < 0.001$); Shimmerlocal ($\beta = 491$; $t(81) = 4.32$; $p < 0.001$); MeanF0Hz ($\beta = 0.197$; $t(81) = 2.87$; $p < 0.01$)
- “A” – JitterPPQ5 ($\beta = 0.546$; $t(82) = 5.89$; $p < 0.001$);
- “S” – SDF0Hz ($\beta = 0.420$; $t(88) = 4.35$; $p < 0.001$).

Quadro 19: Resultados das regressões Lineares múltiplas, com seleção de variáveis *Stepwise*, das medidas acústicas sob as componentes da escala GRBAS, para o software *MDVP*: coeficientes normalizados dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$) (apenas são mostrados os resultados significativos)

	G	R	B	A	S
MeanF0Hz			0.197***		
SDF0Hz	-0.266*		-0.522***		0.420***
HNRdB					
Jitterlocal					
JitterPPQ5			0.692***	0.546***	
Shimmerlocal	0.926***	0.742***	0.491***		
ShimmerAPQ5					
Modelo	***	***	***	***	***
R^2_a	0.53	0.55	0.68	0.29	0.17

* p < 0.05; ** p < 0.01; *** p < 0.001.

R^2_a = Coeficiente de determinação ajustado.

Vozes *Outliers*: G=78,98; R=34,73,90,91,99; B=1,4,13,73; A=3,55,73,94,1,34; S= não tem *outliers*.

Os modelos finais ajustados são, então:

- $G = 30.977 + 4.175 \text{Shimmerlocal} - 0.359 \text{SDF0Hz}$
- $R = 11.824 + 4.504 \text{Shimmerlocal}$
- $B = -6.022 + 21.281 \text{JitterPPQ5} - 0.961 \text{SDF0Hz} + 3.075 \text{Shimmerlocal} + 0.081 \text{MeanF0Hz}$
- $A = 8.081 + 16.731 \text{JitterPPQ5}$
- $S = 16.516 + 0.625 \text{SDF0Hz}$

Em suma, as medidas acústicas que se associam de forma significativa a valores mais baixos das componentes da escala GRBAS (sinal negativo nos coeficientes de regressão) são: SDF0Hz (G) e SDF0Hz (B). Por sua vez, as variáveis que se associam de forma significativa a valores mais elevados das componentes da escala GRBAS (sinal positivo nos coeficientes de regressão) são: Shimmerlocal (G); Shimmerlocal (R); MeanF0Hz, JitterPPQ5, Shimmerlocal (B); JitterPPQ5 (A) e SDF0Hz (S).

Todos os modelos de regressão múltipla foram altamente significativos ($p < 0.001$) e, de uma forma geral, explicam uma proporção moderada das componentes da GRBAS ($0.53 < R^2_a < 0.68$), exceto para A(Astenia) ($R^2_a = 0.29$) e S(Tensão) ($R^2_a = 0.17$).

4.2.6.3. Regressão Linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o software Praat.

No quadro 20 estão representados os coeficientes de regressão normalizados (*standardized coefficients of predictors – β - Beta values*) e os coeficientes de determinação ajustados (*Adjusted R Square - R^2_a*) relativos às regressões lineares múltiplas que possibilitam estimar o poder preditivo das medidas acústicas em estudo nas diferentes componentes da escala GRBAS, para o software *Praat*. Observa-se que os preditores significativos ($p < 0.05$) das componentes da escala GRBAS, são:

- “G” – HNRdB ($\beta = -0.838$; $t(77) = -13.48$; $p < 0.001$);
- “R” – HNRdB ($\beta = -580$; $t(78) = -8.04$; $p < 0.001$), SDF0Hz ($\beta = 0.298$; $t(78) = 4.18$; $p < 0.001$), MeanF0Hz ($\beta = -0.234$; $t(78) = -3.67$; $p < 0.001$);
- “B” – HNRdB ($\beta = -0.836$; $t(83) = -14.15$; $p < 0.001$), MeanF0Hz ($\beta = 0.372$; $t(83) = 6.30$; $p < 0.001$);
- “A” – HNRdB ($\beta = -0.875$; $t(84) = -6.46$; $p < 0.001$), jitterlocal ($\beta = -0.514$; $t(84) = -3.80$; $p < 0.001$);
- “S” – SDF0Hz ($\beta = 0.673$; $t(83) = 8.29$; $p < 0.001$).

Conclui-se que as medidas acústicas que se associam de forma significativa a valores mais baixos das componentes da escala GRBAS (sinal negativo nos coeficientes de regressão) são: HNRdB (G); MeanF0Hz e HNRdB (R); HNRdB (B); HNRdB, Jitterlocal (A). Em oposição, as variáveis que se associam de forma significativa a valores mais elevados das componentes da escala GRBAS (sinal positivo nos coeficientes de regressão) são: SDF0Hz (R); MeanF0Hz (B) e SDF0Hz (S).

Todos os modelos de regressão múltipla foram altamente significativos ($p < 0.001$) e, de uma forma geral, explicam uma proporção boa das componentes da GRB ($0.68 < R^2_a < 0.71$).

Os parâmetros A(Astenia) e S(Tensão) obtiveram valores de correlação regular ($0.33 < R^2_a < 0.45$), respetivamente.

Quadro 20: Resultados das regressões Lineares múltiplas, com seleção de variáveis *Stepwise*, das medidas acústicas sob as componentes da escala GRBAS, para o *software Praat*: coeficientes normalizados dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$) (apenas são mostrados os resultados significativos)

	G	R	B	A	S
MeanF0Hz		-0.234***	0.372***		
SDF0Hz		0.298***			0.673***
HNRRdB	-0.838***	-0.580***	-0.836***	-0.875***	
Jitterlocal				-0.514***	
JitterPPQ5					
Shimmerlocal					
ShimmerAPQ5					
Modelo	***	***	***	***	***
R^2_a	0.70	0.68	0.71	0.33	0.45

* p < 0.05; ** p < 0.01; *** p < 0.001.

R^2_a = Coeficiente de determinação ajustado.

Vozes *Outliers*: G=4,5,40,57,78,90,91,94,96,98,99; R=3,47,58,71,73,90,91,99; B=4,5,40,56; A=55,79,99; S=21,66,79,85,99.

Os modelos finais ajustados são, então:

- $G = 102.166 - 2.640 \text{ HNRdB}$
- $R = 90.856 - 2.102 \text{ HNRdB} + 0.386 \text{ SDF0Hz} - 0.086 \text{ MeanF0Hz}$
- $B = 78.583 - 3.506 \text{ HNRdB} + 0.159 \text{ MeanF0Hz}$
- $A = 94.551 - 3.133 \text{ HNRdB} - 17.282 \text{ Jitterlocal}$
- $S = 13.533 + 0.797 \text{ SDF0Hz}$

4.2.6.4. Regressão Linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas de acordo com o software Voice Studio.

No quadro 21 estão patentes os coeficientes de regressão normalizados (*standardized coefficients of predictors* – β - Beta values) e os coeficientes de determinação ajustados (*Adjusted R Square* - R^2_a) relativos às regressões lineares múltiplas que indicam o poder preditivo das medidas acústicas em estudo nas diferentes componentes da escala GRBAS, para o software *Voice Studio*. Verifica-se que os preditores significativos ($p < 0.05$) das componentes da escala GRBAS, são:

- “G” – ShimmerAPQ5 ($\beta = 0.459$; $t(82) = 4.3$; $p < 0.001$), Jitterlocal ($\beta = 0.344$; $t(82) = 3.2$; $p < 0.001$);
- “R” – Shimmerlocal ($\beta = 0.646$; $t(80) = 8.57$; $p < 0.001$), MeanF0Hz ($\beta = -0.269$; $t(80) = -3.56$; $p < 0.001$);
- “B” – ShimmerAPQ5 ($\beta = 0.696$; $t(84) = 9.23$; $p < 0.001$), MeanF0Hz ($\beta = 0.321$; $t(84) = 4.27$; $p < 0.001$);
- “A” – HNRdB ($\beta = -0.427$; $t(83) = -3.38$; $p < 0.01$), SDF0Hz ($\beta = -0.827$; $t(83) = -4.23$; $p < 0.001$), jitterlocal ($\beta = 0.708$; $t(83) = 3.60$; $p < 0.01$);
- “S” – SDF0Hz ($\beta = 0.526$; $t(79) = 5.67$; $p < 0.001$).

No programa de software *Voice Studio*, as medidas acústicas que se associam de forma significativa a valores mais baixos das componentes da escala GRBAS (sinal negativo nos coeficientes de regressão) são: MeanF0Hz (R); HNRdB e SDF0Hz (A). Por sua vez, as variáveis que se associam de forma significativa a valores mais elevados das componentes da escala GRBAS (sinal positivo nos coeficientes de regressão) são: Jitterlocal e ShimmerAPQ5 (G); Shimmerlocal (R); MeanF0Hz e ShimmerAPQ5 (B); Jitterlocal (A) e SDF0Hz (S).

Todos os modelos de regressão múltipla foram altamente significativos ($p < 0.001$) e, de uma forma geral, explicam uma proporção moderada das componentes da GRBAS ($0.52 < R^2_a < 0.56$), exceto para A ($R^2_a = 0.26$) e S ($R^2_a = 0.27$).

Quadro 21: Resultados das regressões Lineares múltiplas com seleção de variáveis *Stepwise*, das medidas acústicas sob as componentes da escala GRBAS para o software *Voice Studio*: coeficientes normalizados dos preditores (β) e variância total ajustada ($Adj. R^2 = R^2_a$) (apenas são mostrados os resultados significativos)

	G	R	B	A	S
MeanF0Hz		-0.269***	0.321***		
SDF0Hz				-0.827***	0.526***
HNRdB				-0.427***	
Jitterlocal	0.344***			0.708***	
JitterPPQ5					
Shimmerlocal		0.646***			
ShimmerAPQ5	0.459***		0.696***		
Modelo	***	***	***	***	***
R^2_a	0.55	0.56	0.52	0.26	0.27

* p < 0,05; ** p < 0,01; *** p < 0,001.

R^2_a = Coeficiente de determinação ajustado.

Vozes *Outliers*: G=73,78,94,98,99; R=71,73,88,90,91,98,99; B=34,55,94; A=55,94; S=66,79,96,99.

Os modelos finais ajustados são, então:

- $G = 33.141 + 2.852 \text{ShimmerAPQ5} + 2.271 \text{Jitterlocal}$
- $R = 37.793 + 2.651 \text{Shimmerlocal} - 0.096 \text{MeanF0Hz}$
- $B = -8.210 + 5.997 \text{ShimmerAPQ5} + 0.135 \text{MeanF0Hz}$
- $A = 37.050 - 1.141 \text{HNRdB} - 0.733 \text{SDF0Hz} + 6.658 \text{JitterPPQ5}$
- $S = 12.926 + 0.372 \text{SDF0Hz}$

4.3 Discussão dos Resultados

4.3.1. Análise das medidas acústicas de acordo com o tipo de voz (completa ou segmentada) relativamente aos tipos de *software*

A compararação entre as medidas da análise acústica das vozes – completas e segmentadas – com pré-processamento e seleção das regiões estáveis das amostras vocais, ignorando as regiões transitórias de início e fim, foi realizada para três programas de *software*: *Dr. Speech*, *Praat* e *Voice Studio*.

Nas três aplicações todas as medidas acústicas estudadas não apresentam resultados estatisticamente significativos, exceto para o SDF0(Hz) (valores-p > 0.05, com 95% de confiança; Cf. Capítulo 4, Secção 2.1.). Independentemente do *software* usado esta foi a medida que apresentou valores significativamente mais elevados nas vozes completas ou não processadas.

Deste modo, os resultados confirmam a pertinência do procedimento de segmentação que possibilita dados mais consistentes e, como tal, com menor possibilidade de influência disruptiva decorrente dos transitórios do regime fonatório sobre a análise e interpretação dos valores, nomeadamente nos sinais de voz com algum tipo de perturbação – com sub-harmónicos e bifurcações ou mudanças súbitas da qualidade do padrão vibratório (Titze, 1995; Karnell et al., 1997; Dejonckere et al., 2001; Ma & Yiu, 2005; Sprecher et al., 2010; Olszewski et al., 2011). Estas conclusões também foram consideradas outros estudos (Karnell (1991), Bielamowicz et al., (1996), Jones et al., (2001), Yu et al., (2001), Kent et al., (2005), Moers et al., (2012) e Choi et al., (2012b)).

No entanto, todas as restantes medidas acústicas – MeanF0(Hz), Jitter(PPQ5)(%), Jitter(local)(%), Shimmer(APQ5)(%), Shimmer(local)(%) e HNR(dB) – apresentaram valores estatisticamente semelhantes, o que reflete que na amostra de vozes estudada o sinal não foi tão influenciado por ruído e outros fatores acústicos. Note-se contudo que, embora não

estatisticamente significativo, observou-se sempre de forma transversal às três aplicações um aumento da medida HNR(dB) das vozes completas para as segmentadas.

Ao contrário do expectável, dadas as diferenças entre os algoritmos de cálculo usados por cada programa de *software*, a semelhança de resultados reflete também a qualidade média dos sinais vocais estudados que tornam mais robusta a análise de sinal, facto este que poderá derivar dos cuidados metodológicos de recolha das vozes tidos em consideração para a construção do banco de dados de vozes usadas na dissertação, de acordo com o que está preconizado por Karnell (1991), Bielamowicz et al., (1996), Jones et al., (2001), Yu et al, (2001) e Kent et al., (2005), assim como da sua segmentação durante o pré-processamento. Os estudos referidos anteriormente ratificam esta conclusão, e são reforçados com o de Olszewski et al., (2011), que concluíram da pertinência do uso de métodos de selecção da amostra de sinal de modo a conseguir-se uma maior estabilidade e facilidade das medidas acústicas obtidas.

4.3.2. Análise das medidas acústicas (pares de vozes – MDVP-*Multi Dimensional Voice Program*) de acordo com a amplificação das amostras de voz

Em três das aplicações estudadas – *Dr. Speech*, *MDVP- Multi Dimensional Voice Program* e *Voice Studio* – existiam amostras de vozes que não possuíam amplitude suficiente para análise e recolha de medidas acústicas. No *software* com mais ocorrências deste tipo – *MDVP – Multi Dimensional Voice Program* – recorreu-se à amplificação (20dB) de 58 amostras de voz, procedimento este que permitia que cada sinal amplificado atingisse 50-75% da magnitude espectral, sem incorrer em fenómenos de *clipping*.

A comparação das médias e desvios padrão das medidas acústicas das 12 amostras finais – sem e com amplificação – não revelou diferenças estatisticamente significativas (valores-p > 0.05, com 95% de confiança). Aliás, as diferenças eram praticamente nulas (valores-p > 0.000), logo a amplificação gerou mudanças negligenciáveis nos resultados das medidas acústicas em estudo. Este resultado também foi encontrado por Moon et al. (2012).

Este facto reflete o impacto da diferença de amplitude no sinal vocal. Ao contrário do apontado por Orlikoff & Kahane (1991), Baken & Orlikoff (2000), Buder (2000) e Akpinar et al. (2012) que consideraram a influência da amplitude, em especial sobre o *Shimmer*, no programa de *software* MDVP – *Multi Dimensional Voice Program*, e nas amostras em estudo, a manipulação deste parâmetro apenas foi notado (embora residualmente) sobre o Jitter(jitt)%. Logo, concluiu-se que a mudança de amplitude até 20 dB é inócuia para a medição do sinal vocal.

4.3.3. Análise das medidas acústicas de acordo com o programa de *software* e coeficiente de correlação Ró de Spearman

As medidas descritivas para cada tipo de *software*, apresentadas no Capítulo 4, Secção 2.3., revelam que a MeanF0(Hz) não é estatisticamente diferente (valor-p > 0.05), logo a extração deste valor é transversal e equitativo para os quatro algoritmos de processamento de sinal estudados. O cálculo do coeficiente de correlação não paramétrico Ró de Spearman reforça esta conclusão, mostrando a existência de uma associação linear muito elevada, positiva e estatisticamente significativa ($Ró > 0.99$; $p < 0.01$) entre todos os tipos de *software*.

Porém, e de acordo com os resultados apresentados no Capítulo 4, Secção 2.4.2., o SDF0(Hz) já não reúne este consenso entre as aplicações. Saliente-se a grande sensibilidade do *Voice Studio* a este parâmetro (média=13.30Hz), o que sinaliza a sua capacidade para deteção de variações da F0 em torno da média, mensuração esta essencial para a caracterização de alterações da qualidade vocal que incluem o tremor (por exemplo, típico nas disfonias neurológicas) ou outras variações mínimas do sinal. A aplicação com menor ênfase nesta medida foi o *Dr. Speech* (média=1,79Hz). Tal facto deve-se ao algoritmo de cálculo deste programa de *software* ser mais insensível à variabilidade de F0 do sinal (Batalla et al., 2004). Note-se que a determinação do cálculo do coeficiente de correlação Ró de Spearman demonstra uma associação linear muito elevada entre as aplicações de *software* *Dr. Speech* e *Praat* ($Ró=0.91$; $p < 0.01$), e moderada entre os restantes tipos de *software* ($0.57 < Ró < 0.73$).

Relativamente às medidas de perturbação – Jitter(local)% , Jitter(PPQ5)% , Shimmer(local)% e Shimmer(APQ5)% – a mensuração de todas revelou-se diferente, com resultado estatisticamente significativo, entre os quatro programas de *software*. Saliente-se que:

- Para Jitter(local)% e Shimmer(local)% o *software* mais sensível, i.e. com médias mais altas destas medidas, foi o *Voice Studio* (média de Jitter(local)%=3.17% e Shimmer(local)%=5.80%). Em contraposição, o *Dr. Speech* revelou sempre os valores mais baixos (média de Jitter(local)%=3.17% e Shimmer(local)%=5.80%, respetivamente). Note-se que a determinação do cálculo do coeficiente de correlação Ró de Spearman demonstra uma associação linear muito elevada e estatisticamente significativa entre as aplicações de *software Dr. Speech* e *Praat* ($R_o=0.95$; $p < 0.01$), e moderada entre os restantes tipos de *software* ($0.67 < R_o < 0.73$), para o Jitter(local)%. Para a medida Shimmer(local)% o mesmo tipo de estatística inferencial mostra a existência de uma associação linear muito elevada, positiva e estatisticamente significativa ($R_o > 0.91$; $p < 0.01$) entre todos os tipos de *software*.
- Para Jitter(PPQ5)% e Shimmer(APQ5)% a tendência para valores mais altos mantém-se para o *Voice Studio* (Jitter(PPQ5)%=1.99%) e o MDVP (Shimmer(APQ5)%=3.94%), em oposição com os valores inferiores (Jitter(PPQ5)%=0.31% e Shimmer(APQ5)%=2.57%) no *Praat*, respetivamente. Estas conclusões também foram encontradas no estudo de Silva et al. (2009). O *Dr. Speech* não possui estas duas medidas de análise acústica. Note-se que a determinação do cálculo do coeficiente de correlação Ró de Spearman demonstra uma associação linear moderada e estatisticamente significativa entre as três aplicações de *software* ($0.68 < R_o < 0.74$; $p < 0.01$), para o Jitter(PPQ5)%. Para a medida Shimmer(APQ5)% existe uma associação linear elevada/muito elevada ($0.88 < R_o < 0.91$; $p < 0.01$), positiva e estatisticamente significativa entre os três tipos de *software*.

Recorda-se que as medidas dependentes de F0 sofrem a interferência da aperiodicidade do sinal e devem ser cuidadosamente aplicadas e interpretadas (Koike, 1967; Titze, 1995; Vieira et al., 1996; Buder, 2000; Vieira et al., 2002; Camargo et al., 2003; Brasolotto & Rehder, 2011), em especial quando se caracterizam vozes com irregularidades do sinal (Titze, 1995;

Karnell et al., 1997; Dejonckere et al., 2001; Ma & Yiu, 2005; Maryn et al., 2009a; Choi et al., 2012a).

No que concerne à medida de ruído, HNR (dB): o *software* que deteta valores mais baixos foi o *Voice Studio*, seguido do MDVP. O que menos salienta os sinais de ruído na voz é o *Dr. Speech* (média de HNR(dB)=22.85dB). Note-se que a determinação do cálculo do coeficiente de correlação Ró de Spearman demonstra uma associação linear muito elevada e estatisticamente significativa entre as aplicações de *software Dr. Speech* e *Praat* ($Ró=0.99$; $p < 0.01$), e moderada/elevada entre os restantes tipos de *software* ($0.61 < Ró < 0.79$).

4.3.4. Análise descritiva da avaliação áudio-percetiva (GRBAS)

As avaliações percetivas pelos peritos foram caracterizadas quanto às medidas descritivas (Cf. Capítulo 4, Secção 2.5.1.), o que revelou uma discrepância de cotações entre os parâmetros da escala GRBAS. Kreiman & Gerratt (2011) reforçam esta dificuldade de cotação áudio-percetiva em todas as suas publicações. Assim, as médias das cotações foram mais elevadas para os peritos 4 (B, Soprosidade e S, Tensão), 5 (G, Grau Geral e A, Astenia) e 7 (R, Rugosidade). A este propósito observa-se que o perito 4 é aquele que apresenta menor experiência prática (6 anos), e o 5 representa um elemento com contacto com a componente de académica de investigação e docência na área da voz. Quanto ao perito 7, a sua classificação mais elevada do parâmetro *Roughness* (Rugosidade) pode ir de encontro à perspectiva defendida por Moers et al. (2012) que considera esta medida uma supra-classe que aglutina diversos parâmetros percetivos. Em contraponto, as médias mais baixas são encontradas para os peritos 3 e 6, com o denominador comum da prática clínica essencialmente com voz profissional. Sofranko & Prosek (2012) compararam 3 grupos com diferentes experiências de avaliação áudio-percetiva, e concluíram que os professores de canto (em contraposição com os terapeutas da fala) tendem a realizar julgamentos mais estéticos e centrados nas suas experiências académicas/profissionais.

Note-se que a análise individual dos parâmetros da escala GRBAS evidencia que o mais valorizado é o parâmetro súmula por excelência, o G (ou Grau Geral de Disfonia). Esta

conclusão foi também obtida pelos estudos de Hirano et al. (1986), Eskenazi et al. (1990), Feijoo & Hernández (1990), Dejonckere et al. (1996), de Bodt (1997), Wolfe & Martin, (1997), Yu et al. (2001), Heman-Ackah et al. (2002), Shrivastav et al. (2005), Yu et al. (2007), Yamasaki et al. (2008) e Choi et al. (2012a). O parâmetro A (Astenia) foi o menos evidenciado. Reforça-se a dificuldade de cotação áudio-percetiva deste aspeto, tendência esta também encontrada por Dediwit et al. (2008), razão pela qual não faz parte de alguns dos protocolos de classificação áudio-percetiva (*German RBH evaluation scheme*, Nawka et al., 1994; GRB da *European Laryngological Society*, Dejonckere et al. 2001).

