

FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Programa Doutoral em Engenharia Biomédica

***Correlação entre a avaliação
acústica e perceptual na
caracterização de vozes patológicas***

Relatório do Estado-da-Arte

Aluno: Susana Vaz Freitas N^o pdb07004
Orientador: Prof. Dr. Aníbal João de Sousa Ferreira

Outubro/2010

FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Programa Doutoral em Engenharia Biomédica



***Correlação entre a avaliação
acústica e perceptual na
caracterização de vozes patológicas***

Relatório do Estado-da-Arte

Aluno: Susana Vaz Freitas N^o pdb07004
Orientador: Prof. Dr. Aníbal João de Sousa Ferreira

Outubro/2010

Este trabalho foi elaborado pela aluna Susana Alexandre Vaz dos Santos Freitas, número pdb07004, estudante na Faculdade de Engenharia da Universidade do Porto, do Programa Doutoral de Engenharia Biomédica (PRODEB). Foi realizado no âmbito do Relatório do Estado-da-Arte, do 2º ano do PRODEB, como parte integrante dos requisitos para a inscrição a título definitivo como aluna de doutoramento, sob orientação do Professor Dr. Aníbal João de Sousa Ferreira e da Professora Dr. Lúcia Figueiredo Mourão.

Grupo de Acompanhamento do Doutoramento:

- Professor Dr. Diamantino Freitas

- Professora Dr.^a Isabel Guimarães

ÍNDICE DO ESTADO-DA-ARTE

	p.p.
ÍNDICE DO ESTADO DA ARTE	i.
ÍNDICE DE IMAGENS	iii.
ÍNDICE DE TABELAS E QUADROS	iv.
SUMÁRIO	v.
PALAVRAS-CHAVE	vi.
Capítulo 0 – MOTIVAÇÕES E OBJECTIVOS	
0.1. Contextualização	1.
0.2. Motivações	4.
0.3. Objectivos	6.
0.4. Metodologia	7.
Capítulo 1 – A AVALIAÇÃO PERCEPTUAL E ACÚSTICA NA CARACTERIZAÇÃO DE VOZES PATOLÓGICAS	8.
1.1. A avaliação áudio-perceptual	10.
Escalas áudio-perceptuais	12.
O painel de juizes/avaliadores áudio-perceptuais	24.
1.2. Reprodutibilidade dos resultados da avaliação áudio-perceptual	31.
1.3. A avaliação acústica	44.
1.4. A avaliação e análise acústica da voz	52.
1.5. Parâmetros acústicos e implicações clínicas	59.
1.6. Reprodutibilidade dos resultados da avaliação acústica	78.
Capítulo 2 – CORRELAÇÃO ENTRE MEDIDAS DE AVALIAÇÃO ACÚSTICA E PERCEPTUAL	86.

Capítulo 3 – PLANO DE TRABALHOS DA DISSERTAÇÃO DE DOUTORAMENTO	
3.1. Participantes e métodos de recolha dos dados para a Tese	98.
3.2. Tratamento estatístico	104.
3.3 Cronograma	106.
Capítulo 4 – CONCLUSÕES	108.
BIBLIOGRAFIA	111.

ÍNDICE DE IMAGENS

	p.p.
Imagem 1: Aspecto do programa informático criado por Shaheen Awan (<i>in</i> Awan & Lawson, 2009) para avaliação vocal áudio-perceptual	11.
Imagem 2: CAPE-V <i>in</i> Protocolo de Avaliação da Qualidade Vocal da Universidade de Aveiro	21.
Imagem 3: 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)	45.
Imagem 4: Exemplo de Sinal Tipo 1	55.
Imagem 5: Exemplo de Sinal Tipo 2	56.
Imagem 6: Exemplo de Sinal Tipo 3	56.
Imagem 7: 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 é possível a medição objectiva e precisa de alguns parâmetros de perturbação extraídos directamente do sinal acústico	59.
Imagem 8: Espectrograma de um caso com o <i>software</i> Dr. Speech®	61.
Imagem 9: Espectro de banda larga	62.
Imagem 10: Espectro de banda estreita	62.
Imagem 11: Representação esquemática da mudança vocal na adolescência (Probst et al., 2004)	64.
Imagem 11: Sinal de onda complexa – representação isolada da onda sinusoidal e de ruído e a conjugação dos dois componentes	74.

ÍNDICE DE TABELAS E QUADROS

	p.p.
Tabela 1: Definição dos parâmetros da escala GIRBAS (Behlau, 2001; Pouchouli, 2008)	15.
Tabela 2: Definição dos parâmetros da <i>Stockholm Voice Evaluation Approach</i> (traduzido pela autora de Hammarberg, 2000)	20.
Tabela 3: Apresentação comparativa das vantagens e desvantagens das escalas perceptivas – GRBAS (e adaptações mais recentes), VPAS, SVEA, Bufallo III e CAPE-V.	22.
Tabela 4: Principais <i>softwares</i> de análise acústica de voz	47.
Quadro 1: Descrição sumária dos estudos sobre reprodutibilidade dos resultados da avaliação áudio-perceptual da voz	34.
Quadro 2: Descrição sumária dos estudos sobre reprodutibilidade dos resultados da avaliação acústica da voz	81.

SUMÁRIO

O presente trabalho de investigação foi desenvolvido no âmbito do estudo do estado da arte da temática “correlação entre a avaliação acústica e perceptual na caracterização de vozes patológicas”, no Programa Doutoral de Engenharia Biomédica da Faculdade de Engenharia da Universidade do Porto, sob a orientação do Professor Aníbal Ferreira.

Serviu como o momento integrador de conhecimentos adquiridos ao longo da prática clínica da doutoranda, com a pesquisa e análise bibliográfica revista para o resultado que a seguir se apresenta.

A voz, tal como as suas perturbações, é multifacetada. As queixas dos pacientes que procuram o clínico são o ponto de partida para todo o processo terapêutico. Geralmente a descrição mais habitual é “a minha voz está rouca”, e cabe aos profissionais destrinçar a etiologia do problema e a sua rápida e eficaz resolução. Ao longo deste percurso é necessária uma avaliação completa do caso, por forma a estabelecer o seu diagnóstico.

O diagnóstico vocal pode ser realizado tendo como base diferentes protocolos de avaliação, sendo que teve um forte impulso com o *Classification Manual for Voice Disorders I – CMDV-I* (Verdolini, Rosen & Branski, 2006) o qual prevê a relativa independência entre os aspectos vocais e o diagnóstico laríngeo. Esta obra assume-se como um esforço sistémico para organizar a literatura das variadas alterações que podem produzir problemas vocais e criar uma referência-padrão. Pretende, ainda, classificar os distúrbios vocais e facilitar a comunicação entre os profissionais. Cada uma das entradas do manual contém uma descrição da condição ou lesão, distinguindo as características essenciais associadas, com a descrição do impacto vocal produzido, aspectos demográficos dos pacientes, percurso clínico da alteração, factores desencadeantes, além de dados sobre o diagnóstico diferencial e critérios de classificação. Note-se que extrapola a classificação das alterações laríngeas (que

geralmente se confinam ao limite do órgão) e inclui as alterações vocais sistémicas ou extra-laríngeas (como as doenças neurológicas e psicológicas).

Normalmente, em contexto clínico, a avaliação da voz envolve a recolha de informações através de: entrevista (história clínica/*anamnese*); avaliação da fisiologia laríngea (laringoscopia indirecta e/ou endoscopia e/ou estroboscopia e/ou electroglotografia, ou outros); avaliação perceptiva; exame funcional (avaliação músculo-esquelética e aerodinâmica); análise acústica e auto-avaliação do impacto psicossocial da voz, tal como preconizado pelo Comitê de Foniatria da Sociedade Europeia de Laringologia (Dejonckere et al., 2001). Este trabalho e a Dissertação que se lhe seguirá, focar-se-ão na avaliação de alterações da voz do ponto de vista de um terapeuta da fala. Deste modo a avaliação da anatomofisiologia da laringe é excluída. Limitar-nos-emos à descrição, análise e utilização de medidas de caracterização da qualidade vocal pré-existent.

O relatório do estado-da-arte está estruturado da seguinte forma:

- A apresentação das motivações e objectivos da tese.
- Descrição e discussão dos conceitos inerentes ao tema a ser investigado – a avaliação acústica e perceptual na classificação de vozes patológicas – com a análise da reprodutibilidade dos resultados obtidos com o uso de cada uma.
- Contextualização do tema no domínio científico – estado da arte: congregam-se os dados de uma meta-análise sobre os trabalhos mais recentes de investigação acerca dos resultados da correlação entre diferentes medidas de avaliação acústica e formas de classificação áudio-perceptual da voz alterada.
- O plano de trabalhos para cumprir o objectivo da tese: descrição da metodologia de investigação a levar a cabo para a recolha e análise dos dados da tese.
- As conclusões sobre o estado da arte e reflexões acerca dos principais desafios a atingir no final do processo de construção da dissertação de doutoramento.

PALAVRAS-CHAVE

Avaliação áudio-perceptual, voz patológica, método acústico, correlação.

Capítulo 0 – MOTIVAÇÕES E OBJECTIVOS

0.1. Contextualização

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

O conceito de qualidade vocal está directamente relacionado com preceitos fisiológicos, perceptivos e acústicos (Behlau, 2001; Guimarães, 2007; Camargo e Madureira, 2010). A disфонia ou patologia vocal é definida como um distúrbio da comunicação oral, no qual a voz não cumpre a transmissão da mensagem verbal e emocional (Behlau, 2001), representando uma dificuldade ou alteração na emissão da voz, impedindo a sua produção (Behlau & Pontes, 1995 cit in 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 multifactoriais que permitam um conhecimento amplo, adequado e eficaz da função laríngea e da qualidade vocal (Behrman, 2004; Guimarães, 2007; Speyer, 2008 *in* Behlau, 2010).

A elaboração deste trabalho torna-se pertinente na medida em que evidencia a relevância do processo de avaliação na voz patológica, 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-perceptual e o método acústico, e as correlações existentes entre eles. As correlações entre a avaliação perceptivo-auditiva e as medidas objectivas 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, foram apontadas limitações ao estabelecimento de relações directas entre estas duas dimensões da avaliação da voz patológica (Maryn et al., 2009; Camargo & Madureira, 2010).

A avaliação áudio-perceptual é considerada *gold-standard* por alguns investigadores, principalmente aqueles que a usam na rotina clínica e a vêem como soberana em relação às outras (Behram, 2005; Bunton et al., 2007; Kreiman & Gerratt, 2007; Patel & Shrivastav, 2007; Oates, 2009), embora a consistência intra e inter-avaliadores seja entre ligeira-moderada (Kreiman et al., 1990; 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 (Kreiman et al., 1993; Rabinov et al., 1995 in Eadie & Doyle, 2005; 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), da instrução da tarefa (Awan & Lawson, 2009) e experiência do avaliador (Kreiman, 1992; Kreiman et al., 1993; Behlau, 2001; Patel & Shrivastav, 2007).

A análise acústica tradicional 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 citadas 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 (Dajer, 2010). Nas últimas duas décadas a dinâmica não linear tem introduzido uma nova perspectiva para a análise de sistemas que evoluem no tempo e são sensíveis às condições iniciais, ou sejam, embora determinísticos apresentam comportamentos aperiódicos. 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; Tetze et al., 1993; Herzel et al., 1994 in Dajer, 2010).

Refira-se então que – actualmente – 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), ao facilitarem a tarefa de análise e classificação de uma (alteração da) voz, ao longo das várias fases do processo terapêutico (Martens et al., 2007; Awan & Lawson, 2009).

0.2. Motivações

As razões que justificam a realização deste trabalho prendem-se com dificuldades práticas percebidas pela aluna 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 – perceptual e acústica – das alterações da voz que existem? E quais as mais eficazes?
- Que alterações vocais permitem uma descrição fiável com recurso ao método acústico e ao perceptual?
- Quais os parâmetros da avaliação acústica que se associam e corroboram a caracterização áudio-perceptual?

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-perceptual) na avaliação das vozes patológicas. Isto porque, isoladamente, cada uma destas formas de classificação do grau de disfonia é francamente pobre (Behlau, 2010). Por outro lado ainda 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 perceptuais comumente usadas, devido aos índices moderados de consistência no que concerne à reprodutibilidade dos mesmos.

Devido à sua natureza, o trabalho assume ainda uma motivação pedagógica com suporte tecnológico. Isto porque se pretende criar uma base de dados de vozes patológicas, devidamente analisadas acusticamente (por mais do que um software que tenham os mesmos parâmetros acústicos) e classificadas perceptualmente por avaliadores devidamente treinados nesse campo, que se colocará à disposição de

futuros e actuais terapeutas que, deste modo, terão à sua disposição mais um material de treino e preparação para uma mais eficaz avaliação (Awan & Lawson, 2009) e intervenção terapêutica com casos de patologia vocal.

Assim sendo, procurar-se-á seleccionar um grupo ilustrativo de vozes patológicas portuguesas para criar uma base de dados de referência (“âncoras” auditivas) para a caracterização do tipo e severidade de cada parâmetro perceptual – que será cotado por uma amostra significativa e válida de juízes – complementada, também, pelo estudo da consistência inter-avaliações de diferentes parâmetros acústicos, levadas a cabo por distintos softwares *opensource* e comercializados.

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 que, por 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 patológicas (Frohlich et al., 1997; Finizia et al., 1999; Parsa & Jamieson, 2001; Maryn et al., 2009). Contudo, de acordo com a literatiura, a sua correlação com a classificação perceptual varia de razoável a forte para alguns parâmetros (Giovanni et al., 1996; Piccirillo et al., 1998; Wuyts et al., 2000; Butha et al., 2004; Eadie & Baylor, 2006; Ma & Yiu, 2006). Há também a necessidade urgente – e prévia – de enriquecimento e corroboração quantitativa dos dados perceptivos, usados amplamente pelos terapeutas da fala, também eles de reprodutibilidade variável intra e inter-avaliadores.

Para tal, e por último, prevê-se a possibilidade de criar uma medida de quantificação/classificação da qualidade vocal que conjugue os parâmetros da escala GIRBASH – usada ao longo dos trabalhos da dissertação de doutoramento – e os valores estatisticamente mais relevantes da análise acústica levada a cabo.

0.3. Objectivos

O objectivo principal deste trabalho é conhecer o estado-da-arte das temáticas inerentes ao tema da tese de doutoramento, nomeadamente as áreas da engenharia de processamento de sinal e da terapêutica da fala.

Assim, pretende-se ter uma noção actual das áreas de conhecimento relacionadas como 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 o perceptual.

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

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

1. Recolher, caracterizar e analisar a consistência inter-avaliadores e correlação entre os valores da cotação perceptiva de 50 vozes – levada a cabo por 30 alunos de Terapia da Fala/Fonoaudiologia e 30 Terapeutas da Fala/Fonoaudiólogos – de uma base de dados de 36 vozes patológicas.
2. Realizar, descrever e analisar a consistência e correlação entre as avaliações através do método acústico – fornecidas por um software opensource (PRAAT) e três comercializados (Multi Dimensional Voice Program, da Kay Elementrics; VoiceStudio, da Seegnal; Voxmetria, da CTS Informática) – de parâmetros acústicos comuns entre eles, a partir da base de dados de vozes patológicas enunciada no ponto anterior.
3. Levar a cabo, descrever e discutir os resultados da correlação entre a avaliação acústica e perceptual da base de dados de vozes patológicas.

0.4. Metodologia

No que concerne ao estudo do estado-da-arte do âmbito do trabalho, 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 voz patológica.

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 2009/2010, 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 discutidas de forma construtiva, com profissionais de renome, os problemas actuais da avaliação e diagnóstico da patologia vocal.

A construção da base de dados de vozes patológicas – referência para diferentes tipos e graus de severidade de cada parâmetro de perturbação da voz – é da responsabilidade da doutoranda, 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. As mesmas serão analisadas e classificadas perceptualmente por um painel de juizes (4 peritos e, mais tarde, por 30 alunos de Terapia da Fala/Fonoaudiologia e, também, por 30 Fonoaudiólogos/Terapeutas da Fala, com diferentes graus de mestria no método áudio-perceptual). As amostras serão classificadas acusticamente por softwares *open source* e *freeware* vs comercializados. Todo o trabalho culminará com a descrição e interpretação das correlações encontradas entre a avaliação acústica e perceptual das vozes. Ainda se prevê, por último, a construção de um software didáctico com a base de dados de vozes de referência caracterizada ao longo de toda a dissertação.

Capítulo 1 – A AVALIAÇÃO PERCEPTUAL E ACÚSTICA NA CARACTERIZAÇÃO DE VOZES PATOLÓGICAS

A avaliação da voz humana implica uma complexa combinação de dados que permitem a construção de um quadro conceptual que possibilita e/ou suporta a explicação do “como” e “porquê” da perturbação vocal apresentada.

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), 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/projecção 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, a sensação de intensidade e/ou a qualidade vocal são desagradáveis ou inadequadas para a idade, sexo do indivíduo do ponto de vista social e/ou familiar; o indivíduo refere desconforto ou dor ao falar; causa alguma preocupação ao próprio que necessita de ser resolvida (Verdolini, 1994; Guimarães, 2007; Tavares & Martins, 2007).

As informações obtidas sobre a alteração da voz ao longo das diversas provas 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) seleccionar de modo adequado os métodos e técnicas de intervenção; (d) determinar a eficácia da Terapia implementada (Cardig e Horsley, 1992; Ramig & Verdolini, 1998 *in* Freeman & Fawcus, 2004).

Em contexto clínico – e de acordo com o Comitê de Foniatria da Sociedade Europeia de Laringologia (ELS) – a avaliação da voz implica a recolha de informações através de: (1) anamnese ou entrevista clínica; (2) avaliação da fisiologia laríngea (laringoscopia indirecta e/ou endoscopia e/ou estroboscopia e/ou electroglotografia, entre outros); (3) avaliação áudio-perceptual; (4) exame funcional (avaliação musculo-esquelética e aerodinâmica); (5) análise acústica; e, (6) auto-avaliação 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, requisitos mínimos necessários para fazer meta-análises comparativas no tratamento da voz e medidas básicas de medição nas patologias vocais (Brasolotto & Rehder, 2011).

Da revisão da literatura sobre a temática da avaliação da voz patológica verificamos que existe um debate acérrimo sobre o mérito da mesma, com linhas de investigação que visam dar maior segurança e confiabilidade à análise perceptivo-auditiva, com vista à integração dos resultados perceptivos e acústicos, em especial para as vozes disfónicas (Awan & Roy, 2009; Maryn et al., 2009; Brasolotto & Rehder, 2011). De seguida serão descritos os principais procedimentos de recolha, registo e análise dos dados obtidos através da avaliação áudio-perceptual e acústica da voz humana, considerados também na literatura como a análise perceptiva e a objectiva, respectivamente (Pouchoulin, 2008).

1.1. A AVALIAÇÃO ÁUDIO-PERCEPTUAL

A avaliação áudio-perceptual pressupõe que o profissional julgue uma amostra vocal produzida pelo falante, que refere (ou não) queixas de utilização da voz. 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-perceptual da qualidade vocal (Freeman e Fawcus, 2004; Pouchoulin, 2008; Oates, 2009; Brasolotto & Rehder, 2011), dos quais se destacam a:

1. Baixa confiabilidade intra e inter-avaliadores;
2. Não disponibilização de mensurações objectivas;
3. Não existência de uma escala universal de avaliação perceptual.

A revisão da literatura demonstra que a avaliação perceptual faz depender a sua validade de vários factores (Sederholm *et al.*, 1992; Eadie & Baylor, 2006; Ma & Yiu, 2006; Guimarães, 2007; Awan & Lawson, 2009; Awan & Roy, 2009; Brasolotto & Rehder, 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 e a existência de parâmetros vocais externos (por exemplo, fenómenos de co-articulação, características supra-segmentais) que funcionem como ajudas (“âncoras”) ao ouvinte. 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 in Finizia *et al.*, 1999; Yu *et al.* 2001; Dedivitis *et al.*, 2004; Bele, 2005; Eadie *et al.* 2005; Ma & Yiu 2006; Martens *et al.*, 2007; Awan & Lawson, 2009). Awan & Lawson (2009) realizaram uma investigação em que 40 alunos de terapia da fala (sem experiência de avaliação perceptual) 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. Imagem 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 inter-avaliadores (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.

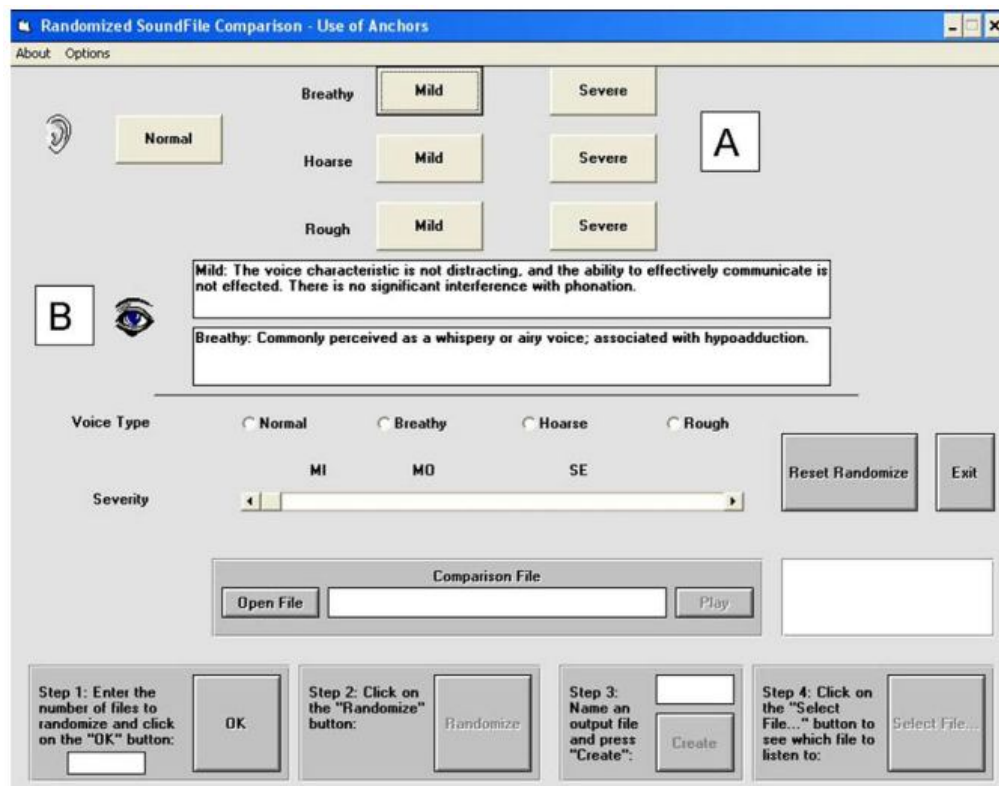


Imagem 1: Aspecto do programa informático criado por Shaheen Awan (in Awan & Lawson, 2009) para avaliação vocal áudio-perceptual.

Para alguns autores a avaliação perceptiva da voz patológica é uma componente central do processo de caracterização das disfonias (Behlau, 2001; Araújo et al., 2002; Cummings, 2008) e, de longe, aquela que é mais usada para descrever a voz em contexto clínico (Hammarberg, 2000): atendendo à sua rapidez de uso, eficiência dos resultados obtidos e ao facto de implicar poucos recursos materiais (custo reduzido) (Orlikoff et al., 1999; Schiavetti & Metz, 2002; Awan & Lawson, 2009).

Esta recolha e análise 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, articulatorio e de ressonância). É um processo integrado que consiste, sucintamente, em ouvir e descrever uma voz, caracterizando-a em termos globais ou com referência a parâmetros específicos, passíveis de associação a características psico-acústicas e fisiopatológicas (Carding et al., 2000; Mathieson & Greene, 2001). Existem inúmeras escalas e protocolos de avaliação áudio-perceptual, no entanto descreveremos com mais pormenor as mais referidas nos artigos científicos consultados para a elaboração deste estado-da-arte.

ESCALAS AUDIO-PERCEPTUAIS: GRBAS (*Grade, Roughness, Breathiness, Asteny, Strain*) (Hirano, 1981), GIRBAS (*Grade, Instability, Roughness, Breathiness, Asteny, Strain*) (Dejonckere, Remacle & Fresnel-Elbaz, 1996), RASAT (*Rouquidão, Aspereza, Sorposidade, Astenia, Tensão*) (Pinho, 2002), RASATI (*Rouquidão, Aspereza, Sorposidade, Astenia, Tensão, Instabilidade*) (Pinho & Pontes, 2008), GRBASH (*Grade, Roughness, Breathiness, Asteny, Strain, Harshness*) (Nemr & Lehn, 2010), VPAS (*Vocal Profile Analysis Scheme*) (Laver, 1980), BVP (*Buffako III – Voice Screening Profile*) (Wilson, 1987), SVEA (*Stockolm Voice Evaluation Approach*) (Hammarberg, 1986, 2000), e o CAPE-V (*Consensus Auditory Perceptual Evaluation of Voice*) (ASHA, 2002 in Kempster et al., 2009).

A análise perceptiva da voz patológica tem como essência o princípio de que voz e audição são indissociáveis (Pouchouli, 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 perceptivos que descrevem a qualidade da voz (Awan & Roy, 2009), o que condiciona o seu uso para efeitos médico-legais. Esta é a vertente qualitativa das escalas perceptivas. 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 diverge entre 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 e tal, 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). Kreiman et al. (1993) consideraram que a escala com 7 níveis é a mais apropriada para avaliar vozes patológicas 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. 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 com a 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 inter-avaliadores. Esta também reflectiu de forma mais eficaz as diferenças de interpretação entre os parâmetros da escala. 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 correlação entre a classificação da GRBAS através da VAS se correlacionava de forma mais forte com as medidas de avaliação acústica,

que as classificações derivadas de níveis de EAI (88% vs 64% de concordância). A investigação de Yamasaki et al. (2008) estabeleceu uma correspondência entre a VAS e a EAI de 4 pontos na avaliação áudio-perceptual 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 ROC. 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 compreendeu 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 4 categorias não obedecem à partição matemática de 25mm iguais (Madazio & Behlau, 2009).

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.

A escala mais conhecida e usada mundialmente é a GRBAS, de Hirano (1981) (Hammarberg, 2000; Cummings, 2008; Wirz & Beck, 1995 *in* Pouchouli, 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. (Pinho, 2008). Cada um dos cinco grafemas é o acrónimo de G – avaliação global da disfonia (*grade*); R – aspereza/rugosidade (*roughness*); B – soprosidade (*breathiness*); A – astenia (*asteny*); S – tensão (*strain*) (Hirano, 1981 *in* Guimarães, 2007). É uma escala compacta e de simples utilização, eficaz para a triagem vocal, 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; Pouchouli, 2008). Os parâmetros avaliados são classificados numa escala de 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.

Em 1996 Dejonckere, Remacle & Fresnel-Elbaz (*in* Behlau, 2001) propuseram acrescentar o factor Instabilidade (I) à escala GRBAS original. Os factores desta escala estão caracterizados na Tabela 1.

Parâmetros	Definição
<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
<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: Definição dos parâmetros da escala GIBAS (Behlau, 2001; Pouchouli, 2008)

Existem críticas ao uso destas escalas, nomeadamente porque não contemplam a classificação da adequação da frequência fundamental e outras características da voz. Estudos identificaram ainda a baixa confiabilidade dos parâmetros *Asteny*-Astenia e *Strain*-Tensão (Dejonckere et al., 2001). Contudo, este é um método que se correlaciona com outras formas de classificação vocal (nomeadamente os questionário de avaliação do impacto da disfonia na qualidade de vida – Jones et al., 2006 e Karnell et al. 2007, *in* Cummings, 2008).

Em 2002 Pinho (*in* Pinho, 2003) sugere a adopção, no Brasil, da escala RASAT a partir da GRBAS. As siglas indicadas correspondem, do ponto de vista anatomofisiológico e perceptual, 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 percebida com ruídos imprevistos produzidos a baixa frequência (Hirano et al, 1990; Hammarberg e Gauffin, 1995; Hammarberg, 2000 *in* Pinho, 2008) que mascaram os harmónicos inferiores no traçado espectrográfico, ou surgem sub-harmónicos (Martens et al., 2007). Este parâmetro verifica-se em casos de: fenda glótica maior ou igual a $0,5\text{mm}^2$, presença isolada de uma alteração orgânica ou fenda de qualquer dimensão com alterações da mucosa vibrante (Isshiki, 1980 *in* Pinho, 2003) das pregas vocais (por exemplo: nódulos, hiperemias ou edemas) (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 Reinke). A voz é seca, sem projecção, com ruídos nas altas frequências pela diminuição da onda mucosa (Hirano et al, 1990; Hammarberg e Gauffin, 1995; Hammarberg, 2000 *in* Pinho, 2008) que implica um maior dispêndio de energia e conseqüente gasto de ar para desencadear a vibração de uma mucosa rígida – notando-se a evidência dos harmónicos inferiores. A aspereza é percebida nas vozes produzidas com esforço excessivo, tensão laríngea e constrição do tracto vocal (Oliveira, 2010). O exemplo clássico de rigidez é a voz dos casos com sulco vocal, quistos, retracções cicatriciais e lesões neoplásicas.

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, papilomas e edemas, e em

pacientes com doenças neurológicas degenerativas (Stemple et al, 1995). O ruído é o parâmetro mais presente nos casos com disфония (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 controle vocal (Behlau, 2001).

Tensão: associada a esforço vocal por aumento da adução glótica (hiperfunção), geralmente inerente ao aumento da actividade da musculatura extrínseca da laringe, com elevação da posição desta (Behlau, 2001). São evidenciados harmónicos em altas frequências (Martens et al., 2007). Exemplo: disфония espasmódica em adução e síndromes de abuso vocal (síndrome de tensão musculoesquelética), com conseqüente 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 disфония de rouquidão moderada, aspereza moderada-severa, sopro leve, sem astenia e com tensão leve-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 confiabilidade (Pinho et al., 2008).