4.3.5. Análise da Consistência Inter-Avaliador (CCIC)

A quantificação do coeficiente de correlação intra-classe (CCIC), de acordo com os resultados descritos no Capítulo 4, Secção 2.5.2., permite concluir que os peritos assumem uma tendência para classificarem cada parâmetro da avaliação áudio-percetiva de modo semelhante. No entanto, o valor numérico atribuído foi diferente, de modo estatisticamente significativo ($p < 0.05$).

Observa-se que os valores dos intervalos de confiança a 95% das correlações entre as classificações médias dos peritos (CCIC) está acima de 0.9 para os parâmetros G(Grau Geral), R(Rugosidade), e B(Soprosidade). O A (Astenia) e S (Tensão) têm uma consistência inferior. Estas conclusões são ratificadas por Dejonckere et al. (1993), De Bodt et al. (1997), Oates (2009) e Moers et al. (2012). Aqui se manifesta, novamente, a tendência de maior precisão na atribuição das avaliações áudio-percetivas para os parâmetros G(Grau Geral) (tal como no estudo de Choi et al., 2012a e Moers et al., 2012), R(Rugosidade) (concomitante com os dados de Wolfe & Steinfatt, 1987; Wolfe & Martin, 1997; Wolfe et al., 1997 e Halberstam, 2004) e B(Soprosidade) (Boucher, 2008), dados estes já revistos (Cf. Capítulo 2, Secção 1.2.), desta dissertação. Os parâmetros A(Astenia) e S(Tensão) são excluídos de algumas das escalas de avaliação devido à reduzida consistência inter-avaliadores e também da análise seguinte da dissertação.

4.3.6. Análise da Consistência Intra-Avaliador

Considerando os resultados do CCIC realizou-se uma análise da diferença entre as cotações obtidas para as vozes “originais” e respetivas “repetições”, já que havia a evidência de que as avaliações das vozes eram variáveis, consoante as suas características. Assim sendo, da análise das 10 vozes repetidas na amostra da dissertação, tal como detalhado no Capítulo 4, Secção 2.5.3., conclui-se da maior vocação para um consenso entre os peritos de um total de 7 vozes, as quais foram depois alvo de análise de dispersão para concluir que 4 dos avaliadores apresentavam diferenças estatisticamente significativas ($p < 0.05$) nas suas cotações das vozes “originais” e “repetições”.

Note-se que os peritos excluídos foram, na sua maioria, pouco consistentes nas cotações aos parâmetros G(Grau) ou B(Soprosidade). Estes fatores são, na bibliografia (Pinho, 2003; Butha et al., 2005; Dogan, 2008; Pinho & Pontes, 2008), considerados de relativamente fácil avaliação áudio-percetiva, o que não aconteceu, facto passível de ser relacionado com a diferença de experiência profissional do painel de juízes em causa, nomeadamente ao contacto privilegiado com casuística de voz profissional, na qual a existência de perturbação da voz nem sempre acontece. É ainda de referir que dois destes peritos são do género masculino. Tal como afirmam Moon et al. (2012), a cotação de uma dada voz pode ser influenciada pelo género do falante. Não é conhecida nenhuma investigação publicada, até à data, que reveja a influência do género do avaliador na cotação dos parâmetros áudio-percetivos.

4.3.7. Regressão linear múltipla para predizer a associação entre as componentes da escala GRBAS e as medidas acústicas

Foi construído um modelo de Regressão Linear Múltipla para cada um dos quatro programas de *software* que avaliou o poder preditivo das medidas acústicas em estudo relativamente às diferentes componentes da escala GRBAS.

Os coeficientes de determinação ajustados (*Adjusted R Square* – R^2_a), em todos os parâmetros, variam de acordo com os programas de *software*, o que é compreensível

atendendo às diferenças de algoritmia adotadas, apontadas também por Silva et al. (2009) e Manfredi et al. (2012) como justificativas para as discrepâncias notadas entre os resultados das suas investigações.

Todos os modelos de regressão múltipla foram altamente significativos ($p<0.001$) e reforçam a noção de que a variabilidade dos parâmetros A(Astenia), ($0.26 < R^2_a < 0.35$) e S(Tensão), ($0.17 < R^2_a < 0.45$) é mal explicada pelas medidas acústicas. Estes dados foram encontrados por outros autores (Hirano et al., 1986; Eskenazi et al., 1990; Feijoo & Hernández, 1990; Dejonckere et al., 1996; de Bodt, 1997; Wolfe & Martin, 1997; Yu et al., 2001; Heman-Ackah et al., 2002; Shrivastav et al., 2005; Yu et al., 2007; Choi et al., 2012a), e suportam a opção de retirada destes parâmetros dos protocolos de avaliação percetiva, nomeadamente os de referência para a Europa (*German RBH evaluation scheme*, Nawka et al., 1994; GRB da *European Laryngological Society*, Dejonckere et al., 2001). Para colmatar esta dificuldade de definição e/ou classificação áudio-percetiva propõe-se uma reformulação da descrição dos mesmos, porém sabendo à partida que são muito difíceis de explicar, ouvir e classificar.

No que concerne ao programa de *software Dr. Speech*, verifica-se que as medidas acústicas que se assumem como preditores significativos são distintas de acordo com cada parâmetro percetivo, exceto o SDF0, que surge associado a três das cinco qualidades áudio-percetivas da GRBAS. Note-se que os modelos ajustados de regressão linear múltipla, neste tipo de *software*, explicam uma proporção boa da variação observada nas componentes G($R^2_a = 0.61$), R($R^2_a = 0.64$) e B($R^2_a = 0.66$).

Da observação dos resultados do programa de *software MDVP*, verifica-se que as medidas acústicas que se assumem como preditores significativos são distintas de acordo com cada parâmetro percetivo, exceto o SDF0(Hz) e o Shimmer(local)%¹, que surgem associados a três das cinco qualidades áudio-percetivas da GRBAS. A análise dos modelos ajustados de regressão linear múltipla, neste *software*, explica uma proporção boa da variação observada na componente B($R^2_a = 0.66$). Também consegue explicar de modo regular os parâmetros R($R^2_a = 0.55$) e o G($R^2_a = 0.53$).

Relativamente ao programa de *software Praat*, verifica-se que as medidas acústicas que se assumem como preditores significativos são distintas de acordo com cada parâmetro áudio-percetivo, exceto o HNR, que surge associado a quatro das cinco qualidades áudio-percetivas da GRBAS. Os modelos ajustados de regressão linear múltipla construídos, neste tipo de *software*, explicam uma proporção boa da variação observada nas componentes B($R^2_a = 0.71$), G ($R^2_a = 0.70$) e R($R^2_a = 0.68$). Note-se que é o único *software* que consegue explicar de modo regular o parâmetro S($R^2_a = 0.45$).

Analizando o programa de *software Voice Studio*, verifica-se de igual modo que as medidas acústicas que se assumem como preditores significativos são distintas de acordo com cada parâmetro áudio-percetivo. Neste há múltiplos parâmetros que se repetem para as diferentes qualidades áudio-percetivas da GRBAS. Saliente-se que os modelos ajustados de regressão linear múltipla, neste tipo de *software*, só explicam uma proporção regular da variação observada nas componentes R($R^2_a = 0.56$), G($R^2_a = 0.55$) e B($R^2_a = 0.52$).

Portanto, a análise destes dados aponta desde já para uma pobre discriminação áudio-percetiva conseguida pelas medidas acústicas SDF0(Hz), HNR(dB) e Shimmer(local)%. De acordo com os dados recolhidos e analisados, conclui-se da maior capacidade de predição acústica relativamente ao parâmetro áudio-percetivo B(Soprosidade), pois foi este que obteve o coeficiente de determinação ajustado (*Adjusted R Square* – R^2_a) mais forte, transversal a todos os programas de *software*. Esta conclusão foi também encontrada em Takahashi & Koike (1976), Eskenazi et al (1990), Feijoo & Hernández (1990), Dejonckere et al. (1993) e Dejonckere (1995).

Também se reforça a noção de impossibilidade de associação direta e exclusiva entre as medidas acústicas e cada um dos parâmetros áudio-percetivos. Esta conclusão foi também apontada por Behrman (2005) e Choi et al. (2012a). Porém, saliente-se que os programas *Praat* e *MDVP* foram os que recorreram a um menor número de medidas acústicas para explicar os parâmetros áudio-percetivos e o *Praat* e o *Dr. Speech* conseguiram coeficientes de determinação ajustados mais elevados. Esta melhor prestação do *Praat* é também enfatizada por Amir et al., (2009) e Silva et al., (2009).

4.4 Material desenvolvido ao longo da elaboração da dissertação:

4.4.1. Descrição da construção e funcionamento do material de treino e teste em termos de avaliação áudio-percetiva (Voice-PE – Voice – Perceptual Evaluation)

O programa Voice-PE (*Voice – Perceptual Evaluation*) foi produzido em C++ e Qt, uma vez que as bibliotecas deste reúnem todas as opções de interface gráfico que o programa apresenta. Surge como uma ferramenta de treino e teste para atuais e futuros profissionais interessados pelo domínio de competências em avaliação áudio-percetiva da voz humana. Pretende-se a sua distribuição gratuita, nomeadamente em escolas que formem terapeutas da fala, locais em que se realize atendimento clínico de casuística com perturbações vocais (serviços de otorrinolaringologia) e, ainda, escolas ou conservatórios de música e outros contextos que interajam com esta área da avaliação e caracterização vocal.

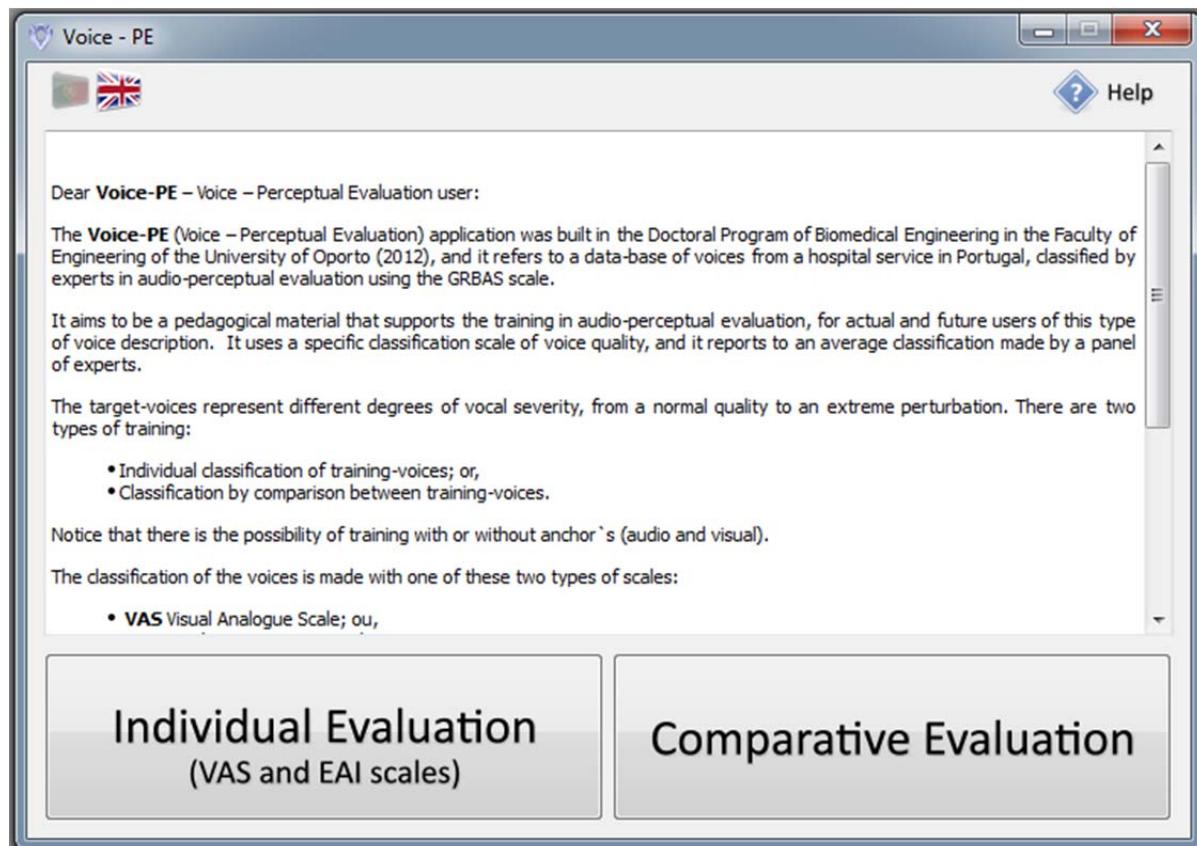


Figura 15: Primeira janela do Voice-PE, em Inglês.

A primeira janela do Voice-PE apresenta uma descrição do programa (*Cf. Figura 15*), com a indicação das suas funcionalidades de treino e teste em avaliação áudio-percetiva. Visualizam-se de imediato os botões representativos das duas formas de treino possíveis: classificação individual de vozes-treino *versus* classificação por comparação entre vozes-treino. Nesta primeira página existe ainda a opção de *Ajuda/Help*, assim como um ícone para a possibilidade de mudança de língua, uma vez que o programa está desenvolvido para o uso em Inglês e Português.

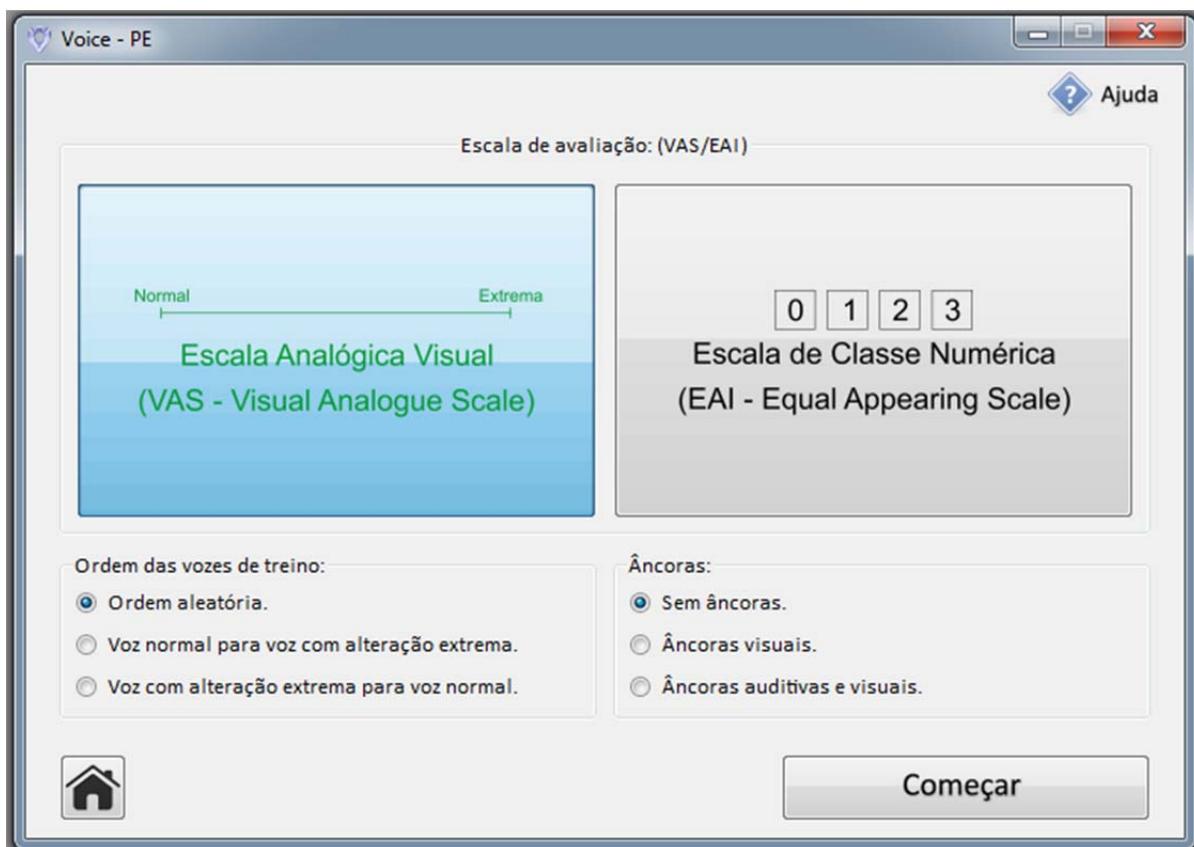


Figura 16: Janela da opção de treino com classificação individual das vozes

Na forma de treino com classificação individual, a mesma pode ser realizada através de duas escalas distintas (*Cf. Figura 16*). Ou é selecionada a escala visual analógica (VAS – *Visual Analogue Scale*), ou a escala de classe numérica (EAI – *Equal Appearing Interval*). Assim, são apresentados ao utilizador grupos de 15 vozes-treino (6 opções possíveis), cuja organização pode ser de 3 formas: ou aleatória, alocada pelo próprio sistema, ou gradativa, com vozes do grau normal ao extremo e vice-versa. Estas duas opções são mais fáceis para o ouvinte, pois este antecipa a graduação da alteração vocal. É ainda dada a possibilidade de recorrer a

âncoras, visuais, visuais e auditivas, ou fazer toda a classificação das vozes-treino sem qualquer tipo de ajuda/referencial de explicação do parâmetro percetivo em análise.

Quando o utilizador opta pela classificação visual analógica (VAS – *Visual Analogue Scale*) surge uma janela (*Cf. Figura 17*) em que pode – através da deslocação de um cursor deslizante sobre uma régua – avaliar cada parâmetro da GRBAS entre 0 e 100. Ao lado direito de cada régua é indicado o valor em que o cursor estaciona, ou seja, a cotação numérica que corresponde à avaliação daquela voz.



Figura 17: Janela da opção de treino com classificação individual das vozes através da escala visual analógica

Para a opção de classificação através de classe numérica (EAI – *Equal Appearing Interval*), a escala assume valores entre zero e três (*Cf. Figura 18*). O utilizador deve ouvir a voz e classificar cada um dos parâmetros clicando sobre o botão com o número que julga corresponder à avaliação daquela amostra sonora. No fim, aciona o botão seguinte para passar para a escuta e classificação de outra voz.

Tanto numa como noutra opção da classificação individual das vozes-treino, o utilizador observa à direita da janela uma barra de estado que representa o ponto do exercício de avaliação que está a ter lugar. Ou seja, a barra vai mudando de preenchimento até à sua totalização no final do grupo de vozes em estudo (Cf. Figura 17 e 18).



Figura 18: Janela da opção de treino com classificação individual das vozes através da escala numérica

É também possível, nas duas opções de classificação, recorrer a âncoras. Estas podem ser inexistentes (*sem âncora*), *visuais* ou *visuais e auditivas*. Deste modo, consoante a escolha realizada previamente pelo utilizador, este pode fazer toda a classificação das vozes-treino sem qualquer tipo de ajuda, ou ter um referencial – com a explicação visual (patente ao clicar sobre o ícone com um ponto de interrogação) ou/e os extremos de gravidade vocal de cada parâmetro percetivo em análise (escutados ao clicar sobre o ícone com um altifalante, de cor verde ou vermelha, para exemplos de qualidade vocal normal e extremamente alterada, respetivamente) (Cf. Figura 17 e 18).

Ao concluir a classificação das 15 vozes-treino é automaticamente criado um relatório. Este reflete a cotação individual de cada voz, para cada um dos parâmetros áudio-percetivos (GRBAS), comparando-a com a média dos seis juízes resultantes do trabalho desenvolvido ao longo da dissertação e com os pontos de corte referenciados por Yamasaki et al. (2007) para discernir entre os distintos graus de perturbação vocal (representados através de uma barra de cores) (Cf. Figura 19 – Resultados da Avaliação com o uso da escala VAS). Nesta fase é ainda possível escutar novamente a voz (clicando sobre o ícone de *play*, junto ao nome da amostra em causa), mas agora observando a pontuação dada pelo utilizador e a dos juízes.

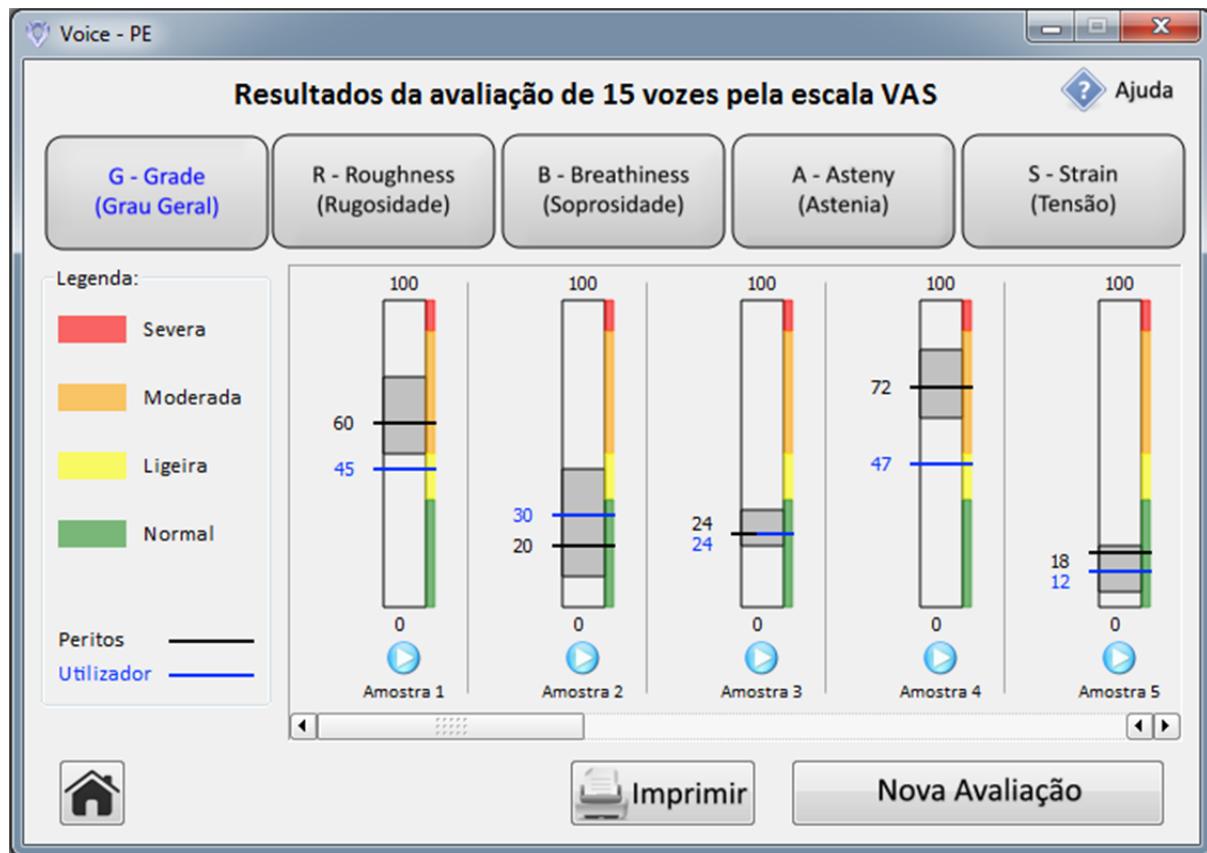


Figura 19: Exemplo do sumário da avaliação áudio-percetiva através da escala visual analógica

Este formato de relatório, criado automaticamente no fim de cada bloco de avaliações áudio-percetivas, é também possível quando se usa a escala EAI (Cf. Figura 20 – Resultados da Avaliação com o uso da escala EAI). As funcionalidades são as mesmas que foram descritas previamente.