Em 2010 foi apresentada a GRBASH (*Grade, Roughness, Breathiness, Asteny, 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 respectiva 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.

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). Pressupõe a descrição fonética da qualidade vocal partindo da referência a uma posição articulatória neutra, definida pela fisio-acústica, na análise de provas de leitura ou fala espontânea. Assim, possibilita a descrição de características laríngeas e supra-glóticas (tracto vocal) relativamente a 31 parâmetros distribuídos por 3 sub-categorias específicas:

- A qualidade vocal: as qualidades laríngeas e supra-laríngeas 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 cadência.

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).

Outra escala comumente usada na avaliação perceptual da voz é a *Buffalo III – Voice Screening Profile* (Wilson, 1987 in Cummings, 2008). Esta classifica numa escala de cinco pontos (em que 1=normal e 5=muito grave) os seguintes parâmetros: tónus 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 dos últimos vinte anos do Séc. XX o departamento de Terapia da Fala do Hospital de Huddinge (Suécia) desenvolveu e aperfeiçoou a *Stockholm Voice Evaluation Approach* (SVEA) (Hammarberg, 1986; Hammarberg et al., 1995, in Hammarberg, 2000). Esta escala partiu da análise de correlações entre 28 variáveis (baseada em 50 termos perceptuais usados pelos clínicos), de que resultaram 13 parâmetros qualitativos propostos para a avaliação perceptiva de vozes normais e patológicas (ver Tabela 2), em 5 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 perceptual 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 cordas vocais e da laringe durante a fonação com sopro insuficiente
Hipofuncional/Laxo	É o oposto da hiperfuncional. Existe tensão vocal insuficiente e actividade 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 falseto

Diplofonia	Dois <i>pitches</i> diferentes podem ser percebidos ao mesmo tempo
Registo Modal/Falseto	Modos de fonação: o registo modal é o modo fonatório normal, standard, com encerramento glótico durante a fonação; em oposição com o falseto, que usualmente está no topo do alcance de frequência fundamental, como resultado do alongamento e afinamento das pregas vocais, com encerramento insuficiente
<i>Pitch</i>	A correlação auditiva com a frequência fundamental
<i>Loudness</i>	A correlação auditiva com nível de pressão sonora da fala

Tabela 2: Definição dos parâmetros da *Stockholm Voice Evaluation Approach* (traduzido pela autora de 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 perceptual suportado por computador (*VISOR – Visual Sort and Rate*) que provou ser um método eficaz para a classificação numa escala analógica visual das cotações perceptivas de vozes patológicas (Granqvist et al., 2003).

Por último, o CAPE-V (*Consensus Auditory Perceptual Evaluation of Voice*) é uma escala de avaliação perceptiva 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. Imagem 2).

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 vs 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).

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 Maria e o avô vivem naquele casarão rosa velho.*
 - b. *Sofia saiu cedo da sala.*
 - c. *A asa do avião andava variada.*
 - 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, aspereza, falseto, astenia, afonia, bitonalidade, tremor, estridente, oglottal fry), entre outros aspectos relevantes) _____

Escala G.R.B.A.S. (G: grau de rouquidão; R: aspereza; B: soprosidade; A: astenia; S: tensão)
 G __ R __ B __ A __ S __ (0=normal; 1=alterações ligeiras; 2=alterações moderadas; 3=alterações severas)

Imagem 2: CAPE-V in Protocolo de Avaliação da Qualidade Vocal da Universidade de Aveiro. Página 21 de 132

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

ESCALA	Vantagens	Desvantagens
GRBAS/ GIRBAS/ RASAT/ RASATI/ GRBASH	<ol style="list-style-type: none"> 1. Baseado em provas clínicas simples (fala espontânea e/ou vogal 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 supra-glóticos) 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 tracto vocal (31 parâmetros) 2. Classificação final em seis níveis 3. Caracteriza factores vocais individuais 4. Adequado a vozes normais e patológicas 5. Relaciona com a função fisiológica 6. Implica programa de treino de 2 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, supra-glóticas 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 supra-glóticas 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 perceptivas – GRBAS (e adaptações mais recentes), VPAS, SVEA, Buffalo III e CAPE-V.

Até à data não existe uma escala universal, pelo que para a realização da Dissertação de Doutoramento a autora seleccionou para estudo os parâmetros GRB (*Grade, Roughness e Breathiness*) para classificar as vozes-alvo usadas ao longo da elaboração da tese (Dejonckere et al., 2001). Cada parâmetro obedece à seguinte descrição:

Parâmetros	Definição
<i>G – Grau (Grade)</i>	Grau de alteração vocal – impressão global da voz, identificação do grau de alteração vocal, como um todo.
<i>I – Instabilidade (Instability)</i>	Instabilidade, flutuação na F0 e/ou na qualidade vocal avaliada.
<i>R – Rugosidade (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. A voz é percebida com ruídos imprevistos produzidos a baixa frequência.
<i>B – Soprosidade (Breathiness)</i>	Presença de turbulência ou ruído audível, escape de ar importante através da glote, sensação de ar na voz.
<i>A –Astenia (Asteny)</i>	Hipofunção, fraqueza vocal, falta/perda de potência na voz relacionada com intensidade fraca, energia vocal reduzida.
<i>S –Tensão (Strain)</i>	Hiperfunção, impressão de estado hiperfuncional associado a esforço vocal, frequência anormalmente aguda, geralmente por aumento da actividade da musculatura intrínseca e elevação da laringe.
<i>H – Aspereza (Harshness)</i>	A voz é seca, sem projecção, com ruídos nas altas frequências pela diminuição da onda mucosa que implica um maior dispêndio de energia e conseqüente gasto de ar para desencadear a vibração de uma mucosa rígida. A aspereza é percebida nas vozes produzidas com esforço excessivo, tensão laríngea e constrição do tracto vocal.

O PAINEL DE JUÍZES/AVALIADORES AUDIO-PERCEPTUAIS:

Na clínica quotidiana espera-se que qualquer protocolo de avaliação seja simples, reprodutível e prático. Os resultados da avaliação perceptual 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).

Porém, existirá sempre um teor subjectivo (nem que seja inconscientemente) das classificações realizadas. Isto porque aquelas não correspondem directamente 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 indirecta 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 perceptivas está sujeito a erros e variabilidade porque: 1) as escalas usadas no contexto clínico e de investigação, por vezes, são impróprias para medir os atributos da qualidade vocal; 2) os avaliadores nem sempre têm o mesmo entendimento sobre os parâmetros a avaliar; 3) os avaliadores nem sempre conseguem discriminar uma só dimensão da escala num estímulo sonoro complexo; 4) os avaliadores têm dificuldade em manter a consistência das classificações, intra e inter-juízes.

O modelo matemático que modeliza a variabilidade da avaliação perceptual 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. Este autor ainda postula que este processo segue uma distribuição normal, logo, é possível estabelecer modelos matemáticos para

caracterizar a resposta dos juizes aos estímulos vocais. Esta forma de descrição da avaliação áudio-perceptual pode ser também considerada uma consequência directa 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 juizes para a avaliação perceptiva visa reunir um conjunto próximo de opiniões, que reduzam a subjectividade (i.e. os erros de medida) deste tipo de avaliação, através do controle dos factores associados ao avaliador; à tarefa de avaliação; ou características resultantes da interacção dos dois parâmetros anteriores (Eadie & Baylor, 2006). Assim, os seus elementos podem ser de profissões distintas, que audio-perceptualmente – escutam e discriminam – 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 reprodutíveis acerca da qualidade vocal em análise (Kreiman et al., 1990, 1992, 1993; de Krom, 1994; De Bodt et al., 1997; Eadie & Baylor, 2006).

ERROS DE MEDIDA POR FACTORES ASSOCIADOS AO AVALIADOR

Estes factores incluem os standards internos do avaliador relativamente à qualidade vocal a ser percebida, aos seus hábitos perceptivos 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 ajuda a modelar estes factores (Eadie & Baylor, 2006). Nesta categoria estão ainda incluídos o cansaço do avaliador, os lapsos de atenção e os erros de transcrição, assumidos como erros aleatórios.

ERROS DE MEDIDA POR FACTORES 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 perceptual. Incluem-se ainda o contexto perceptivo, pois pode alterar o standard interno 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 o seu standard interno.

ERROS DE MEDIDA POR FACTORES ASSOCIADOS À INTERACÇÃO 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 (p.ex. Aspreza) possa ser mais fácil, por ter contactado com mais casos com essa característica evidente. Contudo, podem não estar tão sensíveis a outros parâmetros.

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 inter-avaliadores (Kreiman et al., 1993).

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).

Deste modo, a constituição de um painel de juízes pode depender da experiência dos avaliadores (Kreiman et al., 1990; Anders et al., 1988 in de Krom, 1994), 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 na avaliação de vozes pode ter pressuposto o desenvolvimento de um referencial prévio e crescente de 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 Pouchouli, 2008).

Existem algumas investigações que reflectem 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; Patel & Shrivastav, 2007). Note-se que há estudos que apontam resultados contraditórios e ratificam que a inconsistência inter-avaliadores é muitas vezes o resultado dos métodos usados para a recolha da avaliação e não inerente a problemas nas capacidades perceptivas e inconsistência entre os juízes (Kreiman & Gerratt, 2000; Kreiman et al., 2007; Patel & Shrivastav, 2007; Kreiman & Gerratt, 2010). 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 vs com treino em avaliação perceptual. 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 ($r=0,922$), o que aumentava ligeiramente com duas horas de treino perceptual ($r=0.938$ e $r=0.961$, para as tarefas indicadas, respectivamente). Dejonckere e tal. (1993) demonstraram que a escala GRBAS obtinha adequados valores de consistência intra e inter-juízes, 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 perceptual. 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 corda vocal. Defrontaram-se com valores mais consistentes da correlação bivariada de Pearson para os parâmetros Grau ($r=.828$, $p=.003$), Rugosidade ($r=.718$, $p=.019$) e Soprosidade ($r=.939$, $p<.001$). 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 ($p < .05$), R-rugosidade ($p < .001$) e B-soprosidade ($p < .001$).

Devido à fraca reprodutibilidade dos parâmetros instabilidade, tensão e astenia em diversos estudos enumerados, estes três parâmetros foram retirados do protocolo de avaliação da função vocal da European Laryngological Society (Dejonckere et al., 2001).

Kreiman et al. (1990) demonstraram que os avaliadores com e sem experiência/treino usam estratégias de classificação distintas. Neste estudo, o grupo de peritos apresentou menor acordo sobre os parâmetros de qualidade vocal. Hammarberg (1986) concluiu que os juizes 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 disártrica (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 juizes. 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.

Pouchouli (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 actuam 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 factores intrínsecos e/ou extrínsecos ao avaliador.

Acrescentem-se, então, factores 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 rugosidade, aspereza e sopro em vozes patológicas. Concluíram que houve uma melhoria após treino, contudo 40% das amostras ainda eram classificadas incorrectamente. 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 inter-avaliadores, 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 juizes 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 sopro 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 juizes apenas beneficiou de uma sessão de 20 minutos para apresentação do estudo, material de classificação perceptual (software informático específico) e definição de conceitos (G=grau geral de severidade da alteração vocal). Tinham acesso a um “estímulo-âncora” auditivo e obtiveram um forte grau médio de correlação inter-avaliadores (r de Pearson=0.83, S.D.=0.02) e de acordo inter-avaliadores (média de acordo=81.03%, S.D.= 2.74%).

Os efeitos do treino podem ainda afectar as pistas acústicas e visuais usadas pelos avaliadores para realizarem o treino perceptual (Awan & Lawson, 2009). Martin & Wolfe (1996) concluíram que os ouvintes classificam as vozes patológicas principalmente com base na frequência fundamental (F_0) antes do treino, contudo esta tarefa foi também influenciada após o treino pelas medidas de ruído (HNR ou SNR).

A análise áudio-perceptual 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). Do mesmo modo, as discordâncias podem ser o reflexo das diferenças de treino e experiência profissional (Wolfe & Martin, 1990; Kreiman et al., 1990; Bele, 2005). O sistema de referência interno dos avaliadores é idiossincrático, e varia intra e inter-juízes, de acordo com a experiência prévia e o contexto acústico no qual as análises são levadas a cabo (Kreiman et al., 2007).

Em suma, a análise perceptiva beneficia em termos de reprodutibilidade e consistência inter-avaliadores com a constituição de grupos de classificação que sejam especialistas nesta tarefa, realizando-a em múltiplas sessões de escuta, com recurso a estímulos-“âncora” ou a métodos de emparelhamento comparativo e feedback auditivo, como forma de re-calibração frequente do standard interno dos avaliadores (Chan e Yiu, 2006; Eadie & Baylor, 2006; Awan & Lawson, 2009; Brasolotto & Rehder, 2011). Porém, a sua aplicabilidade – devido a limitações financeiras, de disponibilidade, entre outras – é muitas vezes condicionada. Reforça-se ainda a necessidade de um planeamento metodológico (experimental) e estatístico adequado das investigações sobre avaliação perceptual da voz, com várias avaliações de cada estímulo para cada ouvinte e escalas padronizadas de avaliação (Brasolotto & Rehder, 2011), por forma a conseguir um valor de consistência inter-avaliadores elevada.

1.2. REPRODUTIBILIDADE DOS RESULTADOS DA AVALIAÇÃO AUDIO-PERCEPTUAL

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

- A variabilidade inter-observador, isto é, a variação de julgamentos entre diferentes avaliadores/terapeutas.
- A variabilidade intra-observador, ou seja, a variação de julgamentos de um mesmo avaliador/terapeuta ao longo do tempo.
- A variabilidade intra-sujeitos, entendida como a variação da qualidade vocal do falante em diferentes momentos temporais da recolha e análise da mesma. Este aspecto pode estar associado a factores intrínsecos (emocionais, cansaço) ou extrínsecos, que se não forem controlados condicionam os resultados da avaliação áudio-perceptual.

Este sub-capítulo resulta de uma pesquisa na Pubmed que possibilitou a revisão da literatura publicada (1950-Agosto 2009) sobre os resultados da avaliação áudio-perceptual – formas de uso e respectiva validade dos resultados (*cf.* Quadro 1). 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 estendidos, 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 1 de Agosto/2009
- Relativos a estudos com a qualidade vocal de humanos (objectiva e subjectiva)
- 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: 51

Total de artigos sobre avaliação perceptual: 39 (Quadro 1)

Total de artigos sobre avaliação acústica: 12

No total foram encontradas 39 referências publicadas ao longo de 25 anos (1986-2010) sobre a validade da avaliação áudio-perceptual, enquadráveis nos critérios de inclusão da pesquisa levada a cabo. Note-se que em nenhum dos artigos foi caracterizada a variabilidade intra-sujeito. Todos faziam referência à validade inter-observador e em 32 focavam-se os resultados da validade intra-observador. 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) pré-seleccionadas 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 eles (Granqvist, 2003; Kreiman et al., 1994; Shrivastav, 2006; Wolfe et al., 2000), e numa das investigações as comparações foram realizadas entre duas amostras do mesmo sujeito (Webb et al., 1992).

Foram usadas várias escalas, sendo as mais referenciadas previamente descritas ao longo deste trabalho. A GRBAS está presente em 24 dos artigos. A classificação desta através de escalas categóricas é usada em 19 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). É feita referência à escala analógica visual em 10 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).