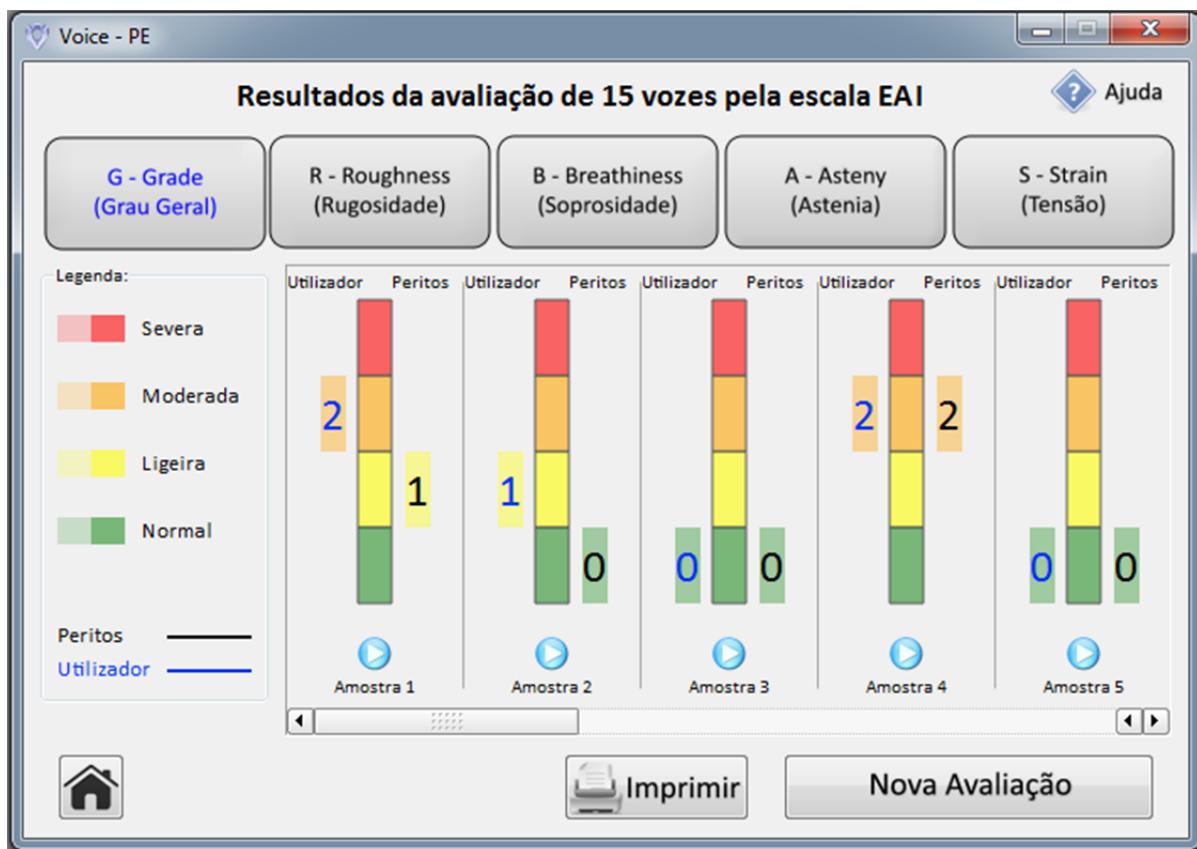


Figura 20: Exemplo do sumário da avaliação áudio-percetiva através da escala numérica

É dada a possibilidade de impressão do relatório, com a oportunidade de identificar o autor da avaliação e a data da mesma, com referência a todos os parâmetros (GRBAS), a cotação do utilizador e a dos juízes (*Cf.* Figura 21).

Na forma de treino com classificação por comparação entre vozes, a mesma implica a individualização e pré-seleção do parâmetro áudio-percetivo da escala GRBAS a treinar (*Cf.* Figura 22). Ou seja, é selecionado um e um só parâmetro (cuja definição escrita – âncora visual – está patente nesta página).

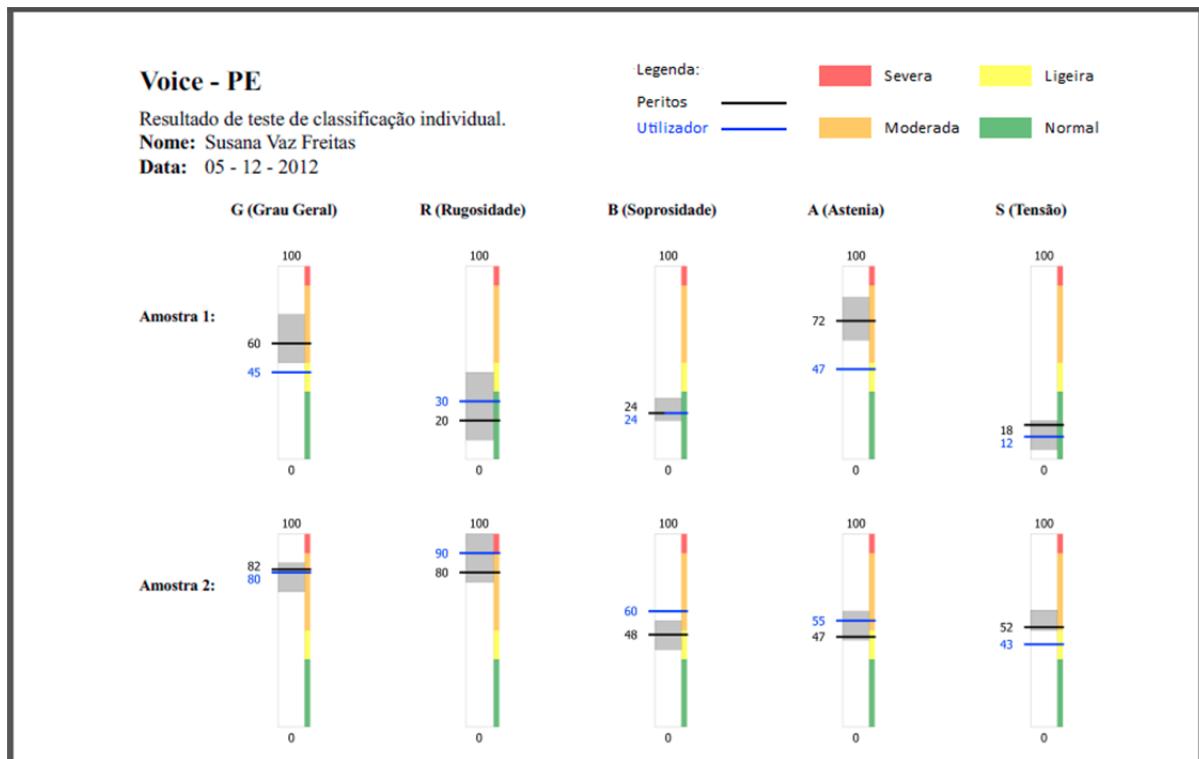


Figura 21: Exemplo parcial do relatório a imprimir sobre a avaliação áudio-percetiva, com uso da escala visual-analógica

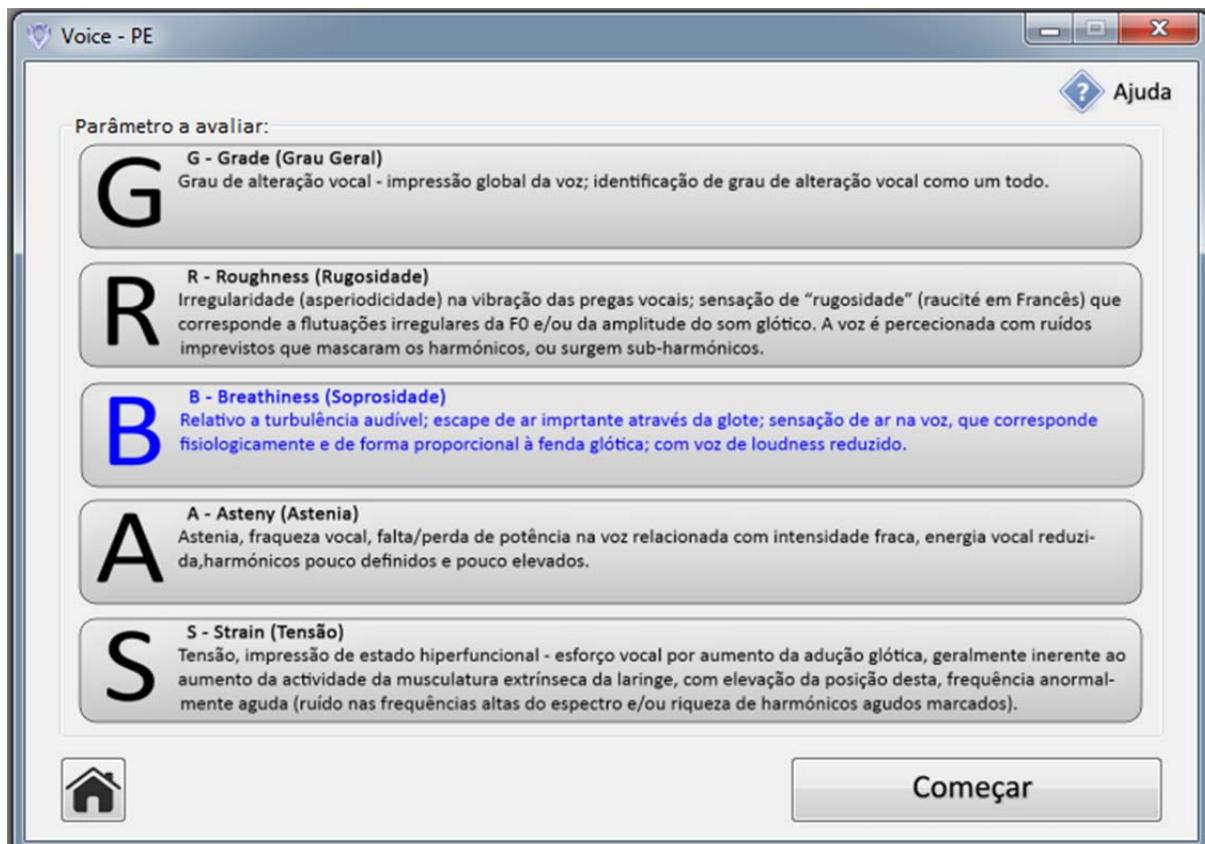


Figura 22: Janela de pré-seleção do parâmetro áudio-percetivo da escala GRBAS a treinar

Inicia-se o treino numa janela que apresenta grupos de 10 vozes, selecionadas pelo sistema aleatoriamente. O utilizador pode escutar cada uma das vozes (clicando sobre o ícone com o formato das pregas vocais), as vezes que necessitar, e organizá-las por comparação entre elas, arrastando o ícone representativo dessa voz, verticalmente (*Cf. Figura 23*). À esquerda da janela está uma régua que assinala os extremos da classificação – qualidade vocal normal/*normal voice quality versus severa/severe* (Yamasaki et al., 2007).

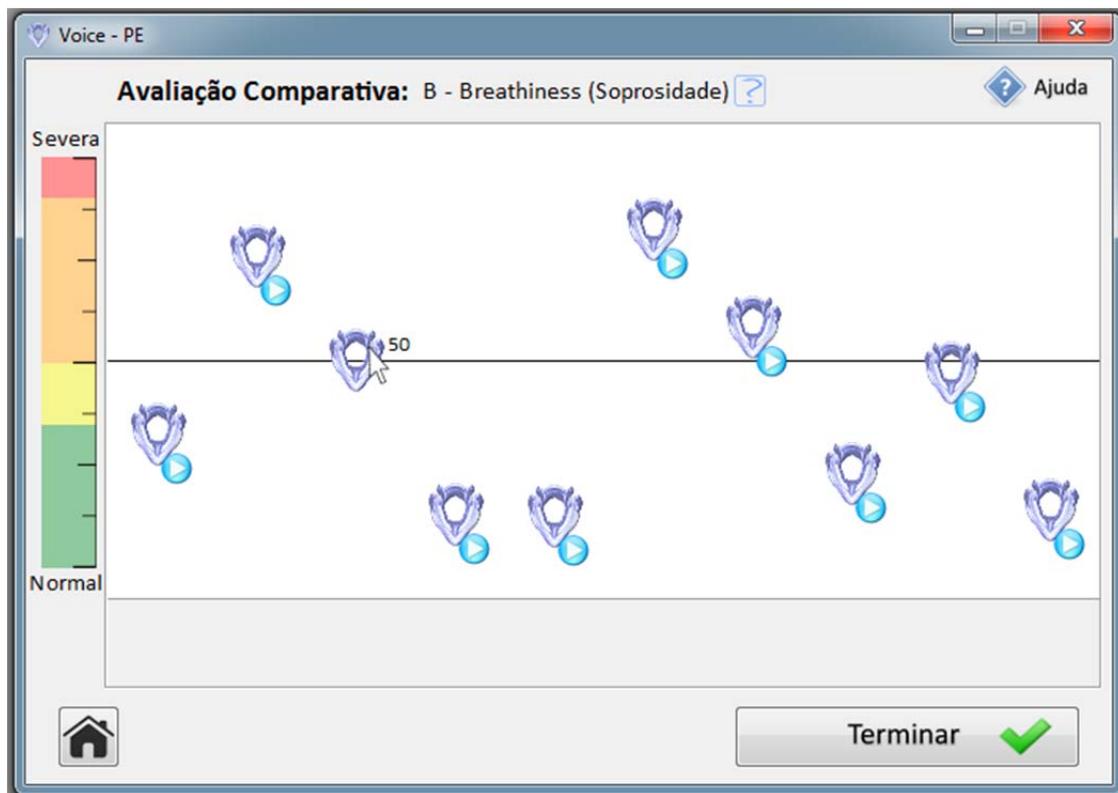


Figura 23: Janela da avaliação comparativa, do parâmetro áudio-percetivo B(Soprosidade)

Todas as vozes têm que ser alvo de alocação na janela classificadora, portanto, fora da barra de apresentação inicial, para finalizar a avaliação e ser criado automaticamente um relatório semelhante ao apresentado na Figura 21. Neste apresenta-se: a cotação individual de cada voz, para cada um dos parâmetros áudio-percetivos (GRBAS), comparando-a com a média dos juízes e com os pontos de corte entre os distintos graus de perturbação vocal (representados através de uma barra de cores).

4.4.2. Estudo sobre práticas profissionais dos terapeutas da fala Portugueses na avaliação da voz humana

- **Data:** 21 e 22 de agosto de 2009

Comunicação Oral – “As práticas profissionais dos Terapeutas da Fala Portugueses no Diagnóstico e Intervenção Terapêutica das Perturbações Vocais”, *em coautoria*, inserido no *4º Simpósio Luso-Brasileiro de Terapia da Fala e 1º Simpósio Luso-Brasileiro de Saúde*, São Paulo.

Artigo Científico: Ricardo Sousa, Susana Vaz Freitas, Aníbal Ferreira – “A Avaliação Acústica da Voz nas Práticas Profissionais dos Terapeutas da Fala Portugueses”, *in Revista da Sociedade Portuguesa de Otorrinolaringologia e Cirurgia Cérvico-Facial*, vol. 49, nº 1, pp. 11-17, 2011.

4.4.3. Publicações/Estudos/Comunicações Orais sobre medidas acústicas e áudio-percetivas

4.4.3.1. Comunicação oral e Artigo sobre a medida HNR

- **Data:** 30 de abril de 2008

Comunicação Oral – “A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa”, *inserida nas I Jornadas sobre Tecnologia e Saúde*, Guarda.

Paper em Ata de Congresso: José Lopes, Susana Freitas, Ricardo Sousa, Joaquim Matos, Filipe Abreu, Aníbal Ferreira – “A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa”, *in I Jornadas sobre Tecnologia e Saúde*.

4.4.3.2. Comunicação oral sobre avaliação acústica no refluxo faringo-laríngeo

- **Data:** 06 de março de 2009

Palestra – “A análise acústica na avaliação e diagnóstico das perturbações vocais associadas ao RFL”, *inserida na Reunião de Serviço de Otorrinolaringologia e Cirurgia Cérvico-Facial do Centro Hospitalar do Porto – Hospital Geral de Santo António*, Porto.

4.4.3.3. Artigo sobre avaliação acústica na Incompetência Glótica

Artigo Científico: Pedro Pestana, Susana Vaz Freitas, Cecília Almeida – “A eficácia da intervenção em Terapia da Fala na paralisia da corda vocal: avaliação objetiva”, *in Revista da Sociedade Portuguesa de Otorrinolaringologia e Cirurgia Cérvico-Facial*, vol. 50, nº 3, pp. 215-220, 2012.

4.4.3.4. Artigo sobre Incompetência Glótica

Artigo Científico: Susana Vaz Freitas, Pedro Pestana – “Incompetência Glótica: Protocolo de Intervenção Baseado na Evidência”, *in Revista Científica da Associação Portuguesa de Terapeutas da Fala* (Submetido).

4.4.3.5. Artigo sobre Disfonia Infantil

Artigo Científico: Susana Vaz Freitas, Pedro Pestana – “Intervenção na Disfonia Infantil: Um desafio de palmo e meio”, *in Revista da Sociedade Brasileira de Fonoaudiologia* (Artigo em finalização, antes da submissão).

4.4.3.6. Comunicações orais com referência à avaliação acústica e áudio-percetiva

- **Data:** 16 de abril de 2009

Palestra/Moderadora – “Expectativas de Intervenção em Voz”, inserida nas Comemorações dia Mundial da Voz – Universidade Fernando Pessoa, Porto.

- **Data:** 21 e 22 de agosto de 2009

Comunicação Oral – “Correlação de Alguns Parâmetros de Avaliação Psicoacústica com Parâmetros de Avaliação Perceptual da Voz”, inserido no 4º Simpósio Luso-Brasileiro de Terapia da Fala e 1º Simpósio Luso-Brasileiro de Saúde, São Paulo.

- **Data:** 25 de maio de 2010

Comunicação Oral – “A importância de um laboratório de voz”, inserida no 1º Curso Teórico-Prático da Voz, Coimbra.

- **Data:** 16 de abril de 2011

Comunicação Oral – “Estudos para a Inovação na Prática Clínica”, inserido no II Congresso Anual de Saúde da Universidade de Aveiro, Aveiro.

- **Data:** 16 de abril de 2011

Comunicação Oral – “A Avaliação Acústica e Percetiva da Voz”, inserido nas Comemorações do Dia Mundial da Voz – Pensar a Voz, Porto.

4.4.3.7. Comunicação oral e Artigo sobre o Vibrato no Fado de Coimbra

- **Data:** 26 de Novembro de 2012

Comunicação Oral – “Vibrato Analysis in Fado de Coimbra”, inserida no 14º Encontro da APEA, Associação Portuguesa de Engenharia de Áudio, Porto.

Paper em Ata de Congresso: Cláudia Fier, Ricardo Sousa, Susana Freitas, Aníbal Ferreira – “Vibrato Analysis in Fado de Coimbra”, inserida no 14º Encontro da APEA.

4.4.4. Artigos propostos para publicação decorrentes da dissertação

4.4.4.1. Susana Vaz Freitas, Pedro Pestana, Vítor Almeida, Aníbal Ferreira – “Acoustic Analisys of Voice Signal: comparison of four aplications software”, in *Biomedical Signal Processing and Control* (Artigo em finalização, antes da submissão).

4.4.4.2. Susana Vaz Freitas, Pedro Pestana, Vítor Almeida, Aníbal Ferreira – “Audio-perceptual evaluation of Portuguese voice disorders – an inter- and intrajudge reliability study”, in *Journal of Voice (In Press)*.

4.4.4.3. Susana Vaz Freitas, Pedro Pestana, Vítor Almeida, Aníbal Ferreira – “Integrating Voice Evaluation: correlation between acoustic and audio perceptual measures”, in *Folia Phoniatrica et Logopaedica*. (Submetido).

4.4.4.4. Susana Vaz Freitas, Vítor Almeida, Aníbal Ferreira – “Software de treino e teste em avaliação áudio percetiva: o Voice-PE (Voice-Perceptual Evaluation)”, in *Revista da Sociedade Portuguesa de Otorrinolaringologia e Cirurgia Cérvico-Facial* (Submetido).

Capítulo 5 – CONCLUSÃO E TRABALHOS FUTUROS

O conhecimento e a adoção na prática dos modelos teóricos que sustentam as representações e medidas percetivas e acústicas da fala e da voz – com vista ao estabelecimento de correspondências entre a percepção e restantes níveis de descrição do sinal sonoro – são de extrema importância, e foram referidos ao longo de toda a dissertação.

Conclui-se que o estudo do sinal da fala ganhou um novo impulso com a possibilidade de concordância entre os seus componentes acústicos e percetivos. A incorporação de recursos instrumentais, especialmente a análise acústica, representou uma nova perspetiva na compreensão dos ajustes subjacentes à produção da fala e da voz, com e sem alterações, o que se repercutiu no processo terapêutico – desde o diagnóstico até à intervenção propriamente dita do terapeuta da fala. No estudo desenvolvido o valor preditivo das medidas acústicas face aos parâmetros áudio-percetivos assumiu magnitudes desde um nível fraco ($R^2_a=0,17$) a bom ($R^2_a=0,71$).