Em 19 das investigações analisadas todos os casos possuíam uma alteração vocal; em 17 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 sistetizadas (Titze et al., 2003; Bergan et al., 2004).

Existe uma variabilidade marcada entre os instrumentos escolhidos (encontramos 11 escalas diferentes, no total). Em alguns casos foi desenvolvido e implementado um novo protocolo de avaliação áudio-perceptual, ajustado a clusters específicos de pacientes (por exemplo, a disfonia espasmódica ou a disartria) (Langeveld et al., 2000; Erickson, 2003).

Em síntese, para além da variabilidade inter-observador, intra-observador e intra-sujeito, a reprodutibilidade da avaliação perceptual pode ainda ser influenciada pelo sistema de avaliação usado, a forma de cotação adoptada e a amostra vocal usada (leitura, fala espontânea ou vogal sustentada).

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.

A maioria dos estudos que usa avaliação perceptiva possui validade inter-observador e intra-observador, de moderada-bom. Por vezes são feitas tentativas de aumento destes valores através da optimização dos sistemas de cotação. Uma vez que cada observador possui o seu *gold standard* interno de adequação da qualidade vocal, é compreensível que os investigadores prevejam 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. A variabilidade intra-sujeito nunca foi descrita em nenhum dos artigos consultados.

Quadro 1: Descrição sumária dos estudos sobre reprodutibilidade dos resultados da avaliação áudio-perceptual da voz.

Referência Bibliográfica	Parâmetro	Cotação	Intra-sujeito	Intra-observador	Inter-Observador	Análise do Artigo/Conclusões
ESTUDOS COMPARATIVOS						
Chan, K. M.; & Yiu, E.M. (2002) – <i>The Effect of anchors and training on the reliability of perceptual voice evaluation.</i>	Rugosidade Soprosidade	VAS	-	+	+	O uso de uma amostra de referência e o treino dos avaliadores mostrou que os standards internos para vozes patológicas não é estável e que tanto o treino como proporcionar referências são requisitos para a sua estabilização. Este estudo também revela que os sinais sintetizados também são boas referências.
Eadie, T.L. & Doyle, P.C. (2002) – <i>Direct magnitude estimation and interval scaling of naturalness and severity in tracheoesophageal (TE) speakers.</i>	Naturalidade Severidade	9 níveis DME	-	+	+	Este estudo sugere que a naturalidade e severidade global, se avaliadas adequadamente, são medidas clínicas válidas para vozes traqueoesofágicas.
Granqvist, S. (2003) – <i>The visual sort and rate method for perceptual evaluation in listening tests.</i>	Soprosidade Rugosidade Componentes em Alta-Frequência	VAS VISOR	-	+	+	O VISOR gerou um aumento da consistência das avaliações denotada em testes não paramétricos: coeficientes de Pearson e Spearman significativamente mais altos com este método. Com testes paramétricos, testes intra-sujeitos, obteve-se um coeficiente de correlação de Pearson mais alto com o VISOR do que para a escala analógica visual em

						papel.
Kreiman, J.; Gerratt, B.R.; Berke, G.S. (1994) – <i>The multidimensional nature of pathologic vocal quality.</i>	Soprosidade Rugosidade	7 níveis	-	+	+	As diferenças na forma como os avaliadores centram a sua atenção em diferentes aspectos das escalas perceptivas, aparentemente, é uma fonte significativa de variabilidade (ruído) nas avaliações da qualidade vocal.
Shrivastav, R. (2006) – <i>Multidimensional scaling of breath voice quality: individual differences in perception.</i>	Soprosidade	7 níveis 5 níveis	-	+	+	Consistência inter-juízes boa para a avaliação perceptual. As diferenças de percepção da qualidade vocal individuais não são tão marcadas como pensado. Um modelo de percepção da qualidade vocal para um avaliador não treinado pode ser uma boa representação para a população geral.
Webb, M.; Starr, C.D.; Moller, K. (1992) – <i>Effects of extended speaking on resonance of patients with cleft palate.</i>	Qualidade Vocal Nasalidade	11 níveis	-	+	+	As medidas de avaliação perceptual deste estudo não proporcionaram medidas muito fidedignas.
Wolfe, V.I.; Martin, D.P.; Palmer, C.I. (2000) – <i>Perception of dysphonic voice quality by naive listeners.</i>	Anormalidade Rugosidade Soprosidade	2 níveis 7 níveis	-	+	+	Para cada uma das três tarefas de escuta a consistência inter-avaliadores foi adequada. Intra-avaliadores: valores obtidos são comparáveis aos de outros estudos.
Yiu, E.M. & Ng, C.Y. (2004) – <i>Equal appearing interval and visual analogue scaling of</i>	Rugosidade Soprosidade	11 níveis VAS	-	+	+	Consistência Intra-avaliadores moderada; significativamente mais alta em EAI; inter-avaliadores semelhante em

<i>perceptual roughness and breathiness.</i>						EAI e VAS.
Kreiman, J.; Gerratt, B.R.; Ito, M. (2007) - <i>When and why listeners disagree in voice quality assessment tasks.</i>	Soprosidade	VAS 8 níveis	-	+	+	A variabilidade inter-avaliador é um problema de desenho do estudo e não de pouca fidedignidade do avaliador. Centraram-se na análise da média, o que pode não corresponder à variabilidade de avaliadores, mas no facto deste tenderem a classificar nos extremos da escala.
Yiu, E.M.; Chan, K.M.; Mok, R.S. (2007) – <i>reliability and confidence in using a paired comparison paradigm in perceptual voice quality evaluation.</i>	Rugosidade Soprosidade	8 níveis	-	+	+	Resultados deste estudo sugerem uma alternativa para as escalas de avaliação perceptual: o paradigma de comparação de pares.
OUTROS						
Bassick, C.J. & Ludlow (1986) – <i>The use of perceptual methods by new clinicians for assessing voice quality.</i>	Darley Rating System	7 níveis	-	+	+	Apesar do amplo período de treino usado, os dados de consistência foram comparáveis com os de estudos em que foram usados avaliadores com experiência. Sugere que a avaliação perceptiva é difícil e requer ampla experiência profissional.
Bele, I.V. (2005) – <i>Reliability in perceptual analysis of voice quality.</i>	16 parâmetros	VAS	-	+	+	Os resultados revelam uma consistência inter-avaliador mais alta para a maioria das características perceptivas. Os avaliadores experientes tendem a ser

						mais consistentes que os estudantes nas avaliações.
Bergan, C.C.; Titze, I.R.; Story, B. (2004) – <i>The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice.</i>	"Ring" "Pressed"	10 níveis	-	+	+	Apesar dos músicos-avaliadores mostrarem um pequena diferença positiva na avaliação das qualidades estudadas, não se obtiveram diferenças significativas face a não músicos para a consistência intra-sujeitos.
Chhetri, D.K. et al. (2008) – <i>Reliability of the perceptual evaluation of adductor spasmodic dysphonia.</i>	Severidade da DEAd	VAS	-	+	+	Os especialistas em voz são capazes de avaliar e acordar a severidade da DEAd de forma consistente.
Damrose, J.F. et al. (2004) – <i>The impact of long-term botulinum toxin injections on symptom severity in patients with spasmodic dysphonia.</i>	Simtomas gerais Severidade	7 níveis	-	+	+	Resultados que sugerem uma boa consistência entre avaliadores no que concerne ao nível da sua experiência.
De Bodt, M.S. et al. (1997) – <i>Test-retest study of the GRBAS scale: influence of experience and Professional background on perceptual rating of voice quality.</i>	GRBAS	4 níveis	-	+	+	Validade de teste-reteste da escala GRBAS com resultados moderados. O parâmetro G(Grade) foi o mais consistente.
de Krom, G. (1994) - <i>Consistency and reliability of voice quality ratings for different types of speech fragments.</i>	GIRBAS	10 níveis	-	+	+	Resultados indicam que os parâmetros grau, rugosidade e sopro para classificar amostras de fala encadeada não são necessariamente mais consistentes ou exactos do que as

						classificações de vogal sustentada, mesmo inferiores a 200ms.
Dejonckere, P.H. et al. (1996) – <i>Differentiated perceptual evaluation of pathological voice quality: Reliability and correlations with acoustic measurements.</i>	GIRBAS	VAS	-	+	+	A consistência da GIRBAS assume-se como muito satisfatória e é recomendada para o uso clínico. Concordância maior para o G(grade).
Erickson, M.L. (2003) – <i>Effects of voicing and synthatic complexity on sign expression in adductor spasmodic dysphonia.</i>	Sinais de DEAd	Contagem	-	+	+	Para o grupo de casos com DEAd: consistência inter-avaliadores de 92% e intra-avaliadores de 90%. Para o grupo de controle: consistência inter-avaliadores de 99% e intra-avaliadores de 97%.
Eskenazi, L; Childers, D.G.; Hicks, D.M. (1990) – <i>Acoustic correlates of voice quality.</i>	Severidade global Rugosidade Soprosidade Aspereza Vocal Fry Excelência da voz normal	7 níveis	-	+	+	Estudo que demonstra uma boa consistência intra-avaliadores, dada a dificuldade das amostras. Os avaliadores tendem a concordar mais na classificação de vozes patológicas e discordar na avaliação de “normais”.
Feijoo, S. & Hernandez, C. (1990) – <i>Short-term stability measures for the evaluation of vocal quality.</i>	Grau Soprosidade	4 níveis	-	+	+	Avaliadores bem treinados são perfeitamente capazes de realizar uma avaliação vocal fidedigna.
Halberstam, B. (2004) – <i>Acoustic and perceptual</i>	Severidade da Rugosidade	7 níveis	-	+	+	Este estudo indica que a percepção de Rugosidade parece ser mais válida na

<i>parameters relating to connected speech and more reliable measures of hoarseness than parameters relating to sustained vowels.</i>						análise de fala encadeada do que em tarefas de vogal sustentada.
Heuer, R. et al. (2000) – <i>The Towne-Heuer Reading Passage – a reliable aid to the evaluation of voice.</i>	Ataque glótico brusco	2 níveis	-	+	+	Atendendo à sua sensibilidade, acessibilidade e facilidade de uso, os autores propõem um texto específico para as provas de avaliação vocal.
Karnell, M.P. et al. (2007) – <i>Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders.</i>	Grau CAPE-V	4 níveis VAS	-	+	+	A percepção dos clínicos acerca da disфонia parece ser fidedigna e não afectada pelo material de classificação. O CAPE-V parece ser mais sensível a pequenas diferenças nos/entre os pacientes que a escala GRBAS.
Kreiman, J. et al. (1993) – <i>Perceptual evaluation of voice quality: review, tutorial, and framework for future research.</i>	Rugosidade	7 níveis VAS	-	+	+	Os resultados sugerem que os métodos tradicionais de avaliação vocal podem nunca conseguir ir de encontro aos standards de consistência interna. Contudo, podem ser desenvolvidos protocolos de classificação para controlar algumas das fontes de variabilidade na percepção da qualidade vocal dos avaliadores.
Kreiman, J. & Gerratt, B.R. (1998) – <i>Validity of rating scale measures of voice quality.</i>	Rugosidade Severidade, semelhança	7 níveis VAS	-	+	+	Os valores baixos de concordância entre avaliadores obtidos por esta investigação reforçam que os protocolos tradicionais de avaliação de qualidades como a

						soprosidade e Rugosidade não são úteis para medir a qualidade vocal percebida.
Langeveld, T.P. et al. (2000) – <i>Perceptual characteristics of adductor spasmodic dysphonia.</i>	GRBAS Afonia Diplofonia Staccato, Tremor Falsetto, Vocal Fry	VAS	-	+	+	O esquema GRBAS aumentado pelos autores deste estudo mostrou adequação e reprodutibilidade na caracterização perceptual da DEAd.
Lee, M. et al. (2005) – <i>The reliability and validity of patient self-rating of their own voice quality.</i>	GRBAS	4 níveis	-	+	+	Os pacientes demonstraram boa validade e consistência ao auto-avaliarem-se com a escala GRBAS. As medidas de consistência em relação à avaliação dos profissionais foi pobre.
Ma, E.P & Yiu, E.M. (2001) – <i>Voice activity and participation profile: assessing the impact of voice disorders on daily activities.</i>	Grau Rugosidade Soprosidade	VAS	-	+	+	Um coeficiente de correlação intra-avaliadores para o parâmetro soprosidade não foi estatisticamente significativo; todos os restantes foram-no.
Morsomme, D. et al. (2001) – <i>Comparison between the GIRBAS Scale and the Acoustic and a Aerodynamic Provided by EVA for the Assessment of Dysphonia following Unilateral Vocal Fold Paralysis.</i>	GIRBAS	VAS	-	+	+	A consistência intra e inter-avaliadores foi próxima do nível bom, no grupo de controle.
Munoz, J. et al. (2002) – <i>Perceptual analysis on</i>	Buffalo III	5 níveis	-	+	+	Concluiu-se que a Buffalo III é um protocolo de avaliação fidedigno para

<i>different voice samples: agreement and reliability.</i>						vozes normais e patológicas espanholas.
Titze, I.R. et al. (2003) – <i>Source and filter adjustments affecting the perception of the vocal qualities twang and yawn.</i>	"Twang" e "Yawn"	10 níveis	-	+	+	Não foi realizado nenhum teste-reteste. Resultados apontam maior variabilidade da avaliação de não-músicos face a músicos.
van der Tom, M. et al. (2002) – <i>Communicative suitability of voice following radiotherapy for T1glottic carcinoma: testing the reliability of a rating instrument.</i>	Adequação comunicativa Qualidade vocal	10 níveis 7 níveis	-	+	+	O conceito de "adequação comunicativa da voz" parece ser basicamente o som. As escalas são consistentes e as medidas possibilitam a diferenciação entre grupos de vozes normais e patológicas.
Webb, A.L. et al. (2004) – <i>The reliability and validity of patient self-rating of their own voice quality.</i>	Buffalo III VPAS GRBAS	5 níveis 3 níveis 4 níveis	-	+	+	Buffalo III: apenas grau geral é fidedigno; VPAS: consistência pobre a moderada; GRBAS: consistente para todos os parâmetros excepto Soprosidade. GRBAS é uma medida simples e fidedigna para uso clínico.
Martens, J.W. et al. (2007) – <i>The effect of visible speech in the perceptual rating of pathological voice.</i>	GIRBAS "Visible Speech"	4 níveis	-	+	+	Estudo mostra que o uso de pistas visuais aumenta claramente a reprodutibilidade entre tarefas para os 3 principais parâmetros: grau, Rugosidade e soprosidade. Consequentemente, enfatiza a consistência e relevância da avaliação perceptual e justifica o seu uso alargado.
Moerman, M. et al. (2006) – <i>The INFVo perceptual rating</i>	INFVo	VAS	-	+	+	O INFVo parece constituir um material fidedigno para a avaliação perceptual de

<i>scale for substitution voicing: development and reliability.</i>						vozes alternativas. O acordo inter-avaliadores foi bom para semi-profissionais e excelente para profissionais.
Murry, T. et al. (2004) – <i>The relationship between ratings of voice quality and quality of life measures.</i>	GRBAS	4 níveis	-	+	+	Os dois profissionais foram consistentes nas suas avaliações da severidade da qualidade vocal com a escala GRBAS.
Shrivastav, R. et al. (2005) – <i>Application of psychometric theory to the measurement of voice quality using rating scales.</i>	Soprosidade	5 níveis	-	+	+	No geral, a probabilidade de um acordo preciso entre classificações (consistência inter-avaliadores) pode ser aumentada se: realizarmos a média de cotações para cada estímulo (minimiza o E_{random}), feito por cada avaliador (de .4 para .9); e, se forem usadas cotações standardizadas, principalmente para comparar dados entre avaliadores (minimiza o $K_{criterion}$).
Stewart, C.F. et al. (1997) – <i>Adductor spasmodic dysphonia: standard evaluation of symptoms and severity.</i>	USDRS	7 níveis 2 níveis	-	+	+	O USDRS é um material bom para quantificar os sintomas vocais de doentes com DEAd.