O aprofundar destes conhecimentos permite uma multiplicidade de atuações terapêuticas, algumas ainda pouco exploradas pelo terapeuta da fala em Portugal. O mais comum é este fazer uso da análise acústica para a avaliação e tratamento das alterações da fala e da voz, contudo, pode-se ainda prever a sua funcionalidade na:

- Assessoria de comunicação e expressividade da fala (tanto em casos de voz falada como cantada);
- Construção e implementação de recursos tecnológicos aplicados às perturbações da comunicação humana (por exemplo, com sistemas de síntese e reconhecimento da fala de Sistemas de Comunicação Aumentativa e Alternativa);
- Construção e implementação de interfaces de comunicação (por exemplo, sistemas de comunicação por telefone, comunicação virtual, localização por satélite, sistemas de autoatendimento);
- Área do reconhecimento de falantes, no âmbito da Medicina Forense.

As considerações finais sobre a temática da avaliação da qualidade vocal reforçam a necessidade de que esta seja uma tarefa multidimensional (Hirano, 1989; Uloza et al., 2011). Deve incluir a avaliação áudio-percetiva (na maior parte das vezes é neste domínio que se manifesta de forma mais evidente uma perturbação da voz) e a análise acústica do sinal sonoro (com diversas medidas e análises combinadas), entre outras, já que os efeitos das mudanças vocais (i.e. os resultados terapêuticos) não se darão necessariamente em todos os aspectos avaliados, nem de forma semelhante. Espera-se que os pacientes não atinjam uma normalidade em todos os domínios da voz, assim como também é expectável uma não melhoria uniforme para todos os parâmetros inicialmente avaliados. Registe-se que as classificações médias dos peritos (CCIC) que colaboraram nos trabalhos da dissertação, apresentaram valores dos intervalos de confiança a 95% das correlações acima de 0.9 para os parâmetros G(Grau Geral), R(Rugosidade), e B(Soprosidade). Os parâmetros A(Astenia) e S(Tensão) tiveram uma consistência inferior.

Em termos de perspetivas de atuação e desenvolvimento futuro, conclui-se que ainda existe um amplo campo de investigação nesta área. São úteis estudos que visem confirmar resultados e desenvolver ferramentas de uso clínico, que consigam contornar as desvantagens dos métodos – acústicos e áudio-percetivos – existentes, propiciando aos pesquisadores, profissionais da saúde e aos próprios falantes, dados mais precisos, obtidos de forma expedita e com informações confiáveis.

A confirmação dos dados deve partir de mais estudos, com cuidados de contrução metodológica que acautelem, nomeadamente, a reformulação da descrição dos parâmetros áudio-percetivos, sabendo que são muito difíceis de explicar, ouvir e, consequentemente, de classificar por parte – até mesmo – de peritos experientes na área da avaliação e intervenção vocal e usuários frequentes de escalas áudio-percetivas. Depois, é necessário estudar novas abordagens técnicas de avaliação da voz com o método acústico, nomeadamente com recurso a modelos que considerem a não-linearidade, e possibilitem a análise de fala espontânea. Ou, e no seguimento dos trabalhos da dissertação, o estabelecimento de medidas de quantificação/classificação da qualidade vocal que conjuguem os parâmetros áudio-percetivos e os valores da análise acústica. Esta componente da dissertação também deve continuar a ser investigada, com o intuito de

detetar medidas acústicas com potencial de classificação da qualidade vocal, em especial aquela apresentada pelos casos que surgem diariamente na prática clínica.

Em suma, e embora se reconheça que os objetivos da dissertação foram ambiciosos, facultarão a resposta a algumas dúvidas da doutoranda e servirão como material de suporte pedagógico para os seus alunos e colegas de trabalho, preveem-se desde já algumas perguntas possíveis para outros estudos futuros, alguns suportados pela aplicação do *software* Voice-PE:

1. Será que testar os mesmos avaliadores com as mesmas amostras de voz, após mais tempo entre sessões/consultas, implica uma alteração na consistência intra-avaliadores? Ou seja, será que as referências internas de avaliação percetiva se mantêm no tempo?
2. Será que avaliar as mesmas amostras de voz, mas por avaliadores provenientes de outras formações profissionais de base, implica uma alteração na consistência intra-avaliadores? E na consistência inter-avaliadores, ou seja, será que as referências internas de avaliação percetiva são diferentes para professores de canto, otorrinolaringologistas, foneticistas/linguistas, engenheiros de sinal, ou outros profissionais?
3. Será que o recurso às “âncoras” auditivas/visuais ao longo do teste de classificação percetiva varia? Isto é, será que os avaliadores usam mais estas pistas no início/meio/final da sessão? Ou mais com determinados parâmetros percetivos? Ou para algum tipo específico de grau de severidade?
4. Será que usar “âncoras” auditivas com vozes naturais tem um impacto distinto do uso de pistas auditivas com amostras de vozes sintetizadas?
5. Será que o género influencia a avaliação, quer: em termos de vozes alteradas do treino (estímulos-“âncora”)? Em termos de vozes alteradas de teste para análise e classificação? Em termos de avaliadores? Ou seja, os avaliadores serão mais permissivos de acordo com seu próprio género? Ou será a voz considerada mais ou menos alterada consoante o sexo do seu falante?

Ao chegar ao final da dissertação, para além de deixar um material de treino e teste em termos de avaliação áudio-percetiva (*Voice-PE – Voice – Perceptual Evaluation*), para atuais

e futuros terapeutas da fala e outros utilizadores deste modo de classificação da voz, prevê-se o seu uso, a curto prazo, na determinação de algumas respostas para as questões supraenumeradas.

Também se definiram fórmulas de quantificação/descrição da qualidade vocal que conjugaram os parâmetros da escala GRBAS e os valores das medidas de análise acústica obtidos para os diferentes tipos de *software* analisados no estudo levado a cabo. Este trabalho de investigação acerca da correspondência entre medidas acústicas e parâmetros áudio-percetivos e a comparação entre programas de *software* distintos deve continuar a ser aprofundado, embora reconhecendo que o estabelecimento de relações diretas e únicas entre a componente de avaliação ácustica e a áudio-percetiva é muito complexa, e quiçá impossível.

Ao terapeuta da fala cabe-lhe apreender e dominar estas ferramentas, no sentido de melhor conseguir traçar estimativas do processo de produção vocal, sem a necessidade de técnicas invasivas. Com o devido suporte teórico das Ciências da Fala e preparação prática, nomeadamente no campo da Fonética Acústica, este é o profissional que congrega facilmente os aspectos da produção e percepção do sinal, beneficiando a sua atuação clínica e podendo, inclusive, fazer incursões noutros campos inovadores ao atuar com as mais variadas manifestações de fala e voz.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Abitbol, J.; Abitbol, P.; Abitbol, B. 1999. Sex Hormones and the Female Voice. *Journal of Voice*. 13: 424-46.
2. Abitbol, Jean; Sataloff, Pr Robert Thayer (Org.). 2002. *7th International Workshop on Laser Voice Surgery and Voice Care*. Paris: Livro de Resumos. 19-20 Abril.
3. Akhtar, S.; Wood, G.; Rubin, J.S.; O`Flynn, P.E.; Ratcliffe, P. 1999. Effect of caffeine on the vocal folds: a pilot study. *Journal of Laryngology and Otology*. 113: 341-345.
4. Akpinar, M.E.; Kocak, I.; Gurpinar, B. and Esen, H.E. 2011. Effects of Soft Palate Implants on Acoustic Characteristics of Voice and Articulation. *Journal of Voice*. 25(3): 381-86.
5. Albertini, G.; Giaquinto, S.; Mignamo, M. 2009. Spectral analysis of the human voice: a potentially useful tool in rehabilitation. *European Journal of Physical Rehabilitation Medicine*. 45(4): 537-45.
6. Hamdam, A-L.; Deeb, R.; Sibai, A.; Rameh, C.; Rifai, H.; Fayyad, J. 2009. Vocal Characteristics in Children With Attention Deficit Hyperactivity Disorder. *Journal of Voice*. 23(2): 190-194.
7. Amir, O.; Wolf, M. and Amir, N. 2009. A clinical comparison between two acoustic analysis software: MDVP and Praat. *Biomedical Signal Processing and Control*. 4: 202-205.
8. Anders, L.; Hollien, H.; Hurme, P.; Sonninen, A. and Wendler, J. 1988. Perceptual evaluation of hoarseness by several classes of listeners. *Folia Phoniatrica et Logopaedica*. 40: 91-100.
9. Andrade e Silva, M.A., Duprat, A.C. 2004. "Voz Cantada". In *Tratado de Fonoaudiologia*, Ferreira, L.P.; Beffi-Lopes, D.M. and Limongi, S.C.O.. São Paulo: Roca.
10. Andrews, M.L. 1995. *Manual of Voice Treatment, Pediatrics Through Geriatrics*. San Diego: Singular.

11. Araújo, S.A. 1999. *Análise acústica da voz normal*. Dissertação de Mestrado em Otorrinolaringologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo.
12. Araújo, S.A.; Grellet, M.; Pereira, J.C. and Rosa, M.O. 2002. Normatização de medidas acústicas da voz normal. *Revista Brasileira de Otorrinolaringologia*. 68: 540-544.
13. Aronson, A. 1990. *Clinical Voice Disorders*. 3rd Edition. New York: Thieme Medical Publishers, Inc.
14. Askenfelt, A.G. and Hammarberg, B. 1986. Speech Waveform perturbation analysis: a perceptual-acoustic comparison of seven measures. *Journal of Speech and Hearing Research*. 29: 50-64.
15. Awan, S.N. 2001. *The voice diagnostic protocol: A practical guide to the diagnosis of voice disorders*. Gaithersburg, M.D.: Aspen.
16. Awan, S.N. and Lawson, L.L. 2009. The Effect of Anchor Modality on the Reliability of Vocal Severity Ratings. *Journal of Voice*. 23(3): 341-352.
17. Awan, S.N. and Roy, N. 2005. Acoustic prediction of voice type in women with functional dysphonia. *Journal of Voice*. 19: 268-282.
18. Awan, S.N. and Roy, N. 2006. Toward the development of an objective index of dysphonia severity: A four-factor model. *Clinical Linguistics & Phonology*. 20: 35-49.
19. Awan, S.N. and Roy, N. 2009. Outcomes Measurement in Voice Disorders: Application of an Acoustic Index of Dysphonia Severity. *Journal of Speech Language and Hearing Research*. 52: 482-499.
20. Awan, S.; Roy, N. and Jiang, J. 2010. Nonlinear Dynamic Analysis of Disordered Voice: The Correlation Dimension (D2) as an index of dysphonia severity following treatment? *Journal of Voice*. 24(3): 285-293.
21. Baken, R.J. 1990. Irregularity of vocal fold period and amplitude: a first approach to the fractal analysis of voice. *Journal of Voice*. 4(3): 185-197.
22. Baken, R.J. and Orlikoff, R.J. 2000. *Clinical Measurement of Speech and Voice*. 2nd Edition. San Diego: Singular Publishing.
23. Balasubramanian, R.K.; Bhatm J.B.; Srivastava, M. and Eldose, A. 2012. Cepstral Analysis of Sexually Appealing Voice. *Journal of Voice*. 26(4): 412-415.

24. Barros, A.P.B. and Carrara De-Angelis, E.C. 2002. "Análise Acústica da Voz". In *Métodos de avaliação e diagnóstico de laringe e voz*, Deditivis, R.A.; Barros, A.P.B.. São Paulo: Lovise Editores.
25. Bassich, C.J. and Ludlow, C.L. 1986. The use of perceptual methods by new clinicians assessing voice quality. *Journal of Speech and Hearing Disorders*. 51: 125-133.
26. Batalla, F.N., Santos, P.C., Santiago, G.S., González, B.S, and Nieto, C.S. 2004. Evaluación perceptual de la disfonía: correlación con los parámetros acústicos y fiabilidad. *Acta Otorrinolaringológica Española*, 55(6): 282-287.
27. Behlau, M. (Org). 2001. *Voz. O Livro do Especialista*. Rio de Janeiro: Livraria e Editora Revinter, Lda. Volume 1.
28. Behlau M., Madazio G., Feijó D., Pontes P. 2001. Avaliação de voz. In: Behlau. MS. *Voz: o livro do especialista*. Rio de Janeiro: Revinter. Volume I.
29. Behlau, M. 2010. "Técnicas Vocais". In *Tratado de Fonoaudiologia*. 2^a Edição. Fernandes, F.D.M. et al.. São Paulo: Roca.
30. Behlau, M. 2004. Consensus auditory-perceptual evaluation of voice (CAPE-V), ASHA 2003. *Revista da Sociedade Brasileira de Fonoaudiologia*. 9(3):187-9.
31. Behlau, M. and Pontes, P. 1995. *Avaliação e Tratamento das Disfonias*. São Paulo: Editora Lovise.
32. Behram, A. 2005. Common Practices of Voice Therapists in the Evaluation of Patients. *Journal of Voice*. 19(3): 454-469.
33. Bele, I.V. 2005. Reliability in Perceptual Analysis of Voice Quality. *Journal of Voice*. 19(4): 555-573.
34. Benninger, M.S.; Jacobson, B.H. and Johnson, A.F. (Edit.). 1994. *Vocal Arts Medicine: The Care and Prevention of Professional Voice Disorders*. New York: Thieme Medical Publishers, Inc.
35. Bielamowicz, S.; Kreiman, J.; Gerratt, B.R.; Dauer, M.S. and Berke, G.S. 1996. Comparison of voice analysis systems for perturbation measurement. *Journal of Voice*. 39: 126-34.
36. Blaj, S.; Madureira, S. and Camargo, Z. 2007. Vocal quality in asthmatic individuals with and without paradoxical vocal fold dysfunction: perceptual, acoustic and physiologic correlates. In *2nd International Symposium of IALP*

- (International Association of Logopedics and Phoniatrics). Abstract Book 2nd International Symposium of IALMP. São Paulo: IALP, V1: 126-127.
37. Bocklet, T.; Toy, H.; Noth, E. et al.. 2009. Automatic evaluation of tracheoesophageal substitute voice: sustained vowel versus standard text. *Folia Phoniatrica et Logopaedica*. 61: 112-116.
 38. Bonatto, M.T.R.L.; Andrada e Silva, M.A. and Costa, O.H. 2004. A relação entre respiração e sistema sensorio-motor oral em crianças disfónicas. *Revista CEFAC*. 6(1): 58-66.
 39. Bhuta, T.; Patrick, L. and Garnett, J.D. 2004. Perceptual Evaluation of voice quality and its correlation with acoustic measurements. *Journal of Voice*. 18: 299-304.
 40. Bistrisky, Y. and Frank, Y. 1981. Efficacy of voice and speech training of prospective elementary school teachers. *Israeli Journal of Speech and Hearing*. 10: 16-32.
 41. Boone, D.R. and McFarlane, S.C. 2000. *The Voice and Voice Therapy*. 6th Edition. Boston: Allyn & Bacon.
 42. Boucher, V.J. 2008. Acoustic Correlates of Fatigue in Laryngeal Muscles: Findings for a Criterion-Based Prevention of Acquired Voice Pathologies. *Journal of Speech Language and Hearing Research*. 51: 1161-1170.
 43. Brasolotto, A.G. and Rehder, M.I. 2011. "Diagnóstico Vocal Fonoaudiológico." In *Disfonia e Disfagia*, Rehder, M.I. and Branco, A.. Rio de Janeiro: Livraria e Editora Revinter, Ltda.
 44. Brindle, B.R. and Morris, H.L. 1979. Prevalence of voice quality deviations in the normal adult populations. *Journal of Communication Disorders*. 12: 439-45.
 45. Broaddus-Lawrence, P.L. ; Treole, K. ; McCabe, R.B. ; Allen, R.L. and Toppin, L. 2000. The Effects of Preventive Vocal Hygiene Education on the Vocal Hygiene Habits and Perceptual Vocal Characteristics of Training Singers. *Journal of Voice*. 14(1): 58-71.
 46. Brochmann, M.; Storck, C.; Carding, P.N. and Drinnan, M.J. 2008. Voice loudness and gender effects on jitter, shimmer in healthy adults. *Journal of Speech Language and Hearing Research*. 51: 1152-1160.

47. Brown, W.S. Jr.; Morris, R.J. and Michel, J.F. 1990. Vocal jitter and fundamental frequency characteristics in aged, female professional singers. *Journal of Voice*. 4: 135-141.
48. Buder, E.H. 2000. "Acoustic Analysis of Voice Quality: A Tabulation of Algorithms 1902-1990." In *Voice Quality Measurements*, Kent, R.D. and Ball, M.J.. San Diego: Singular Publishing Inc..
49. Bunton, K. ; Kent, R.D. ; Duffy, J.R. ; Rosenbeck, J.C. and Kent, J.F. 2007. Listener Agreement for Auditory-Perceptual Ratings of Dysarthria. *Journal of Speech Language and Hearing Research*. 50: 1481-1495.
50. Burnett, T.A. Senner, J.E. and Larson, C.R. 1997. Voice F0 responses to pitch-shifted auditory feedback : a preliminary study. *Journal of Voice*. 11(2) : 202-211.
51. Calas M.; Verhulst J.; Lecoq M.; Dalleas B. and Seilhean M. 1989. La Phatologie Vocale chez L'Enseignant. *Revue de Laryngologie Otologie Rhinologie*. 110: 397-406.
52. Camargo, Z.A. 1996. Parâmetros Vocais e Configurações Laríngeas na Fonação de Indivíduos Submetidos às Laringectomias Parciais Verticais. Dissertação de Mestrado. Pontifícia Universidade Católica de São Paulo.
53. Camargo, Z.A. 2000. "Avaliação Objectiva da Voz". In *A atuação fonoaudiológica no cancer de cabeça e pescoço*. Carrara-de-Angelis, E. et al.. São Paulo: Lovise.
54. Camargo, Z.A. ; Madureira, S. and Tsuji, D.H. 2003. "Analysis of dysphonic voices based on the interpretation of acoustic, physiological and perceptual data". In *6th International Seminar on Speech Production*. Sydney.
55. Camargo, Z.A. and Madureira, S. 2009. Dimensões Perceptivas das Alterações de Qualidade Vocal e Suas Correlações aos Planos da Acústica e da Fisiologia. *D.E.L.T.A.* 25(2): 285-317.
56. Camargo, Z. and Madureira, S. 2010. Análise Acústica: Aplicações na Fonoaudiologia. In *Tratado de Fonoaudiologia*. 2^a Edição, Fernandes, F.D.M. et al.. São Paulo: Roca.
57. Campisi, O.; Tewfik, T.L.; Pelland-Blais, E.; Hussein, M. and Sadegui, N. 2000. Multidimensional Voice Program analysis in children with vocal cord nodules. *Internacional Journal of Otolaryngology*. 29(5): 302-8.

58. Campisi, O.; Tewfik, T.L.; Manoukian, J.J.; Schloss, M.D.; Pelland-Blais, E. and Sadegui, N. 2002. Computer-assisted voice analysis; Establishing a pediatric database. *Archives of otolaryngology-head & neck surgery*. 128(2): 156-160.
59. Cannito, M.P.; Woodson, G.E.; Murry, T. and Bender, D. 2004. Perceptual Analyses of Spasmodic Dysphonia Before and After Treatment. *Archives of otolaryngology-head and neck surgery*. 130: 1393-1399.
60. Carding, P.; Carlson, E.; Epstein, R.; Mathieson, L. and Shewell, C. 2000. Formal perceptual evaluation of voice quality in United Kingdom. *Logopaedic Phoniatic Vocology*. 25: 133-138.
61. Carding, P. and Horsley, I. 1992. An evaluation study of voice therapy in non-organic dysphonia. *European Journal of Disordered Communication*. 27(2): 137-58.
62. Carding, P.; Steen, I.N.; Webb, A.; Mackenzie, K.; Deary, I.J. and Wilson, J.A. 2004. The reliability and sensitivity to change of acoustic measures of voice quality. *Clinical Otolaryngology and Allied Sciences*. 29(5): 538-544.
63. Carmo, R.D.; Camargo, Z. and Nemr, K. 2006. Relação entre Qualidade de Vida e auto-perceção da Qualidade Vocal de pacientes Laringectomizados Totais: Estudo Piloto. *Revista CEFAC*. 8(4): 518-28.
64. Carson, C.; Ingrisano, D. and Eggleston, K. 2003. The effect of noise on computer-aided measures of voice: a comparison of CSpeechSP and the Multi-Dimensional Voice Program software using the CSL 4300B module and Multi-Speech for Windows. *Journal of Voice*. 17: 12-20.
65. Chan, K.M.-K. 2011. Auditory-Perceptual Voice Evaluation: A Practical Approach. In *Handbook of Voice Assessments*, Ma, E. and Yiu, E.. San Diego: Singular Publishing Inc..
66. Chan, R.W.K. 1994. Does the voice improve with vocal hygiene education? A study of some instrumental voice measures in a group of kindergarten teachers. *Journal of Voice*. 8: 279-91.
67. Chan, K.M.K. and Yiu, E.M-L. 2002. The effect of anchors and training on the reliability of perceptual voice evaluation. *Journal of Speech and Hearing Research*. 45: 111-126.

68. Chan, K.M.K. and Yiu, E.M-L. 2006. A Comparison of Two Perceptual Voice Evaluation Training Programs for Naive Listeners. *Journal of Voice*. 20(2): 229-241.
69. Chen, S.H.; Chang, S.Y.; Fue, S.W. and Chang, P. 1996. A Survey on Voice Disorders on Junior High School Teachers in Taipei. Apresentação no XXth Congress of the International Association of Logopedics and Phoniatrics. Tokyo.
70. Childers, D.G.; Skynner, D.P. and Kemerait, R. 1977. The Cepstrum: a guide to processing. *Proceedings of the IEEE*. Annals of the IEEE. 65(10): 1428-1442.
71. Choi, S.H.; Zhang, Y.; Jiang, J.J.; Bless, D.M.; Welham, N.V. 2012a. Nonlinear Dynamic-Based Analysis of Severe Dysphonia in Patients With Vocal Fold Scar and Sulcus Vocalis. *Journal of Voice*. 26(5): 566-76.
72. Choi, S.H.; Lee, J. Y.; Sprecher, A.J. and Jiang, J.J. 2012b. The Effect of Segment Selection on Acoustic Analysis. *Journal of Voice*. 26(1): 1-17.
73. Colton, R.H. and Casper, J. 1996. *Compreendendo os problemas da voz: uma perspectiva fisiológica ao diagnóstico e ao tratamento*. Porto Alegre: Editora Artes Médicas.
74. Comins, R. 1995. Vocal Tuition for professional voice users: a tutor's account. *Journal of the British Voice Association*. 4: 32-43.
75. Cornwell, P.L.; Murdoch, B.E.; Ward, E.C. and Kellie, S. 2004. Acoustic investigation of vocal quality following treatment for childhood cerebellar tumor. *Folia Phoniatrica et Logopaedica*. 56(2): 93-107.
76. Costa, Henrique O.; Duprat, André; Eckley. Cláudia and Silva, Marta A.A. e. 2000. Caracterização do Profissional da Voz para o Laringologista. *Revista Brasileira de Otorrinolaringologia*. 66 (2) Parte 1: 129-134.
77. Cox, N.B.; Ito, M.R. and Morrison, M.D. 1989. Data labeling and sampling effects in harmonic-to-noise ratios. *Journal of the Acoustic Society of America*. 85(5): 2165-2178.
78. Cummings, L. 2008. *Clinical Linguistics*. Edinburgh: Edinburgh University Press.
79. Dajer, M.E. 2010. Análise de Sinais de Voz por Padrões Visuais de Dinâmica Vocal. Dissertação de Doutoramento. Escola de Engenharia de São Carlos da Universidade de São Paulo.

80. Davis, S.B. 1979. "Acoustic characteristics of normal and pathological voices." In *Speech and language: advances in basic research and practice*, Lass, N.J.. New York: Academic Press.
81. De Bodt, M.S.; Van de Heyning, P.H.; Wuyts, F.L. and Lambrechts, L. 1996. The perceptual evaluation of voice disorders. *Acta Otorhinolaryngologie Belgique*. 50: 283-91.
82. De Bodt, M.S.; Wuyts, F.L.; Van de Heyning, P.H. and Croux, C. 1997. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual ratings of voice quality. *Journal of Voice*. 1: 74-80.
83. Dedivitis, R.A.; Barros, A.P.B.; Queija, D.S.; Alexandre, J.C.; Rezende, W.T.; Corazza, V.R., et al. 2004. Interobserver perceptual analysis of smokers' voice. *Clinics Otolaryngology Allied Sciences*. 29: 124-7.
84. Dedivitis, R.A.; Barros, A.P.B.; Queija, D.S.; Pfuetzenreiter Jr., E.G. and Bohn, N.P. 2008. Achados percetivo-auditivos e acústicos em pacientes submetidos à laringectomia fronto-lateral. *Revista Brasileira de Cirurgia de Cabeça e Pescoço*. 37(3): 163-165.
85. Dejonckere, P.H. 1995. Principal components in voice pathology. *Journal of Voice*. 4: 96-105.
86. Dejonckere, P.H. 2007. "Critères acoustiques de fluence pour l`evaluation des dysphonies spasmodiques." In *Voix parlée et chantée*, Klein-Dallant, C. (Edit.). Paris: Editions Ortho.
87. Dejonckere, P.H. 2010. "Assessment of Voice and Respiratory Function." In *Surgery of Larynx and Trachea*, Remacle, M. and Eckel, H.E. (Eds.). Berlin: Springer-Verlag.
88. Dejonckere, P.H. (Edit). 2001. *Occupational Voice – care and cure*. San Diego: Singular Publishing Group.
89. Dejonckere, P.H.; Bradley, P.; Clemente, P. et al. 2001. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Otorhinolaryngology*. 258: 77-82.
90. Dejonckere, P.H.; Giordano, A.; Schoentgen, J.; Frag, S.; Bocchi, L. and Manfredi, C. 2012. To what degree of voice perturbation are jitter measurements valid? A

- novel approach with synthesized vowels and visuo-perceptual pattern recognition. *Computer Methods and Programs in Biomedicine Journal*. 7: 37-42.
91. Dejonckere, P.H. and Lebacq, J. 1996. Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology. *ORL Journal for oto-rhino-laryngology and its related specialities*. 58(6): 326-32.
92. Dejonckere, P.H.; Obbens, C.; Demoor, G.M. and Wienecke, G.H. 1993. Perceptual evaluation of dysphonia: Reliability and relevance. *Folia Phoniatrica et Logopaedica*. 45: 76-83.
93. Dejonckere, P.H.; Remacle, M.; Fresnel-Elbaz, E.; Woisnard, V.; Crevier-Buchman, L. and Millet, B. 1996. Differentiated perceptual evaluation of pathological voice quality reliability and correlations with acoustic measurements. *Revue de Laryngologie Otologie Rhinologie*. 117: 219-224.
94. De Krom, G. 1993. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*. 36: 254-266.
95. De Krom, G. 1994. Consistency and Reliability of Voice Quality Ratings for Different Types os Speech Fragments. *Journal of Speech and Hearing Research*. 37: 985-1000.
96. Deliyski, D. 1993. Acoustic model and evaluation of pathological voice production. *Proceedings of the 3rd Conference on Speech Communication and Technology*. 1969-1972.
97. Deliyski, D.; Shaw, H. and Evans, M. 2005. Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logopaedic Phoniatic Vocology*. 30: 55-62.
98. Deliyski, D.; Shaw, H.; Evans, M. and Vesselinov, R. 2006. Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatrica et Logopaedica*. 58:274-288.
99. Diehl, R. 2000. Searching for an auditory description of vowel categories. *Phonetica*. 57: 267-274.
100. Eadie, T.L. and Doyle, F. 2005. Classification of Dysphonic Voice: Acoustic and Auditory-Perceptual Measures. *Journal of Voice*. 19: 1-14.

101. Eadie, T.L. and Baylor, C.R. 2006. The Effect of Perceptual Training on Inexperienced Listeners` Judgments of Dysphonic Voice. *Journal of Voice*. 20(4): 527-544.
102. Eadie, T.L.; Boven, L.V.; Stubbs, K. and Giannini, E. 2010a. The Effect of Musical Background on Judgments of Dysphonia. *Journal of Voice*. 24(1): 93-101.
103. Eadie, T.L.; Kapsner, M.; Rosenzweig, J.; Waugh, P.; Hillel, A.; Merati, A. 2010b. The role of experience on judgments of dysphonia. *Journal of Voice*. 24(5): 564-573.
104. Eskenazi, L.; Childers, D.G. and Hicks, D.M. 1990. Acoustics correlates of vocal quality. *Journal of Speech and Hearing Research*. 33: 298-30611.
105. Espanhol, D.R. 2004. Determinadores de Pitch. Dissertação de Mestrado. Departamento de Engenharia Eléctrica da Universidade de São Paulo.
106. Fant, G. 1970. *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*. The Hague: Mouton.
107. Ferrand, C.T. 1995. Effects of practice with and without knowledge of results on jitter and shimmer levels in normally speaking women. *Journal of Voice*. 9(4): 419-423.
108. Feijoo, S. and Hernández, C. 1990. Short-term stability measures for the evaluation of vocal quality. *Journal of Speech and Hearing Research*. 33: 324-334.
109. Finizia, C.; Dotevall, H.; Lundström, E. and Lindström, J. 1999. Acoustic and Perceptual Evaluation of Voice and Speech Quality. *Archives of otolaryngology-head & neck surgery*. 125: 157-163.
110. Fioravanti, C. 2003. Prova de Resistência. *Pesquisa FAPESP. Ciência e Tecnologia no Brasil*. 89: 32-37.
111. Fourcin, A. 2000. Patterns, Structure and Measurement in Voice Pathology. *IEEE International Conference on Acoustics, Speech and Signal Processing*. ICASSP'00.
112. Fowler, C.A.; Rubin, P.E.; Remez, R.R. and Turvey, M.T. 1980. "Implications for speech production of a general theory of action". In *Language Production. Vol. I: Speech and Talk*, Butterworth (Ed.). 373-420. New-York: Academic Press.
113. Franco, R.A. and Andrus, J.G. 2009. Aerodynamic and acoustic characteristics of voice before and after adduction arytenopexy and medialization laryngoplasty

- with Gore-Tex in patients with unilateral vocal fold immobility. *Journal of Voice*. 23: 261-267.
114. Freeman, M. and Fawcus, M. 2004. *Distúrbios da Voz e seu Tratamento*, 3^a Edição, São Paulo: Livraria Santos Editora.
115. Freitas, D. 2003. Complicações extra-esofágicas da DRGE. *Endonews*. 6: 21.
116. Fritzell, B. 1996. Voice Disorders and Occupations. *Logopedic Phoniatric Vocology*. 21:7-12.
117. Fröhlich, M.; Michaelis, D.; Strube, H.W. and Kruse, E. 1997. "Acoustic voice quality description: case studies for different regions of the hoarseness diagram". In *Advances in quantitative laryngoscopy*, 2nd Edition, Wittenberg T., Mergell P., Tigges M. and Eysholdt U. (Eds.). 143-150. Berlin: Erlangen.
118. Fröhlich, M.; Michaelis, D.; Strube, H.W. and Kruse, E. 2000. Acoustic voice analysis by means of the hoarseness diagram. *Journal of Speech and Hearing Research*. 43: 706-720.
119. Fukazawa, T.; El-Assuoaty, A. and Honjo, I. 1988. A new index for evaluation of the turbulent noise in pathological voice. *Journal of the Acoustic Society of America*. 83: 1189-1193.
120. Garrett, C.G. 1999. Hoarseness. *Medice-Clinics of North America*. 83: 115-123.
121. Garret, K.L. and Healey, E.C. 1987. An acoustic analysis of fluctuations in the voices of normal adult speakers across three times of day. *Journal of the Acoustic Society of America*. 82: 58-62.
122. Gelzinis, A.; Verikas, A. and Bacauskiene, M. 2008. Automated speech analysis applied to laryngeal disease categorization. *Computer Methods and Programs in Biomedicine Journal*. 91: 36-47.
123. Gerratt, B.R. and Kreiman, J. 2000. Theoretical and methodological development in study of pathological voice quality. *Journal of Phonetics*. 28: 335-42.
124. Gerratt, B.R.; Kreiman, J.; Antonanzaz-Barroso, N. and Berke, G.S. 1993. Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research*. 36: 14-20.

125. Ghoraani, B.; Umapathy, K.; Sugavaneswaran, L.; Krishnan, S. 2012. Pathological speech signal analysis using time-frequency approaches. *Critical Review in Biomedical Engineering*. 40(1): 63-95.
126. Giovanni, A.; Ouaknine, M. and Triglia, J.M. 1999. Determination of largest Lyapunov exponents of vocal signal: a meta-analysis. *Journal of Voice*. 13: 341-354.
127. Giovanni, A.; Robert, D.; Estublier, N.; Teston, B.; Zanaret, M. and Cannoni, M. 1996. Objective evaluation of dysphonia: Preliminary results of a device allowing simultaneous acoustic and aerodynamic measures. *Folia Phoniatrica et Logopaedica*. 48: 175-185.
128. Glaze, L.E.; Bless, D.M. and Susser, R.D. 1990. Acoustic analysis of vowel and loudness differences in children's voice. *Journal of Voice*. 4(1): 37-44.
129. Gliklich, R.E.; Glovsky, R.M. and Montgomery, W.W. 1999. Validation of a Voice Outcome Survey for Unilateral Vocal Cord Paralysis. *Otolaryngology-Head and neck surgery*. 120: 153-8.
130. Gonzalez, J.; Cervera, T. and Miralles, J.L. 2002. Acoustic voice analysis: Reliability of a set of multi-dimensional parameters. *Acta Otorrinolaringológica Española*. 53(4): 256-268.
131. Gotaas, C. and Starr, C.D. 1993. Vocal Fatigue among Teachers. *Folia Phoniatrica et Logopaedica*. 45: 120-9.
132. Gould, J.; Waugh, J.; Carding, P. and Drinnan, M. 2012. A New Voice Rating tool for Clinical Practice. *Journal of Voice*. 26(4): 163-170.
133. Guimarães, I. 2002. An electrolaryngographic study of dysphonic Portuguese speakers. Dissertação de Doutoramento. University of London.
134. Guimarães, I. 2007. *A Ciência e a Arte da Voz Humana*. Alcoitão: ESSA – Escola Superior de Saúde de Alcoitão.
135. Granqvist, S. 2003. The visual sort and rate method for perceptual evaluation in listening tests. *Logopedics Phoniatrics Vocology*. 28: 109-116
136. Hajaiej, Z.; Ouni, K.; Ellouze, N. 2006. Paramétrisation de la parole basée sur une modélisation des filters cochléaires: Application au RAP. Apresentação do XXVIéme Journées d'études sur la Parole. Dinard, França.

137. Hair, J.F.; Anderson, R.E.; Tatham, R.L. and Black, W.C. 1998. *Multivariate Data Analysis*. 5th Edition. New Jersey: Prentice-Hall.
138. Hakkesteegt, M.M.; Brohaar, M.M.P.; Wieringa, M.H. and Feenstra, L. 2008. The Relationship Between Perceptual Evaluation and Objective Multiparametric Evaluation of Dysphonia Severity. *Journal of Voice*. 2: 138-145
139. Halberstam, B. 2004. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL*. 66: 70-73.
140. Hammarberg, B. 1986. Perceptual and acoustic analysis of dysphonia. Dissertação de Doutoramento. Department of Logopedics and Phoniatrics, Huddinge University Hospital, Stockholm.
141. Hammarberg, B. 2000. Voice Research and Clinical Needs. *Folia Phoniatrica et Logopaedica*. 52: 93-102.
142. Hammarberg, B.; Fritzell, B.; Gauffin, J. and Sundberg, J. 1986. Acoustic and perceptual analysis of vocal dysfunction. *Journal of Phonetics*. 14: 533-547.
143. Hammarberg, B.; Gauffin, J. 1995. "Perceptual and acoustics characteristics of quality differences in pathological voices as related to physiological aspects." In *Vocal Fold Physiology*, Fujimura, O. and Hirano, M. 283-303. San Diego: Singular Publishing Group Inc..
144. Harvey, P.L. 1997. The three ages of voice – the young adult patient. *Journal of Voice*. 11: 144-52.
145. Harnisch, W.; Brosch, S.; Schmidt, M. and Hagen, R. 2008. Breathing and voice quality after surgical treatment for bilateral vocal cord paralysis. *Archives of otolaryngology - head & neck surgery*. 134: 278-84.
146. Harrington, J.; Palethorpe, S. And Watson, C.I. 2007. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In *Proceedings of InterSpeech*. Munchen, DE.
147. Heidel S.E. and Torgerson J.K. 1993. Vocal problems among aerobic instructors and aerobic participants. *Journal of Communication Disorders*. 26:179-191.

148. Heman-Ackah, Y.D.; Heuer, R.J.; Michael, D.D. et al. 2003. Cepstral peak prominence: a more reliable measure of dysphonia. *Journal of Speech Language and Hearing Research*. 112: 324-333.
149. Heman-Ackah, Y.D.; Michael, D.D. and Goding, G.S. 2002. The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*. 16: 20-27.
150. Hemler, R.J.B.; Wieneke, G.H. and Dejonckere, P.H. 1997. The Effect of Relative Humidity of Inhaled Air on Acoustic Parameters of Voice in Normal Subjects. *Journal of Voice*. 11(3):295-300.
151. Hess, W. 1983. *Pitch determination of speech signals: Algorithms and devices*. Berlin: Springer-Verlag.
152. Hess, W. 1994. "Pitch determination of speech signals, with special emphasis on time-domain method." In *Proceedings of the Workshop on Acoustic Voice Analysis*, Wong, D. (Ed.). Iowa City: National Center for Voice and Speech.
153. Hillenbrand, J. 1988. Perception of aperiodicities in synthetically generated voices. *Journal of the Acoustic Society of America*. 83: 2361-2371.
154. Hillenbrand, J.; Cleveland, R.A. and Erickson, R.L. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*. 37:769-778.
155. Hillenbrand, J. and Houde, R.A. 1996. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*. 39: 311-321.
156. Hillman, R.; Montgomery, W. and Zeitels, S. 1997. Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Current Opinion Otolaryngology Head Neck Surgery*. 5: 172-175.
157. Hirano, M. 1981. *Clinical examination of voice*. Vienna: Springer-Verlag.
158. Hirano, M. 1989. Objective evaluation of the human voice: clinical aspects. *Folia Phoniatrica et Logopaedica*. 41: 89-144.
159. Hirano, M. and Bless, D.M. 1997. *Exame Videoestroboscópico da laringe*. Porto Alegre: Artes Médicas.
160. Hirano, M.; Hibi, S.; Terasawa, R. and Fujii, M. 1986. Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia. *Journal of Phonetics*. 14: 445-456.

161. Hiroto, I. 1967. Hoarseness: viewpoints of voice physiology. *Japanese Journal of Logopedics & Phoniatrics*. 8: 1-9.
162. Hogikyan, N.D. and Sethuraman, G. 1999. Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice*. 13: 557-569.
163. Hogikyan, N.D.; Wodchis, W.P.; Terrell, J.E.; Bradford, C.R. and Esclamado, R.M. 2000. Voice-related quality of life (V-RQOL) following type I thyroplasty for unilateral vocal fold paralysis. *Journal of Voice*. 14: 378-386.
164. Holmberg, E.; Hillman, R.; Perkell, J. and Gress, C. 1994. Relations between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings. *Journal of Speech and Hearing Research*. 37: 484-495.
165. Holmberg, E.; Doyle, P.; Perkell, J.; Hammarberg, B. and Hillman, R. 2003. Aerodynamic and acoustic voice measurements of patients with vocal nodules: variation in baseline and changes across voice therapy. *Journal of Voice*. 17: 269-282.
166. Horii, Y. 1982. Jitter and shimmer differences among sustained vowel phonations. *Journal of Speech and Hearing Research*. 25: 12-14.
167. Howard, D.M.; Abberton, E.; Fourcin, A. 2012. Disordered voice measurement and auditory analysis. *Speech Communication Journal*. 54(5): 611-621.
168. Hufnagle, J. and Hufnagle, K. 1984. An investigation of the relationship between speaking fundamental frequency and vocal quality improvement. *Journal of Communication Disorders*. 17: 95-100.
169. IALP. 1992. International Association of Logopedics and Phoniatrics (IALP) voice committee discussion of assessment topics. *Journal of Voice*. 6: 194-210.
170. Isshiki, N.; Okamura, H.; Tanabe, M. and Morimoto, M. 1969. Differential diagnosis of hoarseness. *Folia Phoniatrica et Logopaedica*. 21: 9-19.
171. Iwarsson, J. and Petersen, N.R. 2012. Effects of Consensus Training on the Reliability of Auditory Perceptual Ratings of Voice Quality. *Journal of Voice*. 26(3): 304-312.
172. Jacobson, B.H.; Johnson, A.; Grywalski C.; Silbergbeit A; Jacobson, G and Benninger, M.S. 1997. The Voice Handicap Index (VHI): Development and Validation. *American Journal of Speech Language Pathology*. 6(3): 66-70.

173. Jesus, L.; Barney, A.; Santos, R.; Caetano, J.; Jorge, J. and Couto, P.S. 2009- Universidade de Aveiro's Voice Evaluation Protocol. In *Proceedings of InterSpeech*, 971-974. Brighton, UK.
174. Jiang, J.J.; Zhang, Y.; McGilligan, C. 2006. Chaos in voice, form modeling to measurement. *Journal of Voice*. 20: 2-17.
175. Johnson, K. 2003. *Acoustics and Auditory Phonetics*. 2nd Edition. Malden: Blackwell.
176. Jones, K.; Sigmon, J.; Hock L. and Nelson, E. 2002. Prevalence and Risk Factors for Voice Problems among Telemarketers. *Archives of otolaryngology - head & neck surgery*. 128: 571-77.
177. Jones, P.M. 1999. Striking the Right Chord. *Royal College of Speech and Language Therapists Bulletin*. 565: 13-4.
178. Jones, S.M.; Carding, P.N.; Drinnan, M.J. Exploring the relationship between severity of dysphonia and voice-related quality of life. *Clinical Otolaryngology*. 31(5): 411-417.
179. Jones, T.M.; Trabold, M.; Plante, F.; Cheetham, B.M.G. and Earis, J.E. 2001. Objective assessment of hoarseness by measuring jitter. *Clinical Otolaryngology*. 26(1): 29-32.
180. Jónsdottir, V.I. 2002. Cordless amplifying system in classrooms. A descriptive study of teachers and student's options. *Logopedic Phoniatic Vocology*. 27: 29-36.
181. Jónsdottir, V.I.; Boyle, B.E.; Martin, P.J. and Sigurdardottir, G. 2002. A Comparison of the occurrence and nature of vocal symptoms in two groups of Icelandic teachers. *Logopedic Phoniatic Vocology*. 27:98-105.
182. Jónsdottir, V.I.; Laukkanen; A-M. and Vilkman, E. 2002. Changes in Teachers Speech during a Working Day with and without Electric Sound Amplification. *Folia Phoniatica et Logopaedica*. 54:282-287.
183. Karnell, M.P. 1991. Laryngeal perturbation analysis: minimum length of analysis window. *Journal of Speech and Hearing Research*. 34(3): 544-548.
184. Karnell, M.P.; Chang, A.; Smith, A. and Hoffman, H.T. 1997. Impact of signal type of validity on voice perturbation measures. *NCSV Status and Progress Report*. 11: 91-94.

185. Karnell, M.P.; Melton, S.D.; Childe, J.M.; Coleman, T.C.; Dailey, S.A. and Hoffman, H.T. 2007. Reliability of Clinician-Based (GRBAS and CAPE-V) and Patient-Based (V-RQOL and IPVI) Documentation of Voice Disorders." *Journal of Voice*. 21(5): 576-590.
186. Kasuya, H.; Ogawa, S. and Kikuchi, Y. 1986a. An adaptative comb filtering method as applied to acoustic analyses of pathological voices. *IEEE International Conference on Acoustics, Speech and Signal Processing*. ICASSP'86. 11: 669-672.
187. Kasuya, H.; Ogawa, S.; Mashima, K. and Ebihara, S. 1986b. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *Journal of the Acoustic Society of America*. 80(5): 1329-1334.
188. Kempster, G.B.; Gerratt, B.R.; Verdolini, A.K.; Barkmeier-Kraemer, J. and Hillman, R.E. 2009. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech and Language Pathology*. 18: 124-132.
189. Kempster, G.B.; Kistler, D.J.; Hillenbrand, J. 1991. Multidimensional scaling analysis of dysphonia in two speaker groups. *Journal of Speech and Hearing Research*. 34: 534-543.
190. Kent, R.D. and Ball, M.J. 2000. *Voice Quality Measurement*. San Diego: Singular Publishing.
191. Kent, R.D. and Read, C. 1992. *The Acoustics Analysis of Speech*. San Diego: Singular Publishing.
192. Kent, R.D.; Vorperian, H.K.; Kent, J.F. and Duffy, J.R. 2003. Voice dysfunction in dysarthria: applications of the multi-dimensional voice program. *Journal of Communication Disorders*. 36: 281-306.
193. Kent, R.D. 1996. Hearing and believing: some limits to the auditory-perceptual assessment in voice quality perception. *American Journal of Speech Language Pathology*. 5: 7-23.
194. Kent, R.D. 1993. Vocal tract acoustics. *Journal of Voice*. 7(2): 97-117.
195. Kent, R.D. 1997. *The Speech Sciences*. San Diego: Singular Publishing.
196. Kent, R.D. 2004. *The MIT Encyclopedia of Communication Disorders*. Massachusetts: Massachusetts Institute of Technology.