Legenda:	
VAS – Visual Analogue Scale SME – Direct Magnitude Estimation VISOR – Visual Sort and Rate DEAd – Disfonia Espasmódica Adutora	GIRBAS – Grade, Instability, Roughness, Breathiness, Asthenic, Strained CAPE-V – Consensus Auditory Perceptual Evaluation – Voice VPAS – Vocal Profile Analyses Scheme

GRBAS – Grade, Roughness, Breathiness, Asthenic, Strained

INFVo – Impression Inteligibility Noise Fluency Voicing
USDRS – Unified Spasmodic Dysphonia Rating Scale

1.3. A AVALIAÇÃO ACÚSTICA

A avaliação acústica – quando utilizada no âmbito da intervenção terapêutica nas perturbações vocais – permite quantificar, de modo não invasivo, as características da voz humana (Behlau, 2001; Pinho, 2003; Pinho et al., 2006; Awan & Roy, 2009). O método acústico propicia as designadas medidas objectivas, ou seja, dados que são extraídos automaticamente, ou não, através de um processamento computacional adequado. Aquele possibilita a integração de dados fornecidos pela avaliação perceptivo-auditiva com o plano fisiológico, pois permite a especificação detalhada do processo de geração do sinal sonoro: fornece, então, uma estimativa indirecta dos padrões vibratórios das pregas vocais, bem como dos formatos do tracto vocal supraglótico e das respectivas modificações (Hirano & Bless, 1997). Para tal recorre a diferentes parâmetros acústicos que compõem o sinal – periodicidade, amplitude, duração e composição espectral (Murdock, 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 (Camargo, 2000).

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 espectrograma – 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 aspectos do tempo (já referido), a frequência e a intensidade num único gráfico de dois eixos (Kent e Read, 1992). Em 1960, Gunnar Fant postulou a teoria da fonte-filtro linear (ou modelo de fonte-filtro) que explica matematicamente a produção da voz e da fala (Johnson, 2003). Este pressupõe três principais sub-sistemas acústicos para a produção de voz: de conformação do pulso glótico; de caracterização

do tracto vocal e de radiação labial (Kent, 1993, 1997). Na imagem 3 estão ilustrados (à esquerda) o modelo fisiológico dos sistemas envolvidos na produção de voz, assim como a correlação com o modelo linear de fonte-filtro proposto por Fant.

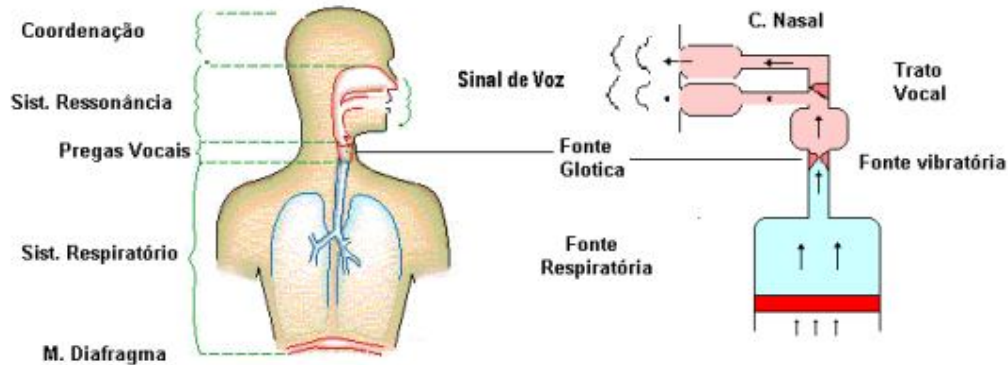


Imagem 3: Desenho esquemático dos sistemas e órgãos envolvidos na produção do sinal de voz e do modelo linear de fonte-filtro proposto por Fant (1970) (adaptado de Dajer, 2010).

A década de 70 marcou o desenvolvimento do processamento digital de sinais, com a divulgação dos primeiros softwares 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). 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* (Araújo, 1999; Baken e Orlikoff, 2000; Childers et al., 1997; Davis, 1979; Pereira e Montagnoli, 1999; Read et al., 1992). Deste modo, as medidas obtidas na análise acústica correspondem a parâmetros físicos definidos. O sinal glótico (sinal da fonte) sofre efeitos ao longo do tracto vocal supraglótico até à saída deste para o meio externo (acção de filtro) (cf. Imagem 2). Acontece um somatório das ondas sonoras provenientes da fonte glótica com outras reflectidas ao longo do tracto vocal, sendo o resultado final (sinal de saída) o sinal irradiado pelos lábios (Fant, 1970; Kent & Read, 1992).

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. Actualmente as propostas – marcas e equipamentos – disponíveis no mercado são inúmeras. Contudo, a sua selecção 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, salientem-se os seus principais objectivos e vantagens de utilização (Barros & Carrara-De Angelis, 2002; Behlau, 2001; Pinho et al., 2006; Camargo & Madureira, 2010):

- Oferece uma maior compreensão acústica do *output* vocal e aproxima formas distintas de avaliação da voz, nomeadamente a análise áudio-perceptual e a acústica ou a laringosestroboscópica e a acústica;
- Proporciona – de modo *expedito* e *user-friendly* – dados normativos para realidades vocais distintas – culturais, profissionais e/ou patológicas;
- Propicia informação importante sobre o impacto do sinal vocal no ouvinte (Weismer, 1984 *in* Murdock, 2005);
- 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 carácter legal;
- 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;
- 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;
- 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;

- Assume-se como um instrumento de detecção precoce de problemas vocais e laringeos, por exemplo em campanhas de triagem, pela detecção de níveis de perturbação fonatória acima dos valores de referência de uma população não-disfónica.

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 4).

<i>Opensource Software</i>	<i>Comercializados</i>
Audacity 1.3.12 (Audacity Team)	CSL, model 4500 (Kay Elemetrics, 1996)
EMU Speech Data System (LMU, 2009)	Dr. Speech 4.0 (Tiger Electronics, 1999)
Praat (Boersma & Weenink, 2003)	IVANS (Avaaz Innovation, 2000)
SFS (Speech Filing System, 1999)	MDVP, model 5105 (Kay Elemetrics, 1996)
Speech Analyser (SIL International, 2007)	MediVoz (TGH Endoscopia SL.)
Sound Forge 10.0b (Sony Creative Software Inc., 2010)	Speech Station2 (Sensimetrics)
Wasp (Waveform Annotations Spectrograms and Pitch)	Speech Studio (Laryngograph Ltd.)
Wavesurfer (Sjolander & Beskow, 2003)	TF32 (Milenkovic, 2003)
	VoiceStudio (Seegnal, 2008)
	Voxmetria (CTS Informática, 2003)

Tabela 4: Principais *softwares* de análise acústica de voz.

A fiabilidade da informação recolhida e analisada está dependente das condições e procedimentos de captação, armazenamento, edição e análise do sinal sonoro (Guimarães, 2007). Isto porque os parâmetros acústicos reflectem a complexa interacção entre a fonte glótica e as cavidades de ressonância do tracto vocal. Assim, dependem das forças biomecânicas e aerodinâmicas da laringe e estruturas supra-glóticas (Mendes, 2003), assim como do complexo controle 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 respectiva severidade. Neste trabalho consideraremos o sinal de fala como “o produto de fontes de voz e de ruído e dos efeitos de ressonância causados pelo trato vocal, efeitos esses que servem de *input* para o sistema auditivo e para o sistema nervoso central e que, portanto, remetem ao conhecimento das bases fisiológicas, acústicas e cognitivas implicadas na produção e percepção da fala” (Madureira, 2007 *in* Camargo e Madureira, 2010).

Tradicionalmente, as vogais sustentadas produzidas com “níveis confortáveis de amplitude e frequência” (Pinho et al., 2006) são usadas como amostra a analisar acusticamente e, assim, caracterizar a qualidade vocal de um indivíduo. 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 co-articulação, ou seja, elimina os factores de confusão inerentes às influências supraglóticas e linguísticas. Martens et al. (2007) sugerem que seja analisada uma porção relativamente estável da vogal sustentada, eliminando os primeiros e últimos 250ms do sinal, o que inclui o *onset* e *offset*. Boucher (2008) também suporta o seu estudo neste princípio. Pecam contudo por não representarem a comunicação verbal e, por isso, moderarem a severidade de uma disфонia (Bassich & Ludlow 1986 *in* Bele, 2005; Klingholtz, 1990; Laver et al., 1992; Yiu et al., 2000; Fourcin, 2000 *in* Guimarães, 2007; Revis et al., 1999; Parsa e Jamieson, 2001; Revis et al., 2002 *in* Pouchoulin, 2008; Maryn et al., 2009). O prolongamento de vogais pode ser usado para obter informação sobre a duração fonatória, o controle pneumofónico, níveis de *pitch* e *loudness* confortáveis, a presença de alterações e a capacidade de manter uma articulação estável. Realizar variações em escala de uma mesma vogal pode ser usado para determinar a faixa de frequência do sujeito. O mesmo raciocínio aplica-se à variação da intensidade vocal (Murdock, 2005). Note-se que a escolha da vogal afecta as medidas de perturbação (Camargo & Madureira, 2010), tal como discutido ao longo deste capítulo. de Krom (1994) usou a escala analógica visual no seu estudo em que

comparou a classificação perceptual em diferentes tipos de estímulos vocais (vogal sustentada vs fala encadeada) e concluiu que o *onset* e/ou toda a vogal possuía uma resolução superior, reflectindo 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 perceptual completa e correcta das características fonatórias, uma vez que os falantes podem sentir-se poucos confortáveis nesta prova, culminando em dados derivados de um comportamento vocal não natural e, assim, incaracterístico (Klingholz, 1990; Qi & Hillman, 1987; *in* Parsa & Jamieson, 2001).

O uso da leitura de um texto foneticamente equilibrado¹ é sugerido por alguns autores como a prova que no laboratório de voz mais se aproxima do discurso espontâneo. 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 (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 co-articulatórias, padrões de entoação, velocidade de fala, características e duração (Murdock, 2005).

O recurso a amostra de discurso espontâneo é, claro, o mais realista (Eadie & Baylor, 2005; Maryn et al., 2009) – em termos fonéticos e prosódicos – embora contenha uma variabilidade marcada, inerente ao contexto comunicativo e características psico-emocionais do sujeito (Bickley & Stevens, 1986; Gobl, 1988; Gobl & Ní Chasaide, 1988; Löfqvist & McGowan, 1991; Rammage et al., 1992 *in* 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, 1973; Muta et al., 1988; Schoentgen, 1989 *in* de Krom, 1994) já que oferecem informação sobre o desempenho ao longo do tempo, a manutenção/estabilidade das qualidades vocais, a

¹ Entenda-se como aquele que é representativo dos fonemas de uma dada língua, em todas as suas combinações possíveis. (Pouchoulin, 2008)

variedade e os efeitos das influências co-articulatórias (Löfqvist & McGowan, 1991; de Krom, 1994; Bele, 2005; Goble et al., 1991 in Eadie & Baylor, 2005; Murdock, 2005). Sabendo que a fala encadeada pressupõe uma maior complexidade de produção fisiológica comparativamente com as vogais sustentadas, os aspectos de alteração vocal estão mais evidenciados, e como tal poderão ser melhor classificados perceptualmente (Bassich & Ludlow, 1986). Saliente-se um estudo que argumenta o contrário: de Krom (1994) encontrou uma consistência inter-avaliadores da classificação áudio-perceptual mais baixa neste tipo de amostra vocal. O autor alega que tal ficou a dever-se a dois factores: 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 extra-glóticos (por exemplo, de dialecto ou velocidade do discurso) que podem ser pistas perceptivas que distorcem e dificultam a classificação final. Esta complexidade da fala encadeada, ao aplicar o método acústico, pressupõe a análise com técnicas mais sofisticadas (Klingholtz, 1990; Laver et al., 1992 in Guimarães, 2007). 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 22.5% de taxa de erro de classificação das vozes patológicas, e na fala encadeada esse erro subia 5.6%.

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 (Campisi et al., 2000; Awan, 2005; Behrman, 2005; Vieira et al., 2005; Ma e Yiu, 2006; Sousa et al., *no prelo*; Brasolotto & Rehder, 2011) que permite o registo e oferece a possibilidade de maior detalhe do processo de geração do sinal sonoro, o qual corresponde a eventos das porções glóticas e supra-glóticas do aparelho fonador (Kent & Read, 1992). Note-se que Weber (2002) estudou um conjunto de patologias laríngeas, associando-as ao grau de disфонia, e reitera que em casos com alterações muito marcadas o programa de análise acústica não é capaz de realizar a avaliação.

Os principais entraves (de Krom, 1994; Rabinov et al., 1995 in Finizia et al., 1999; Butha, 2004; Behram, 2005; Pouchoulin, 2008) 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 associado 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 gastos 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 (1994).

1.4. A AVALIAÇÃO E ANÁLISE ACÚSTICA DA VOZ

A avaliação acústica realiza medições do sinal sonoro vocal (Behlau, 2001). A aplicação de diversas formas de análise está dependente da correcta digitalização do sinal acústico. Vários autores oferecem sugestões ou normas para obter um adequado registo áudio de vozes (Behlau, 2001; Mendes, 2003; Pinho, 2003; Murdock, 2005; Smits et al., 2005; Guimarães, 2007). Assim, conseguir-se-á a sua análise do ponto de vista visual (espectrografia) e quantitativa (extracção de medidas temporais, de periodicidade, amplitude, perturbação, ruído ou outras).

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

a) Quanto a condições de registo:

A revisão da literatura internacional sugere:

- A captação directa da voz para o computador é o ideal (mínimo 16 bit/amostra de placa analógico-digital, para reduzir o chamado ruído de quantização gerado pela placa de som) (Smits et al., 2005);
- As gravações devem ser realizadas em ambientes silenciosos (idealmente em cabine insonorizada e protegida contra sinais eléctricos), com ruído de fundo inferior a 50 dB, em circunstâncias controladas e passíveis de reprodução;
- 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 20.000 amostras/segundo para que, conforme o teorema de Nyquist, frequências de até 10.000Hz possam ser preservadas no sinal digitalizado (Waught, 2000 *in* Guimarães, 2007). Se a extensão da audição humana varia entre 15 a 20 000 Hz, então uma frequência de amostragem de 44 100Hz ($2 \times 20 \text{ KHz} = 40 \text{ KHz}$) por segundo é suficiente para evitar o risco de saturação (Pinho et al., 2006).