197. Klatt, D.H. and Klatt, L.C. 1990. Analysis, synthesis, and perception of voice: development of a standardized clinical protocol. *Journal of the Acoustic Society of America*. 87(2): 820-857.
198. Klingholtz, F. 1990. Acoustic recognition of voice disorders: a comparative study of running speech versus sustained vowels. *Journal of the Acoustic Society of America*. 87: 2218-2224.
199. Koike, Y. 1967. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Journal of the Acoustic Society of America*. 7: 17-23.
200. Kooijman, P.G.; de Jong, F.I.; Oudes, M.J.; Huinck, W.; van Acht, H. and Graamans, K. 2005. Muscular tension and body posture in relation to voice handicap quality in teachers with persistent voice complaints. *Folia Phoniatrica et Logopaedica*. 57: 134-147.
201. Koschke, D.C. 1993. *Voice Disability Index*. Madison: University of Wisconsin Hospital and Clinics.
202. Koufmann J. A. and Isaacson, G. (Edit). 1991. Voice Disorders. *Otolaryngologic Clinics of North America*. 24(5).
203. Koufmann J. and Blalock, D. 1998. Vocal Fatigue and Dysphonia in The Professional Voice User: Bogart-Bacall Syndrome. *Laryngoscope*. 98:493-8.
204. Kreiman, J. and Gerratt, B. 1998. Validity of rating scale measures of voice quality. *Journal of the Acoustic Society of America*. 104: 1598-1608.
205. Kreiman, J. and Gerratt, B. 2000. Sources of listener disagreement in voice quality assessment. *Journal of the Acoustic Society of America*. 108: 1867-1876.
206. Kreiman, J. and Gerratt, B. 2011. "Perceptual Assessment of Voice Quality: Past, Present and Future." In *Handbook of Voice Assessments*. Ma, E.; Yiu, E.. San Diego: Singular Publishing Inc..
207. Kreiman, J.; Gerratt, B. and Antoñanzas-Barroso, N. 2006. *Analysis and Synthesis of Pathological Voice Quality*. Los Angeles: Bureau of Glottal Affairs of the University of California.
208. Kreiman, J.; Gerratt, B. and Antoñanzas-Barroso, N. 2007. Measures of the glottal source spectrum. *Journal of Speech Language and Hearing Research*. 50: 595-610.

209. Kreiman, J.; Gerratt, B.R. and Ito, M. 2007. When and why listeners disagree in voice quality assessments tasks. *Journal of the Acoustic Society of America*. 122: 2354-2364.
210. Kreiman, J.; Gerratt, B.R.; Kempster, G.B.; Ernan, A. and Berke, G.S. 1993. Perceptual evaluation of voice quality: review, tutorial and a framework for future research. *Journal of Speech and Hearing Research*. 36: 21-40.
211. Kreiman, J.; Gerratt, B.R.; Precoda, K. 1990. Listener experience and perception of voice quality. *Journal of Speech and Hearing Research*. 33:103-15.
212. Kreiman, J.; Gerratt, B.R.; Precoda, K. and Berke, G.S. 1992. Individual differences in voice quality perception. *Journal of Speech and Hearing Research*. 35:512-520.
213. Kreiman, J. and Sidtis, D. 2011. *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Boston: Wiley-Blacjwell.
214. Kreiman, J.; Vanlancker-Sidtis, D. and Gerratt, B.R. 2005. "Perception of voice quality." In *Handbook of Speech Perception*. Pisoni, D.B.; Remez, R.E. (Edit.). Walden M.A.: Blackwell.
215. Ladefoged, P. 2006. *A course in Phonetics*. 5^a Edição. Boston: Thomson Wadsworth.
216. Landini, L.; Manfredi, C.; Positano, V.; Santarelli, M.F.; Vanello, N. 2002. Non-linear prediction for oesophageal voice analysis. *Medical Engineering and Physics journal*. 24: 529-533.
217. Languaite J.K. 1972. Adult Voice Screening. *Journal of Speech and Hearing Disorders*. 37: 147-51.
218. Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
219. Laver, J.; Hiller, S.; Mackenzie, J. and Rooney, E. 1986. An acoustic screening system for the detection of laryngeal pathology. *Journal of Phonetics*. 14: 517-524.
220. Lee, C.K. and Childers, D.G. 1991. "Some acoustical, perceptual and physiological aspects of vocal quality." In *Vocal Fold Physiology – acoustic*,

perceptual and physiological aspects of voice mechanisms. Gauffin, J. and Hammarberg, B. (Edit.). San Diego: Singular.

221. Lee, C.F.; Carding. P.N. and Fletcher, M. 2008. The nature and severity of voice disorders in lung cancer patients. *Logopedics Phoniatric Vocology*. 33: 93-103.
222. Le Huche, F. and Allali, A. 2005. *A Voz – Patologia de Orígen Funcional*. 2^a Edição, Porto Alegre: Artmed Editora.
223. Lewis, J.R.; Andreassen, M.L.; Leeper, H.A.; Marcrae, D.L. and Thomas, J. 1993. Vocal characteristics of children with cleft lip/palate and associated velopharyngeal incompetence. *Journal of Otolaryngology*. 22: 113-117.
224. Liberman, A.M. 1957. Some results of research on speech perception. *Journal of the Acoustic Society of America*. 31: 1490-1499.
225. Lieberman, P. 1961. Perturbations in vocal pitch. *Journal of the Acoustic Society of America*. 33: 597-603.
226. Linville, S.E.; Korabic, E.W. and Rosera, M. 1990. Intraproduction variability in jitter measures from elderly speakers. *Journal of Voice*. 4: 45-51.
227. Long J.; Williford H.N.; Olson, M.S. and Wolfe, V. 1998. Voice problems and risk factors among aerobic instructors. *Journal of Voice*. 12: 197-207.
228. Lopes, L.; Vaz-Freitas, S.; Sousa, R.; Matos, J.; Abreu, F. and Ferreira, A. 2008. A medida HNR: sua relevância na análise da voz e sua estimação precisa. Apresentação nas I Jornadas de Tecnologia e Saúde. Guarda: 30 de abril.
229. Ludlow, C.L.; Bassich, C.J.; Connor, N.P.; Coulter, D.C. and Lee, Y.J. 1987. "The validity of using phonatory jitter and shimmer to detect laryngeal pathology." In *Laryngeal Function in Phonation and Respiration*. Baer, T.; Sasakim C.; Harris, K. (Edit). Boston: College Hill.
230. MacCallum, J.K.; Cai, L.; Zhou, L.; Zhang, Y.; Jiang, J.J. 2009. Acoustic analysis of aperiodic voice: perturbation and nonlinear dynamic properties in oesophageal phonation. *Journal of Voice*. 2009; 23: 283-290.
231. Madill, C. and McCabe, P. 2011. "Acoustic Analysis Using Freeware: Praat" In *Handbook of Voice Assessments*. Ma, E. and Yiu, E. San Diego: Singular Publishing Inc..

232. Ma, E.P.-M. 2011. "Voice Range Profile: Phog" In *Handbook of Voice Assessments*. Ma, E. and Yiu, E. San Diego: Singular Publishing Inc..
233. Ma, E.P-M. and Yiu, E.M-L. 2011. *Handbook of Voice Assessments*. San Diego: Singular Publishing Group.
234. Ma, E.P-M. and Yiu, E.M-L. 2001. Voice Activity and Participation Profile: Assessing the Impact of Voice Disorders on Daily Activities. *Journal of Speech Language and Hearing Research*. 44(3): 511-524.
235. Ma, E.P. and Yiu, E.M.L. 2005. Suitability of acoustic perturbation measures in analyzing periodic and nearly periodic voice signals. *Folia Phoniatrica et Logopaedica*. 57(1): 38-47.
236. Ma, E.P-M. and Yiu, E.M.L. 2006. Multiparametric Evaluation of Dysphonic Severity. *Journal of Voice*. 20: 380-390.
237. Macallum, J.K.; Cai, L.; Zhou, L.; Zhang, Y. and Jiang, J.J. 2009. Acoustic analysis of aperiodic voice: perturbation and nonlinear dynamic properties in oesophageal phonation. *Journal of Voice*. 23(3): 283-90.
238. Mackenzie-Beck, J. 2005. "Perceptual analysis of voice quality: the place of vocal profile analysis." In *A figure of speech: a festschrift for John Laver*. Hardcastle, W.J.; Mackenzie-Beck, J. 285-322. Mahwah: Lawrence Erlbaum Associates.
239. Manfredi, C.; Giordano, A.; Schoentgen, J.; Frag, S.; Bocchi, L. and Dejonckere, P.H. 2012. Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools. *Computer Methods and Programs in Biomedicine Journal*. 7: 409-416.
240. McAllister, A.; Sederholm, E.; Ternstrom, S.O. and Sundberg, J. 1996. Perturbation and hoarseness: a pilot study of six children's voices. *Journal of Voice*. 10: 252-261.
241. Madazio, G. and Behlau, M. 2009. "Atualização em Voz", In *(Re)Habilitação Fonoaudiológica. Avaliação da Eficácia*. Berretin-Félix et al. São Paulo: Pulso Editorial Lda..
242. Massachusetts Eye and Ear Enfermary (MEEE) Voice and Speech Lab. 1994. *Disordered Voice Database, version 1,03*. CD-ROM & Operations Manual. New Jersey: Kay Elemetrics Corps.

243. Massaro, D.W. and Stork, D.G. 1998. Speech Recognition and Sensory Integration. *American Scientists Magazine*. 86: 236-244.
244. Maroco, J. 2010. *Análise Estatística – Com utilização do SPSS*, 3^a edição. Lisboa: Edições Sílabo.
245. Martens, J.; Versnel, H. and Dejonckere, P. 2007. The Effect of Visible Speech in the Perceptual Rating of Pathological Voices. *Journal of Voice*. 133: 178-185.
246. Martin, S. 1994. Voice care and development for teachers: survey report. *Journal of the British Voice Association*. 3: 92-8.
247. Martin, D.P.; Fitch, J. and Wolfe, V.I. 1995. Pathological voice type and the acoustic prediction of severity. *Journal of Speech and Hearing Research*. 38: 765-771.
248. Martin, D.P. and Wolfe, V.I. 1996. Effects of perceptual training on ability to use the vocal profile analysis scheme. *Perceptive Motor Skills*. 33: 1291-1298.
249. Maryn, Y.; Corthals, P.; DeBodt, M. and Van Cauwenberge, P. 2009a. Perturbation measures of voice: A comparative study between Multi-Dimensional Voice Program and Praat. *Folia Phoniatrica et Logopaedica*. 61: 217-226.
250. Maryn, Y.; Roy, N.; DeBodt, M.; Van Cauwenberge, P. and Corthals, P. 2009b. Acoustic measurement of overall voice quality: a meta-analysis. *Journal of the Acoustic Society of America*. 126(5): 2619-34.
251. Maryn, Y.; Corthals, P.; Van Cauwenberge, P.; Roy, N. and DeBodt, M. 2010. Toward Improved Ecological Validity in the Acoustic Measurement of Overall Voice Quality: Combining Continuous Speech and Sustained Vowels. *Journal of Voice*. 24(5): 540-55.
252. Mathieson, L.; Hirani, S.P.; Epstein, R.; Baken, R.J.; Wood, G. and Rubin, J.S. 2009. Laryngeal manual therapy: A preliminary study to examine its treatment effects in the management of muscle tension dysphonia. *Journal of Voice*. 23: 353-366.
253. Mattiske, J.A.; Oates, J.M. and Greenwood, K.M. 1998. Vocal Problems among Teachers: A Review of Prevalence, Causes, Prevention and Treatment. *Journal of Voice*. 12 (4): 489-499.

254. Mehta, D.D. and Hillman, R.E. 2008. Voice assessment: updates on perceptual, acoustic, aerodynamic and endoscopic imaging methods. *Current Opinion in Otolaryngology Head and Neck Surgery*. 16: 211-15.
255. Mendes, A.; Rothman, H.; Sapienza, C. and Brown, W. 2003. Effects of Vocal Training on the Acoustic Parameters of the Singing Voice. *Journal of Voice*. 17: 529-543.
256. Miller, M.K. and Verdolini, K. 1995. Frequency and risk factors for voice problems in teachers of singing and control subjects. *Journal of Voice*. 9: 348-62.
257. Michaelis, D.; Gramss, T. and Strube, H.W. 1997. Glottal-to-noise excitation ratio – a new measure for describing pathological voices. *Acustica – Acta Acustica*. 83: 700-706.
258. Michaelis, D.; Frölich, M. and Strube, H.W. 1998. Selection and combination of acoustic features for the description of pathologic voices. *Journal of the Acoustic Society of America*. 103: 1628-1638.
259. Milenkovic, P. and Read, C. 1992. *Cspeech Version 4 User's Manual*. Madison: WI-University of Wisconsin-Madison.
260. Mitev, P. and Hadjитодоров, S. 2000. A method for turbulent noise estimation in voiced signals. *Journal of Medical and Biological Engineering Computers*. 38: 625-631.
261. Mjaavatn, P.E. 1980. *Voice Difficulties among Teachers*. Apresentação do XVIII Congress of the International Association of Logopedics and Phoniatrics. Washington.
262. Moerman, M.; Martens, J.P.; Crevier-Buchman, L.; de Haan, E.; Grand, S.; Tessier, C.; Woisard, V. and Dejonckere, P.H. 2006b. The INFVo perceptual rating scale for substitution voices: development and reliability. *European Archives of Otorhinolaryngology*. 263(5): 435-439.
263. Moerman, M.; Martens, J.P.; Van Der Borgt, M.J.; Pelemans, M.; Gillis, M. and Dejonckere, P.H. 2006b. Perceptual evaluation of substitution voices: development and evaluation of the (I)INFVo rating scale. *European Archives of Otorhinolaryngology*. 263: 183-187.

264. Moers, C.; Möbius, B.; Rosanowski, F.; Noth, E.; Eysholdt, U. and Haderlein, T. 2012. Vowel- and Text-Based Cepstral Analysis of Chronic Hoarseness. *Journal of Voice*. 26(4): 416-424.
265. Monsen, R. and Engebretson, A. 1977. Study of variations on the male and female glottal wave. *Journal of the Acoustic Society of America*. 62: 981-993.
266. Moon, K.R.; Chung, S.M.; Park, H.S. and Kim, H.S. 2012. Materials of Acoustic Analysis: Sustained Vowel Versus Sentence. *Journal of Voice*. 26(5): 563-565.
267. Mora, R.; Jankowska, B.; Mora, F.; Crippa, B.; Dellepiane, M. and Salami, A. 2009. Effects of tonsillectomy on speech spectrum in children. *Journal of Voice*. 23: 614-618.
268. Moran, R.J.; Reilly, R.B.; De Chazal, P., P.D. and Lacy, P.D. 2006. Telephony-based voice pathology assessment using automated speech analysis. *IEEE Transactions on Biomedical Engineering*. 53(3): 468-477.
269. Morely, D.E. 1952. A ten-year survey of speech disorders among university students. *Journal of Speech and Hearing Disorders*. 25-31.
270. Morsomme, D.; Jamart, J.; Werry, C.; Giovanni, A.; Ramacle, M. 2001. Comparison between the GRBAS scale and the acoustic and aerodynamic measures provided by EVA for the assessment of dysphonia following vocal fold paralysis. *Folia Phoniatrica et Logopaedica*. 53(6): 317-325.
271. Morton V. 1995. *Educating Teachers*. Trabalho apresentado no 3rd Voice Symposium of Australia. Sydney.
272. Morton, V. and Watson, D.R. 1998. The Teaching Voice: Problems and Perceptions. *Logopedics Phoniatic Vocology*. 23: 133-139.
273. Murphy, P.J. 2006. Periodicity estimation in synthesized phonation signals using cepstral harmonic peaks. *Speech Communication Journal*. 48: 1704-1713.
274. Murray, L.; Johnston, B.; Lane, A.; Harveu, I.; Donovan, J.; Nair, P. and Harvey, R. 2003. Relation between body mass and gastro-oesophageal reflux symptoms: The Bristol Helicobacter Project. *International Journal of Epidemiology*. 32:645-650
275. Murry, T. and Doherty, E.T. 1980. Selected acoustic characteristics of pathologic and normal speakers. *Journal of Speech and Hearing Research*. 23: 361-369.

276. Murdoch, B. 2005. *Disartria. Uma abordagem Fisiológica para Avaliação e Tratamento*. São Paulo: Editora Lovise, Lta..
277. Nawka, T.; Anders, L.C. and Wendler, J. 1994. Die auditive Beurteilung heiserer Stimmen nach dem RGH-System. *Sprache-Stimme-Gehirn*. 18: 130-33.
278. Nearey, A. 1990. The segment as a unit of speech perception. *Journal of Phonetics*. 18: 347-373.
279. Nemr, K. and Lehn, C. 2010. "Voz em Câncer de Cabeça e PESCOÇO". In *Tratado de Fonoaudiologia*. Fernandes, F.D.M. et al. 2^a Edição. São Paulo: Roca.
280. Newman, P.W.; Harris, R.W. and Hilton, L.M. 1989. Vocal jitter and shimmer in stuttering. *Journal of Fluency Disorders*. 14: 87-95.
281. Nishio, M. and Niimi, S. 2008. Changes in speaking fundamental frequency characteristics with aging. *Folia Phoniatrica et Logopaedica*. 60: 120-127.
282. Nikolov, Z.; Deliyski, D.; Drumeva, L. and Boyanov, B. 1989. Computer system for diagnostics of pathological voices. *Proceedings of the 21st Congress International of Logopedics and Phoniatrics*. 1: 973-976.
283. Nittrouer, S.; McGowan, R.S.; Milenkovic, P.H. and Beehler, D. 1990. Acoustic measurements of men's and women's voices. . *Journal of Speech Language and Hearing Research*. 33: 761-775.
284. Oates, J. 2009. Auditory-perceptual evaluation of disordered vocal quality – pros, cons and future directions. *Folia Phoniatrica et Logopaedica*. 61 (1): 49-56.
285. Oates, J. and Russel, A. 1997. Perceptual Voice Profile. A Sound Judgement. CDROM.
286. Oguz, H.; Demirci, M.; Safak, M.A.; Arslan, N.; Islam, A. and Kargin, S. 2007. Effects of unilateral vocal cord paralysis on objective voice measures obtained by Praat. *European Archives of Otorhinolaryngology*. 264: 257-261.
287. Oliveira, I.B. 2004. Qualidade Vocal: Correlatos acústicos, perceptivo-auditivos e fisiológicos. *Revista Sociedade Brasileira de Fonoaudiologia*. Suplemento Especial.
288. Oliveira, I.B. 2010. "Avaliação Fonoaudiológica da Voz: Reflexões sobre Condutas com Enfoques à Voz Profissional". In *Tratado de Fonoaudiologia*. 2^a Edição. Fernandes, F.D.M et al. São Paulo: Roca.

289. Olszewski, A.E.; Shenm L.B.S. and Jiang, J.J. 2011. Objective Methods of Sample Selection in Acoustic Analysis of Voice. *Annals of Otology, Rhinology and Laryngology*. 120(3): 155-161.
290. Orlikoff, R.F.; Dejonckere, P.H.; Dembowski, J.; Fitch, J.; Gelfer, M.P.; Gerratt, B.R. et al. 1999. The perceived role of voice perception on clinical practice." *Phonoscope*. 2: 89-106.
291. Orlikoff, R.F. and Kahane, J.C. 1991. Influence of mean sound pressure level on jitter and shimmer measures. *Journal of Voice*. 5: 113-119.
292. Orr, R.; de Jong, F. and Cranen, B. 2002. Some objective measures indicate of perceived voice robustness in student teachers. *Logopedic Phoniatic Vocology*. 27: 106-117.
293. Pabon, J.P.H. 1991. Objective acoustic voice-quality parameters in the computer phonogram. *Journal of Voice*. 5: 203-216.
294. Pabon, J.P.H. 1988. Automatic phonogram recording supplemented with acoustical voice quality parameters. *Journal of Speech and Hearing Research*. 31: 710-722.
295. Patel, S. and Shrivastav, R. 2007. *Perception of dysphonic vocal quality: some thoughts and research update. Perspectives on voice and voice disorders*. 17:3-6 (ASHA SID-3).
296. Parsa, V. and Jamieson, D.G. 2001. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *Journal of Speech Language and Hearing Research*.44: 327-339.
297. Pekkarinen, E.; Himberg, L. and Pentti, T. 1992. Prevalence of Vocal symptoms among Teachers compared with nurses: a questionnaire study. *Scandinavian Journal of Logopedics and Phoniatrics*. 17:113-7.
298. Pereira, J.C. and Montagnoli, A.N. 1999. *Análise Acústica dos Sinais de Voz*. Sebenta. São Carlos: EESC/USP.
299. Pettersson, G.E. and Barney, H.L. 1952. Control methods used in a study of the identification of vowels. . *Journal of the Acoustic Society of America*. 24: 175-185.
300. Piccirillo, J.F.; Painter, C.; Fuller, D. and Fredrickson, J.M. 1998. Multivariate analyses of objective vocal function. *Annals of Otology, Rhinology and Laryngology*. 107: 107-112.

301. Piccirillo, J.F.; Painter, C.; Fuller, D.; Haiduk, A. and Fredrickson, J.M. 1998. Assessment of two objective voice function indices. *Annals of Otology, Rhinology and Laryngology*. 107: 175-185.
302. Pinho, S.M.R. 2003. *Fundamentos em Fonoaudiologia. Tratando os Distúrbios da Voz*, 2^a Edição. Rio de Janeiro: Editora Guanabara Kooogan.
303. Pinho, S.M.R. and Pontes, P. 2002. Escala de Avaliação Perceptiva da Fonte Glótica: RASAT. *Vox Brasilis*. 3: 11-13.
304. Pinho, S.M.R. and Pontes, P. 2008. *Músculos intrínsecos da Laringe e Dinâmica Vocal*. (Série Desvendando os Segredos da Voz). Volume 1. Rio de Janeiro: Revinter.
305. Pinho, S.; Tsuji, D. and Bohadana, S. 2006. *Fundamentos em Laringologia e Voz*. Rio de Janeiro: Editora Guanabara Koogan, S.A..
306. Pouchoulin, G. 2008. *Approche Statistique pour L'Analyse Objective et la Caractérisation de la Voix Dysphonique*. Dissertação de Doutoramento. Université d'Avignon et des Pays de Vaucluse.
307. Probst, R.; Grevers, G. and Iro, H. 2004. *Basic Otorhinolaryngology. A Step-by-Step Learning Guide*. New York: Thieme Publishing.
308. Qi, Y. And Hillman, R.E. 1997. Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *Journal of the Acoustic Society of America*. 102: 537-543.
309. Qi, Y.; Milstein, C. and Hillman, R.E. 1999. The estimation of signal-to-noise ratio in continuous speech for disordered voices. *Journal of the Acoustic Society of America*. 105: 2532-2535.
310. Rahn, D.A. 3rd; Chou, M.; Jiang, J.J. and Zhang, Y. 2007. Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. *Journal of Voice*. 21: 64-71.
311. Rabinov, C.R.; Kreiman, J.; Gerratt, B. and Bielamowicz, S. 1995. Comparing reliability of perceptual ratings and acoustic measures of jitter. *Journal of Speech and Hearing Research*. 38: 26-32.
312. Rammage, L.A.; Peppard, R.C.; Bless, D.M. 1982. Aerodynamic, laryngoscopic and perceptual-acoustic characteristics in dysphonic females with posterior glottal chinks: A retrospective study. *Journal of Voice*. 6(1):64-78.

313. Ramig, L.O. and Shipp, T. 1987. Comparative measures of vocal tremor and vocal vibrato. *Journal of Voice*. 1: 162-167.
314. Ramig, L.O. and Verdolini, K. 1998. Treatment Efficacy: voice disorders. *Journal of Speech Language and Hearing Research*. 41: S101-S116.
315. Ringel, R.L. and Chodzko-Zaiko, W.J. 1997. Vocal indices of biological age. *Journal of Voice*. 1: 31-37.
316. Repp, B.H. 1988. Integration and segregation in speech perception. *Language and Speech*. 31(3): 239-271.
317. Revis, J.; Giovanni, A.; Triglia, J.M. 2002. Influence of Voice Onset on the Perceptual Analysis od Dysphonia. *Folia Phoniatrica et Logopaedica*. 54(1): 19-25.
318. Revis, J.; Giovanni, A.; Wuyts, F.; Triglia, J.M. 1999. Comparision of Different Voice Samples for Perceptual Analysis. *Folia Phoniatrica et Logopaedica*. 51(3): 108-116.
319. Rothenberg, M. 1973. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *Journal of the Acoustic Society of America*. 53(6): 1632-1645.
320. Roy, N.; Gray, S.D.; Simon, M.; Dove, H.; Corbin-Lewis, K. and Stemple, J.C. 2001. An Evaluation of the Effects of Two Treatment Approaches for Teachers With Voice Disorders: A Prospective Randomized Clinical Trial. *Journal of Speech Language and Hearing Research*. 44: 286-96.
321. Roy, N.; Weinrich, B.; Gray, S.D.; Tanner, K.; Toledo, S.W.; Dove, H.; Corbin-Lewis, K. and Stemple, J.C. 2002. Voice amplification Versus Vocal Hygiene Instruction for Teachers With Voice Disorders: A Treatment Outcomes Study. *Journal of Speech Language and Hearing Research*. 45: 625-38.
322. Russell A.; Oates J. and Greenwood K.M. 1998. Prevalence of voice problems in teachers. *Journal of Voice*. 12:467-479.
323. Sakata, T.; Kubota, N.; Yonekawa, H.; Imaizumi, A.; Niimi, S. 1994. GRBAS evaluation of running speech and sustained phonations. *Annual Bulletin – Research Institute of Logopedics and Phoniatrics*. 28: 51-56.
324. Sader, R.C.M. and Hanayama, E.M. 2004. Considerações teóricas sobre a abordagem acústica da voz infantil. *Revista CEFAC*. 6(3): 312-8.

325. Sala, E.; Laine, A.; Simberg, S.; Pentti, J.; Suonpää, J. 2001. The Prevalence of Voice Disorders Among Day Care Center Teachers Compared with Nurses: A Questionnaire and Clinical Study. *Journal of Voice*. 15(3): 413-423.
326. Sapienza, C.M.; Crandell, C.C. and Curtis, B. 2003. Effects of Sound-Field Frequency Modulation Amplification on reducing Teachers Sound Pressure Level in the Classroom. *Journal of Voice*; 13(3): 375-381.
327. Sapir, S. 1993. Vocal Attrition in Voice Students: Survey Findings". *Journal of Voice*. 7: 69-74.
328. Sapir, S. 1995. Psychogenic spasmodic dysphonia: a case study with expert opinions. *Journal of Voice*. 9: 270-281.
329. Sapir, S.; Keidar, A. and Mathers-Schmidt, B. 1993. Vocal Attrition in Teachers: Survey Findings. *European Journal of Disorders of Communication*. 28: 177-85.
330. Sapir, S.; Mathers-Schmidt B. and Larson G.W. 1996. Singer`s and non-singer`s vocal health. Vocal behaviors and attitudes towards voice and singing: indirect findings from a questionnaire. *European Journal of Disorders of Communication*. 31: 193-209.
331. Sarfati, J. 1989. Readaptation vocale des enseignants. *Revue Laryngologie*. 110: 393-5.
332. Sataloff, R.T. 1997. *Professional Voice*. San Diego: Singular Publishing Group.
333. Scherer, K.R. 1981. "Vocal indicators of stress". In *Speech evaluation in psychiatry*. Darby, J.K. (Ed.). New York: Grune & Statton.
334. Scherer, K. R. 1986. Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. 99: 143-165.
335. Scherer, R.C.; Vail, V.J. and Guo, C. 1995. Required number of tokens to determine representative voice perturbation values. *Journal of Speech and Hearing Research*. 38: 1260-1269.
336. Schiavetti, N. and Metz, D.E. 2002. *Evaluating Research in Communicative Disorders*. 4th Edition. Boston: Allyn & Bacon.
337. Schoentgen, J. 1989. Jitter in sustained vowels and isolated sentences produced by dysphonic speakers. *Speech Communication Journal*. 8(1): 61-79.
338. Sederholm, E.; McAllister, A.; Sundber, J. and Dalkvist, J. 1992. Perceptual Analysis of Child Hoarseness using Continuous Scales. *STL-QPSR*. 1: 99-113.

339. Seikel, J.A.; King, D.W. and Drumright, D.G. 2010. *Anatomy and Physiology for Speech, Language and Hearing*. 4th Edition. New York: Delmar.
340. Shewell, C. 1998. The effect of perceptual training on ability to use the Vocal Profile Analysis scheme. *International Journal of Language and Communication Disorders*. 33: 322-326.
341. Shrivastav, R. 2011. "Evaluating Voice Quality." In *Handbook of Voice Assessments*. Ma, E.; Yiu, E.. San Diego: Singular Publishing Inc..
342. Shrivastav, R. 2006. Multidimensional scaling of breathy voice quality: individual differences in perception. *Journal of Voice*. 20: 211-222.
343. Shrivastav, R. 2003. The use of an auditory model in predicting perceptual ratings of breathy voice quality. *Journal of Voice*. 17: 503-512.
344. Shrivastav, R. and Camacho, A. 2010. A computational model to predict changes in breathiness resulting from variations in aspiration noise level. *Journal of Voice*. 24(4): 395-405.
345. Shrivastav, R.; Sapienza, C.M. and Nandur, V. 2005. Application of psychometric theory to the measurement of voice quality using rating scales. *Journal of Speech Language and Hearing Research*. 48: 323-335.
346. Shrivastav, R. and Sapienza, C.M. 2003. Objective measures of breathy voice quality obtained using an auditory model. *Journal of the Acoustic Society of America*. 114(4): 2217-2224.
347. Silva, D.; Oliveira, L.; Andrea, M. 2009. Jitter Estimation Algorithms for Detection of Pathological Voices. *EURASIP – Journal Advisory Signal Processing*. Article ID: 567875, 9 págs.
348. Simberg, S.; Laine, A.; Sala, E. and Rönnemaa, A. 2000. Prevalence of Voice Disorders Among Future Teachers. *Journal of Voice*. 14(2): 231-5.
349. Simberg, S.; Sala, E.; Laine, A. and Rönnemaa, A. 2001. A fast and easy method for voice disorders among teachers students. *Logopedic Phoniatic Vocology*. 26: 10-6.
350. Smith E.; Gray, S.D.; Dove, H.; Kirchner, L. and Heras, H. 1997. Frequency and Effects of Teachers Voice Problems. *Journal of Voice*. 11(1): 81-7.

351. Smith, E.; Kirchner, H.L; Taylor, M.; Hoffman, H. and Lemke, J.H. 1998. Voice Problems Among Teachers: Differences by Gender and Teaching Characteristics. *Journal of Voice*. 12 (3): 328-334.
352. Smith, E.; Lemke, J.H.; Taylor, M.; Kirchner, H.L. and Hoffman, H. 1998. Frequency of Voice Problems Among Teachers and Other Occupations. *Journal of Voice*. 12 (4): 480-88.
353. Smith, E.; Verdolini, K.; Gray, S. et al. 1996. Effects of Voice Disorders on Quality of Life. *Journal of Speech and Language Pathology*. 4: 223-44.
354. Smits, I.; Ceuppens, P. and De Bodt, M. 2005. A Comparative Study of Acoustic Voice Measurements by Means of Dr. Speech and Computerized Speech Lab. *Journal of Voice*. 19(2): 187-196.
355. Sofranko, J.L. and Prosek, R.A. 2012. The Effect of Experience on Classification of Voice Quality. *Journal of Voice*. 26(3): 299-303.
356. Speyer, R.; Wieneke, G.H. and Dejonckere, P.H. 2004a. The use of acoustic parameters for the evaluation of voice therapy for dysphonic patients. *Acta Acoustica united with Acustica*. 90: 520-527.
357. Speyer, R.; Wieneke, G.H. and Dejonckere, P.H. 2004b. Documentation of progress in voice therapy: perceptual, acoustic and laryngostroboscopic findings pretherapy and posttherapy. *Journal of Voice*. 18: 325-339.
358. Speyer, R.; Wieneke, G.H. and Dejonckere, P.H. 2004c. Self-assessment of voice therapy for chronic dysphonia. *Clinical Otolaryngology*. 29: 66-74.
359. Spielman, J.L.; Staar, A.C. and Hunter, E.J. 2006. *Recommendations for the creation of a recording laboratory*. Iowa City: National Center for Voice and Speech.
360. Sousa, R.; Vaz-Freitas, S. and Ferreira, A. 2011. Práticas Profissionais dos Terapeutas da Fala Portugueses nas Perturbações da Voz. *Revista da Sociedade Portuguesa de ORL*. 49(1): 11-17.
361. Souza, L.B.R. 2010. *Atuação Fonoaudiológica em Voz*. Rio de Janeiro: Livraria e Editora Revinter, Lda.
362. Sprecher, A., Olszewski, A., Jiang, J. J. and Zhang, Y. 2010. Updating signal typing in voice: addition of type 4 signals. *Journal of the Acoustic Society of America*. 127: 3710-16.

363. Stayner-Katusic, S.; Horga, D. and Zrinski, K.V. 2008. A longitudinal study of voice before and after phonosurgery for removal of a polyp. *Clinical Linguistics and Phonology*. 22(10-11): 857-863.
364. Stemple, J.C.; Glaze, L.E. and Ferdeman, B.K. 1995. *Clinical Voice Pathology: Theory and Management*. 2nd Edition. San Diego: Singular.
365. Stevens, K.N. 1989. On the quantal nature of speech." *Journal of Phonetics*. 17: 3-46.
366. Stevens, K.N. 1998. *Acoustic Phonetics*. Cambridge, Massachusetts: The MIT Press.
367. Subramaniam, V. and Kumar, P. 2009. Impact of tonsillectomy with or without adenoidectomy on the acoustic parameters of voice: a comparative study. *Archives of Otolaryngology - head & neck surgery*. 135: 966-969.
368. Sweeney, T. and Sell, D. 2008. Relationship between perceptual ratings of nasality and nasometry in children/adolescents with cleft and/or velopharyngeal dysfunction. *International Journal of Language and Communication Disorders*. 43(3): 265-82.
369. Takahashi, H. and Koike, Y. 1976. Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Oto-Laryngologica*. 338: 2-22.
370. Tavares, E.L.M. and Martins, R.H.G. 2007. Vocal Evaluation in Teachers With or Without Symptoms. *Journal of Voice*. 21(4): 407-414.
371. Teager, H.M. and Teager, S.M. 1990. "Evidence for nonlinear sound production mechanisms in the vocal tract". In *Speech Production and Speech Modeling*. Hardcastle, W.J. and Marchal, A. Dordrecht: Kluwer.
372. Timmermans, B.; De Bodt, M.S.; Wuyls, F.L. et al. 2002. Poor quality on future elite vocal performers and professional voice users. *Journal of Voice*. 16: 372-382.
373. Timmermans, B.; De Bodt, M.S.; Wuyls, F.L. and Van de Heyning, P. 2004. Voice quality changes in future professional voice users after 18 months of voice training. *Folia Phoniatrica et Logopaedica*. 261: 1-5.
374. Timmermans, B.; De Bodt, M.S.; Wuyls, F.L. and Van de Heyning, P. 2005. Voice quality changes in future professional voice users after 9 months of voice training. *European Archives of Otorhinolaryngology*. 19: 511-518.

375. Titze, I.R. 1989. On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustic Society of America*. 85: 901-912.
376. Titze, I.R. 1993. Standards in acoustic analysis of voice. *Journal of Voice*. 8(1): 1-7.
377. Titze, I.R. 1994. *Principles of Voice Production*. Englewood Cliffs: Prentice-Hall.
378. Titze, I.R. 1995. *Summary Statement. Workshop on Acoustic Voice Analysis*. Denver, C.O.: The National Center for Voice and Speech.
379. Titze, I.R.; Bergan, C.C.; Hunter, E.J. and Story, B. 2003. Source and filter adjustments affecting the perception of the vocal qualities twang and yawn. *Logopedics Phoniatric Vocology*. 28: 147-155.
380. Titze, I.R.; Lemke, J. and Montequin, D. 1997. Population in the US workforce who rely on voice as a primary tool of trade: a preliminary report. *Journal of Voice*. 11: 254:259.
381. Titze, I.R. and Liang, H. 1993. Comparison of F0 extraction methods for high-precision voice perturbation measurements. *Journal of Speech and Hearing Research*. 36(6): 1120-1133.
382. Titze, I.R. and Winholtz, W.S. 1993. Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research*. 36(6): 1177-1190.
383. Toner, M.A.; Emanuel, F.W. and Parker, D. 1990. Relationship of spectral noise levels to psychophysical scaling of vowel roughness. *Journal of Speech and Hearing Research*. 33: 238-244.
384. Torre III, P. and Barlow, J.A. 2009. Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*. 42: 324-333.
385. Uloza, V.; Verikas, A.; Bacauskiene, M.; Gelzinis, A.; Pribuisienne, R.; Keseta, M. and Saferis, V. 2011. Categorizing Normal and Pathological Voices: Automated and Perceptual Categorization. *Journal of Voice*. 25: 700-708.
386. Umapathy, K.; Krishnan, S.; Parsa, V. and Jamieson, D.G. 2005. Discrimination of pathological voices using a time-frequency approach. *IEEE Trans Biomedical Engineering*. 52(3): 421-430.

387. Urrutikoetxea, A.; Ispizua, A.; Matellanes, F. and Aurrekoetxea, J. 1995. *Prevalence of Vocal Nodules in Female Teachers*. Apresentação em Vídeo no 1º Congresso Mundial da Voz. Porto.
388. van Gogh, V.D.L.; Verdonck-de Leeuw, I.M.; Landendijk, J.A. and Kuik, D.J.; Mahieu, H.F. 2012. Long-Term Efficacy of Voice Therapy in Patients With Voice Problems After Treatment of Early Glottic Cancer. *Journal of Voice*. 26(3): 198-401.
389. Van Lierde, K.M.; Vinck, B.M.; Baudonck, N.; De Vel, E. and Dhooge, I. 2005a. Comparison of the overall intelligibility, articulation, resonance, and voice characteristics between children using cochlear implants and those using bilateral hearing aids: a pilot study. *International Journal of Audiology*. 44: 452-465.
390. Van Lierde, K.M.; Vinck, B.; De Ley, S.; Clement, G. and Van Cauwenberge, P. 2005b. Genetics of vocal quality characteristics in monozygotic twins: a multiparameter approach. *Journal of Voice*. 19: 511-518.
391. Van Lierde, K.M.; Claeys, S.; De Bodt, M. and Van Cauwenberge, P. 2004a. Vocal quality characteristics on children with cleft palate: a multiparameter approach. *Journal of Voice*. 18: 254-362.
392. Van Lierde, K.M.; De Ley, S.; Clement, G; De Bodt, M. and Van Cauwenberge, P. 2004b. Outcome of laryngeal manual therapy in four Dutch adults with persistent moderato-to-severe vocal hyperfunction: a pilot study. *Journal of Voice*. 18: 467-474.
393. Van Lierde, K.M.; De Ley, S.; De Bodt, M. and Van Cauwenberge, P. 2007. Long-Term Outcome of Hyperfunctional Voice Disorders Based on a Multiparameter Approach. *Journal of Voice*. 21(2): 179-188.
394. Vasiliakis, M. and Stylianau, Y. 2009. Voice pathology detection based on short-term jitter estimations in running speech. *Folia Phoniatrica et Logopaedica*. 61(3): 153-170.
395. Vaziri, G.; Almassganj, F. and Behroozmand, R. 2010. Pathological assessment of patients 'speech signals using nonlinear dynamical analysis. *Computer in Biology and Medicine Journal*. 40(1): 54-63.

396. Verdolini, K. 1994. "Voice Disorders." In Tomblin, B.; Morris, H.L.; Spriesterbach, D.C. (Edit). *Diagnosis in Speech-Language Pathology*. San Diego: Singular Publishing Group.
397. Verdolini, K. and Ramig, L. O. 2001. Review: Occupational Risks for Voice Problems. *Logopedic Phoniatic Vocology*. 26: 37-46.
398. Verdolini, K.; Rosen, C.A. and Branski, R. 2006. *Classification manual for voice disorders – I. SID 3, Voice and Voice Disorders*. ASHA.
399. Verdolini-Marston, K.; Titze, I.R. and Druker, D.G. 1990. Changes in phonation threshold pressure with induced conditions of hydration. *Journal of Voice*. 4: 142-51.
400. Verdonck-de Leeuw, I.M. and Mathieu, H.F. 2004. Vocal aging and the impact on daily life: A longitudinal study. *Journal of Voice*. 18(2): 193-202.
401. Vertigan, A.E.; Theodoros, D.G.; Winkworth, A.L.; Gibson, P.G.. 2008. A Comparison of Two Approaches to the Treatment of Chronic Cough: Perceptual, Acoustic, and Electroglossographic Outcomes. *Journal of Voice*. 22(5): 581-589.
402. Vieira, M.N.; Maran, A.G.; McInnes, F.G. and Jack, M.A. 1996. Detecting arytenoids cartilage misplacement through acoustic and electroglossographic jitter analysis. In *4th International Conference on Spoken Language Processing*. Philadelphia.
403. Vieira, M.N.; McInnes, F.R. and Jack, M.A. 2002. On the influence of laryngeal pathologies on acoustic and electroglossographic jitter measures. *Journal of the Acoustic Society of America*. 111(2): 1045-55.
404. Vieira, V.P.; De Biase, N. and Pontes, P. 2005. Análise Acústica e Perceptivo-Auditiva vs Coaptação Glótica em Alteração Estrutural Mínima. *Acta ORL*. 6-12.
405. Vilkman, E. 2000. Voice Problems at work: a challenge for occupational safety and health arrangement. *Folia Phoniatica et Logopaedica*. 52: 120-125.
406. Vogel, A.P. 2011. "Multidimensional Analysis of Voice: Computerized Speech Lab." In *Handbook of Voice Assessments*. Ma, E. and Yiu, E. San Diego: Singular Publishing Inc..
407. Vogel, A.P. and Maruff, P. 2008. Comparison of voice acquisition methodologies in speech research. *Behaviour Research Methods*. 40(4): 982-987.

408. Vogel, A.P.; Maruff, P.; Snyder, P.J. and Mundt, J.C. 2009. Standardization of pitch-range settings in voice acoustic analysis. *Behavior Research Methods*. 41(2): 318-324.
409. Vogel, A.P. and Morgan, A.T. 2009. Factors affecting the quality of sound recording for speech and voice analysis. *International Journal of Speech-Language Pathology*. 11(6): 431-437.
410. Waugh, I. 2000. *Digital Audio Recording*. Kent: PC Publishing.
411. Webb, A.L.; Carding, P.N.; Deary, I.J.; Markenzie, K.; Steen, N. and Wilson, J.A. 2004. The reliability of three perceptual evaluation scales for dysphonia. *European Archives of Otorhinolaryngology*. 261: 429-34.
412. Weber, S.A.T. 2002. *Estudo das relações laringeas e do grau de disfonia em pacientes com paracoccidioidomicose*. Dissertação de Mestrado. Universidade do Estado de São Paulo. Botucatu.
413. Wertzner, H.F.; Schreiverm S. and Amaro, L. 2005. Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders. *Brazilian Journal of Otorhinolaryngology*. 71: 582-588.
414. Wilson, D.K. 1987. *Buffalo III Voice Screening Profile*. Baltimore: Williams & Wilkins.
415. Wirz, S. and Beck, J.M. 1995. *Assessment of Voice Quality: The vocal profile analysis scheme*. London: Whurr.
416. Wolfe, V.; Cornell, R. and Palmer, C. 1991. Acoustic correlates of pathologic voice types. *Journal of Speech, Language and Hearing Research*. 34: 509-516.
417. Wolfe, V.; Fitch, J. and Cornell, R. 1995. Acoustic prediction of severity in commonly occurring voice problems. *Journal of Speech, Language and Hearing Research*. 38(2): 273-279.
418. Wolfe, V. and Martin, D. 1997. Acoustic correlates of dysphonia: type and severity. *Journal of Communication Disorders*. 30: 403-416.
419. Wolfe, V.; Martin, D. and Palmer, C. 2000. Perception of dysphonic voice quality by naïve listeners. *Journal of Speech, Language and Hearing Research*. 43: 697-705.
420. Wolfe, V. and Steinfatt, T. 1987. Prediction of vocal severity within and across voice types. *Journal of Speech, Language and Hearing Research*. 30: 230-40.