- Qualquer método de aquisição e armazenamento de sinal pode afectar, substancialmente, a qualidade do mesmo;
- Vozes importadas – seja de que sistema for – mesmo digitais, podem ter sido significativamente alteradas por ruídos, adicionados aquando da sua captura;
- Uma verificação parcial do nível de ruído introduzido no sinal pelo sistema de gravação ou captura de voz pode ser realizada, comparando-se dois registos: gravação com microfone desligado e outra com ele ligado;
- O indivíduo deve estar, preferencialmente, de pé, para melhor controle postural e aproximação ao uso habitual da voz;
- O microfone deve ser instalado num apoio fixo, num ângulo de 45 a 90° da boca do falante, a uma distância inferior a 10 cm (entre 3 a 4 cm) – para recolha de vogais sustentadas (Vieira et al., 2005); 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). Outros autores propõem 15 cm (6 pol.), preferencialmente controlados com um microfone de cabeça (Murdock, 2005; Eadie & Baylor, 2006). 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);
- Verificar a impedância do microfone, para não haver distorções na gravação;
- 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.

b) Quanto a 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, projecção vocal ou outra; e, (3) na duração – tempo ou unidade de fala, com consequências na validade e fiabilidade das medições (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, pelo menos uma vogal sustentada (de Krom, 1994) e uma sequência automática de fala encadeada;
- A análise da frequência fundamental e os seus índices de perturbação é, geralmente, realizada com base nas vogais /i/, /u/ e /a/;
- A análise da qualidade vocal suporta-se numa amostra de fala encadeada (Hammarberg, 2000; Martens et al., 2007);
- A necessidade de um maior número de análises em casos em que o desvio vocal é mais acentuado. 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.;
- A verificação imediata da qualidade da gravação, logo após a sua realização, antes que o paciente saia do consultório.

c) Quanto a sistemas de gravação:

O principal aspecto a ter em atenção é a escolha do microfone para a captação de voz. Assim, este deve ser seleccionado quanto ao tipo (dinâmicos), direcionalidade (omnidireccional), frequência de resposta (linear), impedância (baixa), modo de uso e distância da fonte (Guimarães, 2007). Contudo, atente-se a que:

- Ao usar um gravador este deve ser profissional, equipado com um microfone, condensador, mono, unidireccional, com sensibilidade mínima de -60dB, deslocado do corpo da unidade de gravação para evitar a captação de ruído do hardware;
- O *minidisc* revela uma diferença reduzida da qualidade de gravação face à sua realização directa para o computador. Contudo, a edição e registo das amostras, assim como a sua passagem para outros formatos, introduz factores de compressão e 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 em cassetes VHS apenas podem ser analisadas perceptualmente, uma vez que o microfone está acoplado ao corpo da câmara e isso acrescenta ruídos extrínsecos ao sinal acústico.

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. Imagem 4).

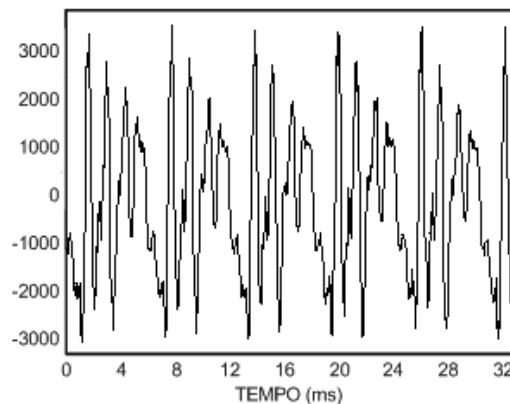
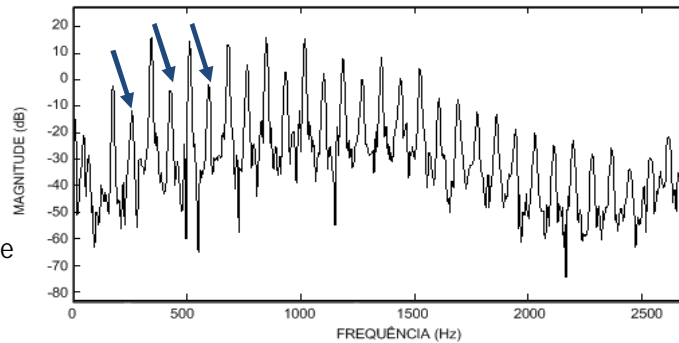


Imagem 4: Exemplo de Sinal Tipo 1.

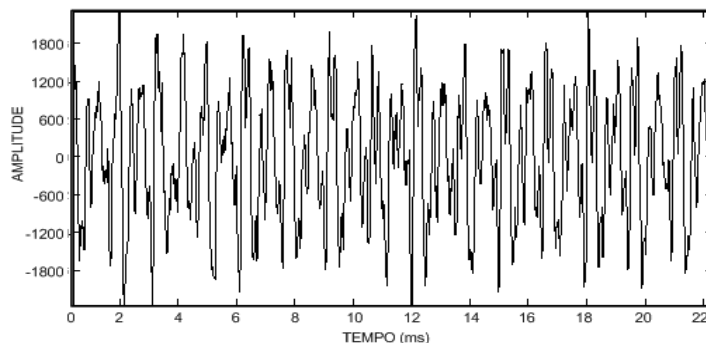
- **Sinal Tipo 2:** periódico e com estrutura sub-harmónica e modular – é um sinal acústico que apresenta alterações qualitativas no segmento analisado, ou seja, bifurcações, intermitências, sub-harmónicos (cf. Imagem 5) e modulações. A frequência dos sub-harmónicos ou modulações é de energia considerável e, por isso, não existe uma frequência fundamental única no segmento em questão. Embora se possam seleccionar 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.

Imagem 5: Exemplo de Sinal Tipo 2.



- **Sinal Tipo 3:** sinal não periódico – não tem estrutura periódica estável (cf. Imagem 6), 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 (Côrtes & Gama, 2010) como o sinal anterior. Giovanni et al. (1999) propõem alguns sistemas de medição não linear como a análise fractal, atratores ou o expoente de Lyapunov para a análise acústica não linear deste tipo de sinal glótico (Yu et al, 2000; Behlau, 2001; Maacallum, et al., 2009; Vaziri et al., 2010), a qual salienta o estudo pormenorizado das bifurcações ou mudanças súbitas da qualidade do padrão vibratório.

Imagem 6: Exemplo de Sinal Tipo 3.



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. 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 in 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 tão extremos que condicionam a identificação de um período regular para análise. Neste domínio há

particularidades temporais e espectrais do sinal directamente relacionadas com a disфонia que não podem ser nem descartados, nem excessivamente predominantes no processo de parametrização (Herzel e tal., 1994; Sataloff & Hawkshaw, 2001 in Kent, 2004).

Deste modo, outra dificuldade adicional na análise acústica do sinal vocal é o sistema empregue na extracção 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 (Titze, 1995; Parsa & Jamieson, 2001). No estudo de Espanhol (2004) foi demonstrado que os “determinadores” disponíveis 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 vs 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 estimado de SNR discriminar entre disфонia-voz normal, com uma correlação de 0.78, e uma capacidade de classificação 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 spectral*) e o SFR (*spectral flatness ratio*) explicavam 96% dos resultados.

São encontradas grandes diferenças nos valores das medidas para os diferentes softwares disponíveis, tal compromete a viabilidade e confiabilidade dos resultados encontrados na literatura e limita fortemente a possibilidade de comparação inter-estudos (Carson et al., 2003; Smits et al., 2005; Lopes et al., 2008).

Embora as investigações da última década e a sua aplicação clínica tenham trazido inúmeras contribuições, importa ressaltar que este é um procedimento complexo (Brasolotto & Rehder, 2011) que requer um suporte teórico e treino para o manuseio dos recursos computadorizados disponíveis actualmente. Note-se que muitos investigadores consideram, consensualmente, que a análise acústica da voz patológica – baseada apenas em parâmetros de perturbação – não tem confiabilidade, pois o sinal vocal possui uma aperiodicidade de grau extremamente variável, que muitos dos *softwares* não são capazes de estimar de forma eficaz (Titze, 1995; Bielałowicz et al., 1996; Yiu, 1999 *in* Yiu et al, 2000; Ma & Yiu, 2005). Assim, recomenda-se cautela na utilização deste recurso de avaliação vocal isolado (Eadie & Baylor, 2005; Ma & Yiu, 2006; Maryn et al., 2009; Camargo & Madureira, 2010; Rahn et al. *in* Dajer, 2010) por forma a conseguir que ele tenha um verdadeiro significado clínico.

1.5. 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 factores de perturbação objectivos, associados à forma de onda do vozeamento (cf. Imagem 7).

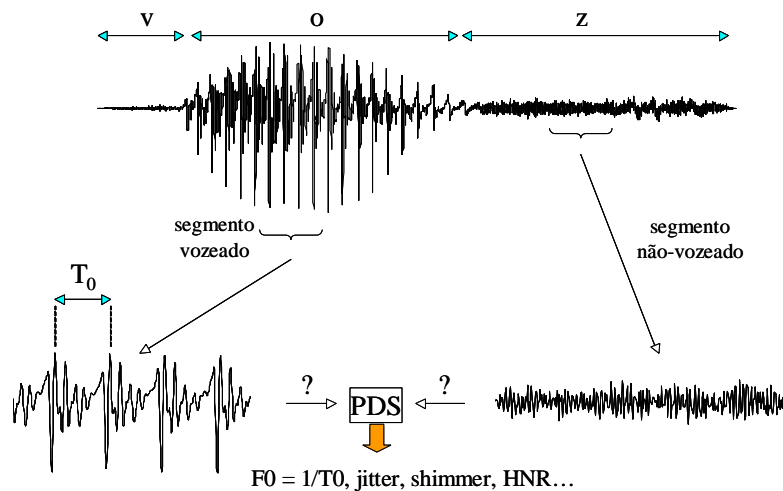


Imagem 7: Ilustração do sinal de voz captado por um microfone e correspondente à palavra voz. Destaca-se a região vozeada do sinal e a região não-vozeada. Usando técnicas de Processamento Digital de Sinal é possível a medição objectiva e precisa de alguns parâmetros de perturbação extraídos directamente do sinal acústico.

O sinal da voz pode ser analisado de duas formas distintas: directas e indirectas.

- **Formas directas do sinal:** são aquelas em que a análise do sinal se baseia na onda de pressão que irradia da boca, isto é, consiste na técnica que interpreta o sinal vocal como um fenómeno de pressão sonora.

- **Formas indirectas do sinal:** pressupõem uma versão modificada do sinal vocal, como por exemplo o espectro, o *cepstrum* e o sinal residual. Podem incluir a separação – através de fórmulas matemáticas de modelagem do tracto vocal – entre a fonte glótica e o filtro. Há técnicas específicas de filtragem inversa que retiram todo o efeito do

tracto. Existem 2 métodos a enumerar: técnica de filtragem inversa residual e técnica de filtragem inversa glotal (Davis, 1979; Pereira e Montagnoli, 1999).

A técnica de filtragem inversa residual remove os efeitos acústicos do tracto vocal sobre as características do sinal de voz, de forma a especificar, de forma aproximada, aquele que resulta da excitação glótica. Ou seja, inclui um conjunto de técnicas que subtraem um modelo fixo dos sub-sistemas glótico e supra-glótico, mantendo as informações do primeiro. As técnicas de filtragem inversa glotal retiram os efeitos acústicos da radiação labial e do tracto 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 tracto vocal, uma vez que cada um resulta em diferenças acústicas conhecidas. O tracto vocal é responsável pelas ressonâncias (formantes), enquanto que 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* (Dliyski, 1993); VTI – *Voice Turbulence Index* (Dliyski, 1993); TNI – *Turbulent Noise Index* (Mitev & Hadjitodorov, 2000); SRA – *Sum of Rahmonics Amplitudes* (Murphy, 2006); DVB – *Degree of Voice Breaks* (Nikolov et al., 1989); DSH – *Degree of Sub-Harmonics* (Deliyski, 1993)) porém, nesta dissertação realizaremos a descrição e análise de medidas computacionais que recorrem a técnicas de Processamento Digital de Sinal (PDS) que permitem a medição de parâmetros objectivos como a frequência fundamental (F_0) da voz (ou *pitch*²), parâmetros de perturbação como o *jitter* ou *shimmer*, ou de qualidade como a relação harmónico-ruído (HNR), os quais são transversais aos programas de software seleccionados para a análise das vozes-alvo.

² Em rigor, o *pitch* denota o correspondente psicofísico (i.e., perceptivo) da frequência fundamental (F_0) e é condicionado por outros factores objectivos do sinal de voz como seja a sua intensidade. Contudo, para simplificar a discussão, consideramos neste trabalho que *pitch* e F_0 são sinónimos.

ESPECTROGRAFIA

Esta é uma das técnicas mais usadas na 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 computadorizada 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 – um múltiplo 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 (Pinho et al., 2006).

A informação disponibilizada por esta técnica na sua representação, designada por espectrograma, 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.* Imagem 8).

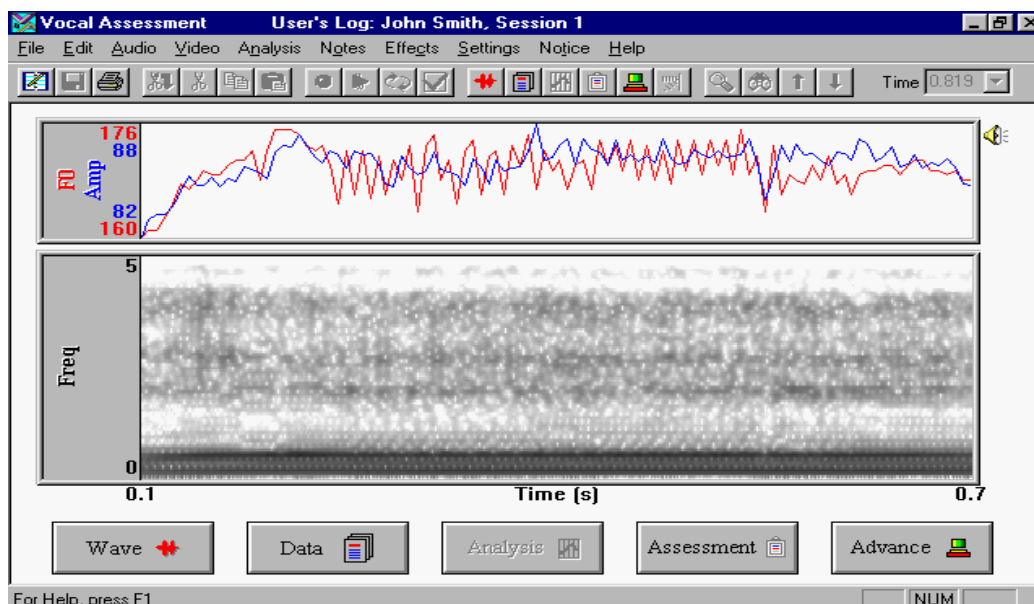


Imagem 8: Espectrograma de um caso com o *software* Dr. Speech®.

A vantagem da espectrografia é o facto de reflectir 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 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, primeiro, 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).

Sons quase-periódicos – como das vogais – são representados pela sobreposição de várias componentes sinusoidais, cada uma com a sua frequências, amplitude e defasamento (atraso). (Pinho et al., 2006) O som complexo como o da voz humana pode ser decomposto em componentes harmónicos através do uso de filtros de banda larga (*cf.* Imagem 9) ou de banda estreita (*cf.* Imagem 10). Se for usado um filtro de banda larga é obtida uma melhor resolução temporal, enquanto que o de banda estreita potencia a resolução da frequência.

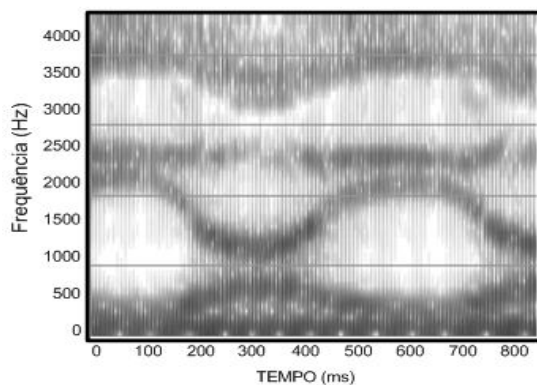


Imagem 9: Espectro de banda larga.

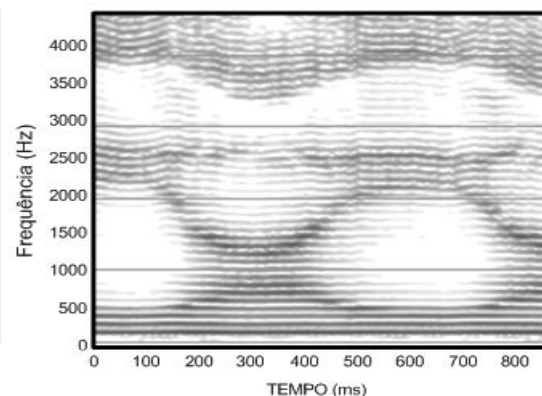


Imagem 10: Espectro de banda estreita.

O espectrograma de banda larga resulta da aplicação de um filtro de banda larga e, geralmente, tem uma faixa de frequências de 200-500Hz (Murdock, 2005). É um gráfico de resolução temporal disposto em estrias verticais, que salienta as regiões de

energia do sinal, apontando a existência de actividades 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 directa com a configuração do tracto vocal durante a emissão do som – variável de acordo com características de postura dos articuladores – faríngeos e bocais – associados (Souza, 2010).

O espectrograma de banda estreita é retratado com estrias horizontais, cada uma representa um harmónico. 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 (Murdock, 2005). Tem como principal finalidade a representação da estrutura harmónica do sinal da fala (Souza, 2010).

Yanagihara, 1967 (*in* Guimarães, 2007) estudou as perturbações vocais com base na interacção entre os componentes de ruído do formante principal, componentes do ruído de altas frequências (acima dos 3 kHz) e perda de componentes harmónicos de alta frequência. Concluiu que existem quatro categorias para classificação do grau de severidade da disфонia, com base nos traçados espectrográficos de vogais sustentadas:

- GRAU I: os componentes harmónicos misturam-se com os elementos de ruído, principalmente na região dos formantes das vogais;
- GRAU II: os componentes de ruído predominam sobre os harmónicos do segundo formante. Existem também ligeiros componentes de ruído de alta frequência acima dos 3 kHz;
- GRAU III: o segundo formante é totalmente substituído por ruído e o componente de ruído de alta frequência (acima dos 3 kHz) intensifica a sua energia e expande a sua extensão;
- GRAU IV: o primeiro formante perde os seus componentes periódicos e o segundo é substituído por componentes de ruído. Nas altas frequências o ruído intensifica-se ainda mais.

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 (Behlau, 2001). Assim, reflecte o número de ciclos vibratórios produzidos pelas pregas vocais, num segundo (Pinho, 2003). É classificada como uma medida acústica no domínio do tempo.

Reflecte a eficiência do sistema fonatório, a biomecânica laríngea (comprimento natural da prega vocal, alongamento, massa em vibração e tensão envolvida) e a sua interacção com a aerodinâmica (pressão subglótica resultante da passagem do ar pulmonar).

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 afectada pelo género, idade (*cf.* Imagem 11), comportamento vocal avaliado, hábitos pessoais e estilo de vida (tabaco, álcool), o uso de voz profissional e o tipo de perturbação vocal apresentada (Kent & Read, 1992).

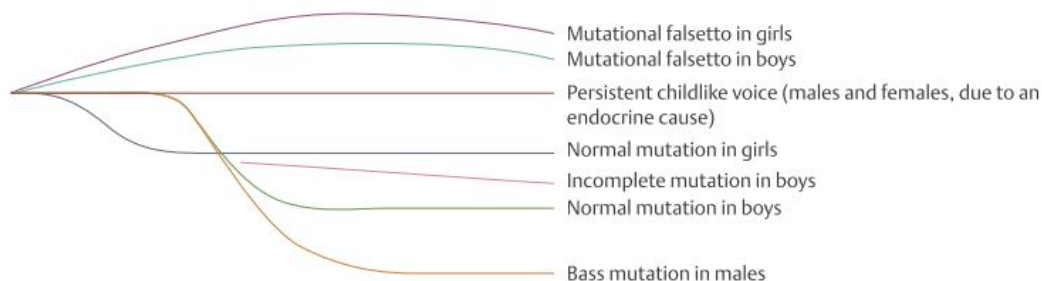


Imagem 11: Representação esquemática da mudança vocal na adolescência (Probst et al., 2004).

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, respectivamente. Wuyts et al. (1996) definiram para homens gamas de frequência entre 78-166Hz e para as mulheres 167-258Hz.

Quando extraímos o valor da F0 estamos a considerar um referencial médio, obtido por média aritmética, que é a soma das medidas de F0 dividida pelo número de ondas captadas.

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).

Um outro método de extracção da frequência fundamental que requer, contudo, um processador de alta velocidade, é o *cepstrum*³. 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; Mehta & Hilman, 2008; Awan & Roy, 2009), que se baseia no facto do sinal vocal ser a convolução do sinal da fonte (laringe) e a resposta do tracto vocal (sistema de ressonância). Este processo significa a interacção entre dois conjuntos de propriedades espectrais (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 confundidas por variáveis como a técnica de gravação das amostras, volume de gravação ou aperiodicidade do sinal de voz (Carding et al., 2004).

³ O *cepstrum* aqui considerado (*cepstrum real*) consiste na transformada de Fourier inversa do logaritmo do espectro (o que explica a designação de 'ceps' como inverso de 'spec'). Remete portanto para um domínio do tempo que caracteriza a periodicidade existente no espectro. Em termos práticos, é útil por exemplo para calcular o período fundamental (em segundos) de uma estrutura harmónica.

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. Neste sentido as medidas dependentes de F0 sofrem a interferência da aperiodicidade do sinal e devem ser cuidadosamente aplicadas (Vieira et al., 1996; Vieira et al., 2002; Camargo et al., 2003; Brasolotto & Rehder, 2011).

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);
- 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;
- 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).

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);
- A frequência fundamental pode ser influenciada pelos factores comprimento, alongamento, massa e tensão das pregas vocais e possui integração com a pressão sub-glótica (Brasolotto & Rehder, 2011);
- A F0 parece não mudar significativamente com a Terapia da Fala, salvo excepções;
- Vozes com crepitação e rugosidade tendem a apresentar um F0 grave, enquanto que vozes ásperas caracterizam-se com F0 agudo;

- Situações de extrema tensão psicológica podem implicar vozes excepcionalmente agudas;
- A voz dos surdos torna a extracção da F0 difícil.

Pinho et al. (2006) argumentam que contrariamente à análise espectral (onde os algoritmos são amplamente conhecidos e os programas, em geral, confiáveis) a extracção automática da F0 é um problema delicado, potencialmente sujeito a erros que se propagam às medidas de perturbação a curto-prazo, que dela derivam.

DESVIO PADRÃO DA FREQUÊNCIA FUNDAMENTAL (*pitch sigma*)

Na fala normal é esperado um certo teor de variabilidade da frequência fundamental. Todavia, as situações extremas são indesejáveis.

Esta medida é geralmente expressa em semitons e 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 internacional também pode aparecer com a designação *pitch sigma* (Ladefoged, 2006).

Sugestões para obtenção de medidas de variabilidade de F0

- Pode medir-se com base numa vogal sustentada ou fala encadeada;
- Quando medida em vogais sustentadas não ultrapassa os 2 Hz 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 Elementrics) considera que para homens o F0 S.D. varia entre 0-2.9Hz, e para mulheres 0-6.6.Hz (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.

Variabilidade da frequência fundamental nas perturbações vocais (Behlau, 2001)

- As disfonias de etiologia neurológica associam-se, frequentemente, a grande variabilidade de F0;
- Momentos de stress e ansiedade influenciam (aumentam) o desvio padrão da F0 (salvo raras exceções);
- Os gagos apresentam este parâmetro e um gama tonal mais restrita;
- O surdo pode apresentar emissões com grande variabilidade de F0.
- Estudos consideram esta medida eficaz para predizer o grau geral de alteração vocal (Wolfe & Steinfatt, 1987; Callan et al., 1999 *in* Awan & Roy, 2009)

JITTER – PERTURBAÇÃO DA FREQUÊNCIA A CURTO PRAZO

O *jitter* é uma medida de curto termo (ciclo a ciclo), de variabilidade não voluntária na F0 (Pinho et al., 2006), que permite determinar o grau de variabilidade do sistema fonatório. 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) glóticos⁴ consecutivos de frequências F0(1) e F0(2) pode ser calculado por:

$$J (\%) = |F0(1) - F0(2)| / (\frac{1}{2}[F0(1) + F0(2)]) \times 100$$

A extracção dos valores de *jitter* pode ser feita através de medidas absolutas ou relativas. Estas correlacionam-se com a média da F0. As primeiras ignoram a F0 do indivíduo. Sumariamente, podemos indicar que a partir deste parâmetro são derivadas as seguintes medidas acústicas (MEEI, 1994; Moran et al., 2006):

- *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.

⁴ 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).

- *Absolute Jitter* [μ s] (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 do desvio-padrão relativo da frequência fundamental calculado período a período.

- *F0-Tremor Intensity Index* [%] (FTRI): é calculada através do quociente 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 quociente 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 quociente 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 (Guimarães, 2007), em especial por razões neurológicas, emocionais e biomecânicas. 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..

A maioria dos investigadores considera como intervalo de referência 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; Guimarães, 2007; Brasolotto & Rehder, 2011):

- Acoplamento entre a região glótica e supra-glótica;

- Histologia da prega vocal e assimetria mecânica;
- Distribuição de muco sobre as pregas vocais durante a vibração;
- Pequena variação de massa ou de tensão nas pregas vocais;
- Sensação táctil laríngea.

Sugestões para 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, 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; Vieira et al., 1997 *in* Pinho et al., 2006). Deve ser indicado no protocolo de avaliação qual a usada;
- Os valores de *jitter* são influenciados pelo género e idade do falante: são encontrados valores mais elevados em crianças (Albertini et al., 2009);
- O *jitter* não indica as grandes variações da sustentação da frequência, mas sim a pequena variabilidade entre os sucessivos ciclos glóticos. Assim, o *jitter* é uma medida de aperiodicidade que diminui a confiabilidade quando a periodicidade aumenta (Laver, 1980);
- É aumentado artificialmente por factores 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 extracção de F0 (Pinho et al., 2006).
- 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*.
- 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).

- Para diminuir erros na demarcação dos períodos calcula-se o *jitter* não apenas baseado em dois ciclos, mas em três ou mais – medida de perturbação média relativa (RAP – *Relative Average Perturbation*).

***Jitter* nas perturbações vocais**

- Altera-se, principalmente, com a falta de controle da vibração das pregas vocais, como acontece nas disfonias de causa neurológica (Behlau, 2001);
- Existem diferenças estatisticamente significativas entre os valores de *jitter* de pacientes com disfonia *versus* eufonia quando se avaliam vogais sustentadas (Ma e Yiu, 2006; Zhang e Jiang, 2008), e em amostras de fala espontânea (Vasilakis, M. & Stylianau, Y., 2009);
- Existem diferenças ao longo do ciclo vital – com valores esperados de 0,59% ± 0,54% para adultos-jovens vs 2,10% ± 1,55% em homens idosos (presbifonia). Para as mulheres obtiveram-se 5,34% ± 4,51% e 2,02% ± 2,03%, respectivamente. (Xue e Deliyski, 2001)
- O Computer Speech Lab da Kay Elemetrics aponta como valores referência 0-2.1% para homens e 0-3.1% para mulheres (Wuyts et al., 1996).
- O *jitter* correlaciona-se com a aspereza e/ou rugosidade (Yumoto et al., 1984; Wolfe & Steinfatt, 1987; Dejonckere et al., 1993; Deal e Emanuel, 1978 *in* Yiu et al., 2000; Ma & Yiu, 2006; Koike, 1973 e Lieberman, 1961 *in* Pinho et al., 2006) e a soproidade (Eskenazi et al, 1990; Dejonckere, 1995).

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. Se o sistema fonador fosse completamente estável ele seria zero. 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:

$$S (\%) = |A(1) - A(2)| / (\frac{1}{2}[A(1) + A(2)]) \times 100$$

As investigações sobre esta temática sugerem que esta medida é inversamente proporcional à intensidade⁵ média, ou seja, quanto maior for esta, menor o valor de *shimmer*, e vice-versa. Oferece-nos uma percepção indirecta do ruído na produção vocal. Assim, nas perturbações da voz o *shimmer* surge mais alterado nas frequências graves e intensidade fraca.

As medidas relativas de *shimmer* podem ser apresentadas de diferentes formas (MEEI, 1994; Moran et al., 2006):

- *Shimmer* [%] (Shim): em percentagem (factor de perturbação direcciona), cujo valor limite usado habitualmente é 3.0% (Behlau, 2001), ou então 5.0% (0,44 dB) (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, isto é, mede o número de vezes que a diferença de amplitude entre ciclos consecutivos muda de direcção.

- *Shimmer* [dB] (ShdB): em dB, com o coeficiente logarítmico da amplitude de ciclos consecutivos, através da fórmula $S (sB) = 20 \log_{10} [A(1)/A(2)]$, em que os valores desta medida são 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 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 do desvio-padrão relativo da amplitude calculada período a período.

⁵ A intensidade vocal relaciona-se directamente com a pressão subglótica, a resistência das pregas vocais a essa pressão e a configuração do tracto vocal. (Boone e McFarlane, 2000; Colton e Casper, 1996).