421. Wong, D.Y.; Markel, J.D. and Junior, A.H.G. 1979. Latest squares glottal inverse filtering from the acoustic speech waveform. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 27:350-355.
422. Wong, D.Y.; Ito, M.R.; Cox, N.B. and Titze, I.R. 1991. Observation of perturbations in a lumped-element model of the vocal folds with applications to some pathological cases. *Journal of the Acoustic Society of America*. 89(1): 383-394.
423. Wong, D.Y.; Lange, R.; Titze, I.R. and Guo, C.G. 1995. A qualitative study of mechanisms of jitter-induced shimmer in the voice. *Journal of the Acoustic Society of America*. 97(5): 3421-3421.
424. Wuyts, F.L.; De Bodt, M.S.; Bruckers, L. and Molenberghs, G. 1996. Normative data stem. *Acta Otorhinolaryngologica Belgica*. 50: 331-341.
425. Wuyts, F.L.; De Bodt, M.S.; Molenberghs, G.; Remacle, M. Heyler, L. and Millett, B. 2000. The Dysphonic Severity Index: an objective measure of vocal quality based on a multiparameter approach. *Journal of Speech, Language and Hearing Research*. 43: 796-809.
426. Wuyts, F.L.; De Bodt, M.S. and Van de Heyning, P.H. 1999. Is the Reliability of a Visual Analog Scale Higher Than an ordinal Scale? An Experiment with the GRBAS Scale for the Perceptual Evaluation of Dysphonia. *Journal of Voice*. 13(4): 508-517.
427. Xue, S.A. and Deliyski, D. 2002. Effects of aging on selected acoustic voice parameters: preliminary normative data and educational implications. *Education and Gerontology*. 22(2): 159-168.
428. Yamaguchi, H.; Shrivastav, R.; Andrews, M.L. and Niimi, S. 2003. A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatrica et Logopaedica*. 55: 147-157.
429. Yamasaki, R.; Leão, S.H.S; Madazio, G.; Padovani, M. and Azevedo, R. 2007. "Análise perceptivo-auditiva de vozes normais e alteradas: Escala analógica visual". In *XV Congresso Brasileiro de Fonoaudiologia e VII Congresso Internacional de Fonoaudiologia*. Outubro; Gramado – RS.
430. Yamasaki, R.; Leão, S.H.S; Madazio, G.; Padovani, M.; Azevedo, R. and Behlau, M. 2008. Correspondência entre Escala Analógico-Visual e a Escala Numérica na

Avaliação Perceptivo-Auditiva de Vozes. In *XVI Congresso Brasileiro de Fonoaudiologia*. Outubro; Campos de Jordão – SP.

431. Yamauchi, E.J.; Imaizumi, S.; Maruyama, H. and Haji, T. 2010. Perceptual evaluation of pathological voice quality: a comparative analysis between the RASATI and GRBASI scales. *Logopedics Phoniatrics Vocology*. 35(3): 121-8.
432. Yiu, E.M.-L. 1999. Limitations of perturbation measures in clinical acoustic voice analysis. *Asia Pacific Journal of Speech, Language and Hearing*. 4: 157-168.
433. Yiu, E.M.-L.; Murdoch, B.; Hird, K. and Lau, P. 2002. Perception of synthesized voice quality in connected speech by Cantonese speakers. *Journal of the Acoustic Society of America*. 112(3 Pt. 1): 1091-1101.
434. Yiu, E.M.-L. and Ng, C.-Y. 2004. Equal appearing interval and visual analogue scaling of perceptual roughness and breathiness. *Clinical Linguistics & Phonetics*. 18(3): 211-229.
435. Yiu, E.M.-L.; Worrall, L.; Longland, J. and Mitchell, C. 2000. Analyzing vocal quality of connected speech using Kay's computerized speech lab: a preliminary finding. *Clinical Linguistics & Phonetics*. 14 (4): 295-305.
436. Yiu, E.M.-L. 2002. Impact and Prevention of Voice Problems in the Teaching Profession: Embracing the Consumer's View. *Journal of Voice*. 16(2): 215-28.
437. Yu, P., Garrel, R., Nicollas, R., Ouaknine, M., and Giovanni, A. 2007. Objective voice analysis in dysphonic patients: new data including nonlinear measurements. *Folia Phoniatrica et Logopaedica*, 59(1): 20-30.
438. Yu, P.; Ouaknine, M. and Giovanni, A. 2001. Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements. *Journal of Voice*. 15: 529-542.
439. Yu, P.; Revis, J.; Wuyts, F.L.; Zanaret, M. and Giovanni, A. 2002. Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analogue scale. *Folia Phoniatrica et Logopaedica*. 54: 271-281.
440. Yumoto, E. 1983. The quantitative evaluation of hoarseness. *Archives of otolaryngology - head & neck surgery*. 109(1): 48-52.
441. Yumoto, E.; Gould, W.J. and Baer, T. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustic Society of America*. 71(6): 1544-1549.

442. Yumoto, E.; Sasaki, Y. and Okamura, H. 1984. Harmonics-to-noise ratio and psychological measurement of the degree of harshness. *Journal of Speech and Hearing Research*. 27(1): 2-6.
443. Zhang, Y. and Jiang, J.J. 2008. Acoustic Analyses of Sustained and Running Voices of Patients with Laryngeal Pathologies. *Journal of Voice*. 22: 1-9.
444. Zhang, Y.; Jiang, J.J.; Biazzo, L. and Jorgensen, M. 2005. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *Journal of Voice*. 19(4): 519-28.
445. Zhang, Y.; Jiang, J.J. and Rahn, D.A. 3rd. 2005. Studying vocal fold vibrations in Parkinson's disease with a nonlinear model. *Chaos*. 15: 33903.
446. Zraick, R.R.; Gentry, M.A.; Smith-Olinde, L. and Gregg, B.A. 2006. The effect of speaking context on elicitation of habitual pitch. *Journal of Voice*. 20(4): 545-559.
447. Zraick, R.R.; Skaggs, S.D. and Montague, J.C. 2000. The effect of task on determination of habitual pitch. *Journal of Voice*. 14: 484-489.
448. Zraick, R.R.; Wenderl, K. and Smith-Olinde, L. 2005. The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *Journal of Voice*. 19: 574-581.

ANEXOS

Anexo I – CAPE-V – Consensus Auditory-Perceptual Evaluation of Voice in Protocolo de Avaliação da Qualidade Vocal da Universidade de Aveiro (Jesus et al., 2009).

Consensus Auditory-Perceptual Evaluation of Voice (CAPE – V)

Os parâmetros da qualidade vocal que se seguem, devem ser avaliados com recurso às seguintes tarefas:

1. Vogais sustentadas /a/ e /i/ (três repetições com a duração de 3-5 segundos cada)
2. Produção de Frases:
 - a. *A Marta e o avô vivem naquele casarão rosa velho.*
 - b. *Sofia saiu cedo da sala.*
 - c. *A aia do avião andava avariada.*
 - d. *Agora é hora de acabar.*
 - e. *A minha mãe mandou-me embora.*
 - f. *O Tiago comeu quatro peras.*
3. Fala espontânea (mínimo 20 segundos):
 - a. Novo paciente: "Fale-me como começou o seu problema de voz, quando o notou e o que fez em relação a isso.";
 - b. Paciente em acompanhamento: "Diga-me como está a sua voz."

Legenda: C = Consistente I = Inconsistente
 AL: Alterações ligeiras
 AM: Alterações moderadas
 AS: Alterações severas

Pontuação

Grau de Severidade Global AL AM AS C I ____/100

Rouquidão AL AM AS C I ____/100

Soprosidade AL AM AS C I ____/100

Tensão AL AM AS C I ____/100

Altura Tonal (Indicar o tipo de alteração): _____
 AL AM AS C I ____/100

Intensidade (Indicar o tipo de alteração): _____
 AL AM AS C I ____/100

 AL AM AS C I ____/100

 AL AM AS C I ____/100

RESSONÂNCIA: Normal Alterada (breve descrição): _____

OUTROS PARÂMETROS (por exemplo, diplofonia, asperezza, falseto, astenia, afonia, bitonalidade, tremor, estridente, «glottal fry», entre outros aspectos relevantes) _____

Escala G.R.B.A.S. (G: grau de rouquidão; R: asperezza; B: soprosidade; A: astenia; S: tensão)

G ___ R ___ B ___ A ___ S ___ (0=normal; 1=alterações ligeiras; 2=alterações moderadas; 3=alterações severas)

Anexo II – Testes de *Kolmogorov-Smirnov* para as distribuições das medidas acústicas de acordo com a amplificação da voz (valores-p > 0.05, logo não se rejeita a normalidade das distribuições).

Tests of Normality

Ampilificação da voz	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Mean F0 (Hz)	Não Amplificada	.201	12	.193	.903	12
	Amplificada (ampl20db)	.201	12	.193	.903	12
SDFO (Hz)	Não Amplificada	.165	12	.200 [*]	.953	12
	Amplificada (ampl20db)	.166	12	.200 [*]	.953	12
Jitter (PPQ5) %	Não Amplificada	.183	12	.200 [*]	.903	12
	Amplificada (ampl20db)	.184	12	.200 [*]	.903	12
Jitter (Jitt) %	Não Amplificada	.177	12	.200 [*]	.904	12
	Amplificada (ampl20db)	.200	12	.199	.927	12
Shimmer (APQ5) %	Não Amplificada	.154	12	.200 [*]	.913	12
	Amplificada (ampl20db)	.154	12	.200 [*]	.913	12
Shimmer (Shim) %	Não Amplificada	.162	12	.200 [*]	.919	12
	Amplificada (ampl20db)	.162	12	.200 [*]	.919	12
NHR (dB)	Não Amplificada	.209	12	.155	.883	12
	Amplificada (ampl20db)	.209	12	.155	.883	12

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Anexo III - Testes de *Kolmogorov-Smirnov* para as distribuições das medidas acústicas de acordo com o programa de *software* (praticamente para a totalidade dos casos em estudo, os valores-p < 0.05, logo rejeita-se a normalidade das distribuições). Numa primeira fase, de forma a transformar as distribuições das medidas acústicas, foram efetuadas diversas transformações para a simetria utilizando o “log10” e a “raiz quadrada”, no entanto, este procedimento não trouxe a pretendida simetria das distribuições. Foi ainda executada a eliminação dos valores extremos severos (*outliers*), mas, de igual forma, não foi conseguida a simetria das distribuições.

Tests of Normality^{b,c}

Tipo de software	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MeanF0(Hz)	Dr. Speech	.104	86	.022	.978	.157
	MDVP	.098	90	.033	.983	.300
	Praat	.105	90	.015	.974	.069
	Voice Studio	.106	90	.015	.973	.055
SDF0(Hz)	Dr. Speech	.140	86	.000	.762	.000
	MDVP	.363	90	.000	.294	.000
	Praat	.461	90	.000	.363	.000
	Voice Studio	.315	90	.000	.462	.000
HNR(dB)	Dr. Speech	.090	86	.081	.952	.003
	MDVP	.223	90	.000	.729	.000
	Praat	.081	90	.191	.950	.002
	Voice Studio	.143	90	.000	.913	.000
jitter(local)%	Dr. Speech	.213	86	.000	.635	.000
	MDVP	.191	90	.000	.730	.000
	Praat	.267	90	.000	.520	.000
	Voice Studio	.224	90	.000	.653	.000
Jitter(PPQ5)%	MDVP	.207	90	.000	.685	.000
	Praat	.246	90	.000	.572	.000
	Voice Studio	.239	90	.000	.623	.000
Shimmer(local)%	Dr. Speech	.157	86	.000	.762	.000
	MDVP	.223	90	.000	.709	.000
	Praat	.181	90	.000	.736	.000
	Voice Studio	.212	90	.000	.701	.000
Shimmer(APQ5)%	MDVP	.250	90	.000	.633	.000
	Praat	.176	90	.000	.755	.000
	Voice Studio	.204	90	.000	.706	.000

a. Lilliefors Significance Correction

b. There are no valid cases for Jitter(PPQ5)% when Tipo de software = 1.000. Statistics cannot be computed for this level.

c. There are no valid cases for Shimmer(APQ5)% when Tipo de software = 1.000. Statistics cannot be computed for this level.

Anexo IV - Testes de *Kolmogorov-Smirnov* para as distribuições das medidas acústicas de acordo com o tipo de voz, para cada programa de *software* (praticamente para a totalidade dos casos em estudo, os valores-p < 0.05, logo rejeita-se a normalidade das distribuições).

3.1. Dr. Speech

Tests of Normality

Tipo de Voz	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MeanF0(Hz)	.078	85	.200 ^b	.982	85	.292
	.104	86	.022	.978	86	.157
SDF0(Hz)	.165	85	.000	.757	85	.000
	.140	86	.000	.762	86	.000
jitter(local)%	.244	85	.000	.571	85	.000
	.213	86	.000	.635	86	.000
Shimmer(local)%	.202	85	.000	.763	85	.000
	.157	86	.000	.762	86	.000
HNR(dB)	.117	85	.006	.944	85	.001
	.090	86	.081	.952	86	.003

a. Lilliefors Significance Correction

^b. This is a lower bound of the true significance.

3.2. Praat

Tests of Normality

Tipo de Voz	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MeanF0(Hz)	.083	90	.169	.973	90	.060
	.105	90	.015	.974	90	.069
SDF0(Hz)	.424	90	.000	.415	90	.000
	.461	90	.000	.363	90	.000
Jitter(PPQ5)%	.251	90	.000	.576	90	.000
	.246	90	.000	.572	90	.000
Jitter(local)%	.260	90	.000	.531	90	.000
	.267	90	.000	.520	90	.000
Shimmer(APQ5)%	.170	90	.000	.779	90	.000
	.176	90	.000	.755	90	.000
Shimmer(local)%	.180	90	.000	.750	90	.000
	.181	90	.000	.736	90	.000
HNR(dB)	.080	90	.200 ^b	.949	90	.002
	.081	90	.191	.950	90	.002

a. Lilliefors Significance Correction

^b. This is a lower bound of the true significance.

3.3. Voice Studio

Tests of Normality

Tipo de Voz	Kolmogorov-Smirnov ^a			Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
MeanF0(Hz)	Inteira	.089	90	.077	.975	90	.076
	Segmentada	.106	90	.015	.973	90	.055
SDFO(Hz)	Inteira	.330	90	.000	.413	90	.000
	Segmentada	.315	90	.000	.462	90	.000
Jitter(PPQ5)%	Inteira	.211	90	.000	.695	90	.000
	Segmentada	.239	90	.000	.623	90	.000
Jitter(local)%	Inteira	.212	90	.000	.682	90	.000
	Segmentada	.224	90	.000	.653	90	.000
Shimmer(APQ5)%	Inteira	.187	90	.000	.740	90	.000
	Segmentada	.204	90	.000	.706	90	.000
Shimmer(local)%	Inteira	.189	90	.000	.728	90	.000
	Segmentada	.212	90	.000	.701	90	.000
HNR(dB)	Inteira	.136	90	.000	.922	90	.000
	Segmentada	.143	90	.000	.913	90	.000

a. Lilliefors Significance Correction

Anexo V – Simetria da distribuição das diferenças D_i . Considera-se que uma distribuição é simétrica se tiver um valor entre aproximadamente -2 e +2. O quadro seguinte, mostra-nos que a larga maioria das distribuições da diferença da GRBAS para os peritos, têm um valor entre aproximadamente -2 e +2, assumindo-se assim a simetria das distribuições, o que permite assumir a informação proveniente do *teste de Wilcoxon* como fidedigna.

	Perito									
Diferenças	1	2	3	4	5	6	7	8	9	10
dif_G	-0.1	0.0	-0.8	2.0	-0.5	-2.9	0.1	1.3	0.2	2.4
dif_R	1.7	-0.7	0.2	2.5	0.5	-1.9	-0.7	1.4	1.3	0.7
dif_S	-1.4	4.6	0.4	2.0	3.1	2.7	-1.5	-0.1	1.0	1.4
dif_A	3.3	-1.2	-1.2	2.0	0.3	-2.4	-1.2	-1.1	-0.1	-4.1
dif_T	1.3	0.6	-3.1	-1.2	-1.4	1.9	1.3	-2.0	-3.9	-2.0

Simetria = (*Skewness/Std. Error Skewness*)

Anexo VI – Regressões Lineares Múltiplas.

5.1. Dr. Speech

Variável independente	Coef. não normalizados		Coef. normalizados		t	valor-p
	B	Erro Padrão	Beta			
G ($R^2_a = 0,61$)						
(Constante)	100.83	6.74			14.97	0.000
HNRdB	-2.59	0.26	-0.687		-9.97	0.000
SDF0Hz	6.26	1.31	0.329		4.77	0.000
R ($R^2_a = 0,64$)						
(Constante)	30.44	5.97			5.10	0.000
Shimmerlocal	8.47	0.92	0.650		9.25	0.000
MeanF0Hz	-0.13	0.03	-0.344		-4.67	0.000
SDF0Hz	4.22	1.36	0.230		3.11	0.003
B ($R^2_a = 0,66$)						
(Constante)	88.01	16.47			5.34	0.000
HNRdB	-3.46	0.52	-0.784		-6.62	0.000
MeanF0Hz	0.16	0.03	0.357		5.44	0.000
jitterlocal	-43.35	11.03	-0.455		-3.93	0.000
Shimmerlocal	4.48	1.71	0.372		2.62	0.011
A ($R^2_a = 0,35$)						
(Constante)	105.48	15.18			6.95	0.000
HNRdB	-3.20	0.54	-0.817		-5.96	0.000
jitterlocal	-27.36	11.16	-0.336		-2.45	0.016
S ($R^2_a = 0,23$)						
(Constante)	2.59	3.72			0.70	0.488
SDF0Hz	9.07	1.77	0.492		5.12	0.000

R^2_a = Coeficiente de determinação ajustado

Vozes Outliers: G=1,5,40; R=47,71,75,90,91,99; B=14,56,78,93; A=55,79,94,99; S=66,99

5.2. MDVP

Variável independente	Coef. não normalizados		Coef. normalizados	t	valor-p
	B	Erro Padrão	Beta		
G ($R^2_a = 0,53$)					
(Constante)	30.98	2.58		11.99	0.000
Shimmerlocal	4.18	0.53	0.926	7.94	0.000
SDF0Hz	-0.36	0.16	-0.266	-2.28	0.025
R ($R^2_a = 0,55$)					
(Constante)	11.82	2.57		4.60	0.000
Shimmerlocal	4.50	0.45	0.742	10.08	0.000
B ($R^2_a = 0,68$)					
(Constant)	-6.02	6.46		-0.93	0.354
JitterPPQ5	21.28	3.23	0.692	6.60	0.000
SDF0Hz	-0.96	0.17	-0.522	-5.61	0.000
Shimmerlocal	3.08	0.71	0.491	4.32	0.000
MeanF0Hz	0.08	0.03	0.197	2.87	0.005
A ($R^2_a = 0,29$)					
(Constante)	8.08	3.02		2.67	0.009
JitterPPQ5	16.73	2.84	0.546	5.89	0.000
S ($R^2_a = 0,17$)					
(Constante)	16.52	2.21		7.47	0.000
SDF0Hz	0.63	0.14	0.420	4.35	0.000

R^2_a = Coeficiente de determinação ajustado

Vozes Outliers: G=78,98; R=34,73,90,91,99; B=1,4,13,73; A=3,55,73,94,1,34.

5.3. Praat

Variável independente	Coef. não normalizados		Coef. normalizados		t	valor-p		
	B	Erro Padrão						
			Beta					
G ($R^2_a = 0,70$)								
(Constante)	102.17	4.01			25.47	0.000		
HNRRdB	-2.64	0.20	-0.838		-13.48	0.000		
R ($R^2_a = 0,68$)								
(Constante)	90.86	6.80			13.36	0.000		
HNRRdB	-2.10	0.26	-0.580		-8.04	0.000		
SDF0Hz	0.39	0.09	0.298		4.18	0.000		
MeanF0Hz	-0.09	0.02	-0.234		-3.67	0.000		
B ($R^2_a = 0,71$)								
(Constante)	78.58	6.41			12.26	0.000		
HNRRdB	-3.51	0.25	-0.836		-14.15	0.000		
MeanF0Hz	0.16	0.03	0.372		6.30	0.000		
A ($R^2_a = 0,33$)								
(Constant)	94.55	11.60			8.15	0.000		
HNRRdB	-3.13	0.49	-0.875		-6.46	0.000		
jitterlocal	-17.28	4.55	-0.514		-3.80	0.000		
S ($R^2_a = 0,45$)								
(Constant)	13.53	1.71			7.93	0.000		
SDF0Hz	0.80	0.10	0.673		8.29	0.000		

R^2_a = Coeficiente de determinação ajustado

Vozes Outliers: G=4,5,40,57,78,90,91,94,96,98,99; R=3,47,58,71,73,90,91,99; B=4,5,40,56; A=55,79,99; S= 21,66,79,85,99

5.4. Voice Studio

Variável independente	Coef. não normalizados		Coef. normalizados	t	valor-p
	B	Erro Padrão	Beta		
G ($R^2_a = 0,55$)					
(Constante)	33.14	2.12		15.6	0.000
ShimmerAPQ5	2.85	0.66	0.459	4.3	0.000
jitterlocal	2.27	0.71	0.344	3.2	0.002
R ($R^2_a = 0,56$)					
(Constante)	37.79	6.08		6.22	0.000
Shimmerlocal	2.65	0.31	0.646	8.57	0.000
MeanF0Hz	-0.10	0.03	-0.269	-3.56	0.001
B ($R^2_a = 0,52$)					
(Constante)	-8.21	7.12		-1.15	0.252
ShimmerAPQ5	6.00	0.65	0.696	9.23	0.000
MeanF0Hz	0.14	0.03	0.321	4.27	0.000
A ($R^2_a = 0,26$)					
(Constante)	38.37	7.40		5.18	0.000
HNRRdB	-1.21	0.36	-0.427	-3.38	0.001
SDF0Hz	-0.66	0.16	-0.827	-4.23	0.000
jitterlocal	3.84	1.07	0.708	3.60	0.001
S ($R^2_a = 0,27$)					
(Constante)	12.93	1.96		6.59	0.000
SDF0Hz	0.37	0.07	0.526	5.67	0.000

R^2_a = Coeficiente de determinação ajustado

Vozes Outliers: G=73,78,94,98,99; R=71,73,88,90,91,98,99; B=34,55,94; A=55,94; T=66,79,96,99