- *Amplitude Tremor Intensity Index* [%] (ATRI): em percentagem, é calculada através do quociente 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 quociente 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 quociente 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 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;
- Há variações intrínsecas a cada vogal – deve ser indicado no protocolo de avaliação qual a usada;
- Os valores de *shimmer* são influenciados pelo género e idade do falante: são obtidos valores superiores nos homens (Albertini et al., 2009);
- 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;
- A forma de captação do sinal é importante e deve ser indicada.
- Há evidências práticas e científicas que indicam que o ambiente de gravação (número de interlocutores, tipo de comunicação necessária, ...) influenciam a intensidade usada (Behlau, 2001).
- Estudos sugerem que para obter este parâmetro é necessário um mínimo de 130 ciclos glóticos (Karnell, 1991 *in* Zhang e Jiang, 2008).

Shimmer nas perturbações vocais (Behlau, 2001; Pinho, 2003; Pinho et al., 2006)

- Altera-se, principalmente, nas situações de redução da resistência glótica (exemplo: paralisias/parésias das pregas vocais com variações na velocidade de fechamento glótico ou outras fendas glóticas);
- Também é influenciado por lesões de massa, mesmo as de tamanho reduzido;
- Existem diferenças ao longo do ciclo vital – com valores esperados de $2,52\% \pm 1,00\%$ para adultos-jovens vs $5,54\% \pm 3,51\%$ em homens idosos (presbifonia). Para as mulheres obtiveram-se $2,00\% \pm 0,79\%$ e $5,34\% \pm 4,51\%$, respectivamente (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%.
- Relaciona-se com o grau geral de disfonia (*grade*) e com a soproidade (Dejonckere et al., 1993), ou com a rugosidade (Eadie & Baylor, 2006);
- Parece ser menos sensível à associação com as perturbações vocais que o *jitter*.

HNR – PROPORÇÃO HARMÓNICO RÚIDO

A medida HNR (proposta por Yumoto & Gould, 1982 *in* Pouchoulin, 2008) é uma avaliação objectiva, isto é, de base matemática, que relaciona componente periódica e aperiódica (Guimarães, 2007) que compõem um segmento de fala sustentada (*cf.* Imagem 12).

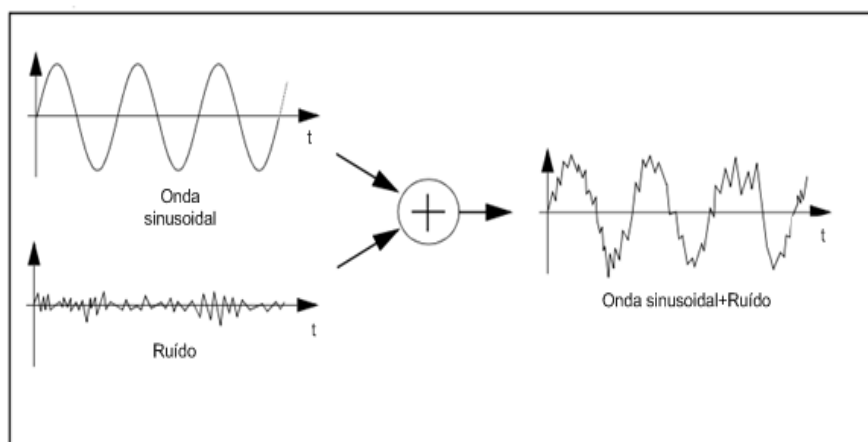


Imagem 12: Sinal de onda complexa – representação isolada da onda sinusoidal e de ruído e a conjugação dos dois componentes.

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

A relação entre as duas componentes 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, 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.

Tal como para as restantes medidas, existem diversas formas de cálculo matemático da proporção harmónico-ruído (Lopes et al., 2009), sendo a mais recente proposta por Krom (Krom, 1993) e subsequentemente modificado por Qi (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. Apesar de mais directa, esta abordagem é vulnerável à natureza dos sinais de voz e, em particular, os seus resultados dependem muito da frequência fundamental. Estes problemas foram subsequentemente minimizados em novos resultados publicados por Murphy (Murphy, 2007) (*in* Lopes et al., 2009).

Uma outra medida de ruído designa-se por NNE (Noise Normalized Energy – Energia Normalizada de Ruído), proposto por Kasuya et al., (1986 *in* Pouchoulin, 2008) que corresponde a uma média da relação ruído-harmónico (e não de harmónico-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 (Pinho et al., 2006).

Sugestões para obtenção de medidas de HNR (Behlau, 2001)

- Os valores das medidas de ruído são apresentados em número de dB;
- O componente ruído é tão maior quanto menor os componentes harmónicos na emissão;
- O HNR é menor nos homens e maior nas mulheres (Behlau, 2001), provavelmente devido à fenda glótica posterior (fisiológica) das mulheres (inerente a factores anatómicos do arcaboço laríngeo) que acarreta maior turbulência glótica aquando da produção de voz;
- O HNR é maior no registo falsete⁶, seguido pelo modal⁷ e, finalmente, o basal⁸.

HNR nas perturbações vocais (Behlau, 2001)

- Indivíduos com grandes lesões de massa, difusas, podem apresentar valores de HNR muito altos;
- Pequenas fendas glóticas podem resultar num HNR baixo, sem possibilidade de correlação com o grau de disфонia ou a avaliação áudio-perceptual;
- Valores de HNR inferiores a 7dB são, necessariamente, patológicos. Contudo, os valores de referência são variáveis (=22 dB para Vieira, 1997 *in* Pinho et al., 2006), de acordo com o *software* de análise acústica em uso.

⁶ Por registo de falsete entende-se o tipo de fonação, ocorrido na fala normal, mas com tendência para o registo mais elevado (Fawcus, 1991 *in* Freeman e Fawcus, 2004)

⁷ O registo modal, descrito por Hollien (1974 *in* Freeman e Fawcus, 2004) ocorre com mais frequência na fala normal e decorre da adução glótica completa, com rigidez suficiente para interromper momentaneamente o fluxo de ar pulmonar. Tal resulta na sucessão de impulsos glóticos que decorrem, nos adultos do sexo masculino, por volta dos 100Hz; e, no sexo feminino e crianças em torno dos 200Hz (*idem*).

⁸ Também designado por *glottal fry* ou registo de pulso (Hollien, 1974 *in* Freeman e Fawcus, 2004) ocorre em frequência mais graves que a voz modal e caracteriza-se por um índice de vibração relativamente aleatório.

- Vários autores encontraram forte correlação entre o HNR e o parâmetro rugosidade (Eskenazi et al., 1990; Krom, 1993; Martin et al., 1995) e soproidade (Krom, 1995; Kojima, Gould, Lambinanse and Isshiki, 1980; Lee & Childers, 1991; Yumoto, Sasaki and Okamura, 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 (Brasolotto & Rehder, 2011):

- Vogal sustentada para extracção da frequência fundamental (F0), desvio padrão da F0, *jitter*, *shimmer* e medidas de ruído; é aconselhável obter 3 amostras;
- Fala encadeada para extracção da frequência fundamental média, mínima e máxima;
- Emissão de fala encadeada e vogal sustentada em frequências 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 o perfil de extensão da fala.
- Emissões repetidas o mais rápido possível de / a / e/ou / i /, como prova de diadococinésia laríngea (tarefa de coordenação, alternância ou dissociação de movimentos glóticos).

1.6. REPRODUTIBILIDADE DOS RESULTADOS DA AVALIAÇÃO ACÚSTICA

O uso clínico de um determinado teste ou escala depende da reprodutibilidade dos seus parâmetros. Este aspecto sujeita-se a três factores:

- A variabilidade inter-observador, isto é, a variação de julgamentos entre diferentes softwares de avaliação com recurso ao método acústico.
- A variabilidade intra-observador, ou seja, a variação de julgamentos de um mesmo avaliador ao longo do tempo. Este aspecto não se coloca no uso de softwares de avaliação acústica, uma vez que o algoritmo de análise é constante num mesmo programa.
- A variabilidade intra-sujeitos, entendida como a variação da qualidade vocal do doente em diferentes momentos temporais da recolha e análise da mesma. Este aspecto pode estar associado a factores intrínsecos (emocionais, cansaço) ou extrínsecos, que se não forem controlados condicionam os resultados da avaliação através do método acústico.

Este sub-capítulo resulta de uma pesquisa na Pubmed que possibilitou a revisão da literatura publicada (1950-Agosto 2009) sobre os resultados da avaliação acústica – formas de uso e respectiva validade dos resultados (*cf.* Quadro 2). Os termos de pesquisa usados foram (voice quality OR voice disorders) AND (reproducibility of results OR retest OR variability OR variation OR repetability OR intrasubject OR intra-subject).

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 1 de Agosto/2009
- Relativos a estudos com a qualidade vocal de humanos (objectiva e subjectiva)
- 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: 51

Total de artigos sobre avaliação perceptual: 39

Total de artigos sobre avaliação acústica: 12 (Quadro 2)

Os doze artigos que se enquadraram nos critérios de inclusão foram publicados num período de 16 anos (entre 1989 e 2004, Quadro 2). Aparentemente as medidas objectivas de quantificação das perturbações da voz são usadas há cerca de 20 anos, porém são poucas as publicações que analisam a validade e consistência interna destas medidas (12 estudos).

Ao tentar esboçar uma comparação entre os resultados dos estudos, deparamo-nos com as seguintes dificuldades: as medidas/parâmetros analisados são muito díspares, embora com um claro predomínio das medidas de perturbação; os softwares usados nas publicações são também diferentes e com algoritmos de quantificação das medidas nem sempre conhecidos; a concepção metodológica das investigações é igualmente muito variável.

A maioria das investigações analisadas centra os resultados na interpretação da consistência da medida/parâmetro em estudo (Higgings & 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). 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 um dos estudos (Gramming et al., 1991).

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 controles (Carding et al., 2004).

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. Esta falta de uniformidade pode ser ainda a razão que justifique 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 factor moderado-pobre. Nenhum dos estudos analisados centra muita atenção nas diferenças inter-observadores, 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 inter-observadores.

Quadro 2: Descrição sumária dos estudos sobre reprodutibilidade dos resultados da avaliação acústica da voz.

Referência Bibliográfica	N Repetições C=Pacientes Ct=Controles	Intervalo de tempo entre medições	Parâmetros	Análise Estatística	Resultados	Análise do Artigo/Conclusões
Bough, I.D. Jr. et al. (1996) – <i>Intrasubject variability of objective voice measures.</i>	C: 14 (Sonograph e Visi-Pitch)	15 amostras em 15 dias	FO Jitter Shimmer HNR Perturbação	Coeficiente de correlação interclasses	No mesmo dia: 0.041-0.999 Entre os dias: 0.106-0.994	A utilidade das medidas objectivas de avaliação dependem da consistência e validade do instrumento usado para as obter.
Carding, P.N. et al. (2004) – <i>The reliability and sensitivity to change of acoustic measures of voice quality.</i>	C: 181 Ct: 50 (MDVP)	2 horas	Jitter Shimmer NHR	Coeficiente de correlação interclasses	C: 0.33-0.46 Ct: 0.68-0.73	A medição não pode ser aplicada clinicamente. Consistência pobre a moderada.
Dwire, A. & MaCauley, R. (1995) – Repeated measures of vocal fundamental frequency perturbation obtained using the Visi-Pitch.	C: 49 (Visi-Pitch)	1 semana	RAP	Coeficiente de correlação de Pearson	Homens: 0.83-0.90 Mulheres: 0.36-0.58	A relação entre os valores das medições entre sessões para os indivíduos da amostra foi marcada para homens mas apenas moderada para as mulheres.
Gelfer, M.P. (1989) – <i>Stability in phonational frequency range.</i>	C: 20 (MDVP)	3 horas em 3 meses	Gama de frequência fonatória	Análise de variância Média individual Variância para o	Varição: FO_baixa: 2.15 st FO_alta: 2.5 st FO_gama: 3.6 st	Todas as medidas variaram bastantes semitons. Sugerem mais investigação para

				desvia padrão		determinar a fonte da variabilidade intra-sujeitos antes de aplicar estas medidas na clínica.
Gramming, P. et al. (1991) – <i>Variability of phonetograms.</i>	C: 2	15x Em 3 semanas (2 investigadores)	Fonetograma	Média, Desvio padrão (suave e forte). Coeficiente de correlação	Mulheres: 2.7 e 2.4 dB Homens: 3.0 e 2.4 dB Suave: 0.998 Forte: 0.985	A utilização de dois avaliadores não revelou grande impacto nos resultados do estudo. A variabilidade vocal diária ou variação SPL foi de aproximadamente 3 dB, na fonação suave e forte. Atendendo a estes efeitos pequenos, o fonetograma para ser uma ferramenta útil na clínica.
Higgins, M.B. & Saman, J.H. (1989) – <i>A comparison of intrasubject variation across sessions of three vocal frequency perturbation indices.</i>	C: 15 (MDVP)	17 x Em 33 dias	JF PPA DPF	Coeficiente de variação	Homens: 11-46 Mulheres: 31-62	Observou-se uma variabilidade considerável entre sessões.
Higgins, M.B. et al. (1994) – <i>Aerodynamic and electroglottographic</i>	Ct: 21	4x Em 2 semanas	Respiração, Pressão intra-oral, sinal EGG, ...	Coeficiente de variação	Média intra-sujeitos: 2-22	São dados valores de referência para corte que devem ser considerados como

<i>measures of normal voice production: intrasubject variability within and across sessions.</i>						limites de normalidade vocal.
Kania, R.E. et al. (2004) – <i>Variability of electroglottographic glottal closed quotients: necessity of standardization to obtain normative values.</i>	Ct: 20	3 amostras de cada sujeito Não é referido se repetem	GCO	Análise de variância para medidas repetidas	O efeito do tempo da medida não é significativo	A variabilidade do GCO com a F0 e a intensidade contribuem para a consistência e para os problemas de standardização do GCO como uma medida de encerramento glótico
Lee, L. et al. (1999) – <i>Consistency of acoustic and aerodynamic measures of voice production over 28 days under various testing conditions.</i>	Ct: 70 divididos por 3 grupos/provas (MDVP)	28 dias	F0, jitter, TMF, Volume fonatório, Taxa expiratória	Análise de variância	São apresentados resultados para todos os parâmetros	Manter uma intensidade e frequência constantes ao longo dos testes ajuda a assegurar similitudes entre as condições de avaliação. Quando estas variáveis se mantêm constantes, as mudanças nas medidas acústicas e aerodinâmicas podem ser atribuídas com mais consistência à patologia de base
Mendoza, E. et al. (1996) – <i>The long-term</i>	Ct: 17	5x leram texto standard	LTAS	Análise de variância	Valores absolutos:	O facto das diferenças entre sessões

<i>average spectrum as a measure of voice stability.</i>		Em 2 semanas		unifactorial	diferença entre sessões significativa. Valores relativos: não significativos	desaparecerem quando se usam medidas relativas pode dever-se ao facto de a utilização das mesmas eliminar as fontes de erro sistemático ou aleatório introduzido durante a gravação ou no espaço de tempo entre duas sessões consecutivas.
Stassen, H.H. (1991) – <i>Affective state and voice: the specific properties of overtone distributions.</i>	C: 187	14 dias	Padrões espectrais	Reprodutibilidade	Localização e intensidade dos três primeiros formantes: reproduzidos exactamente “Bandwith” e “height” dos formantes: altamente reprodutível	Os valores de referência resultam de uma amostra de 187 sujeitos saudáveis, para assim obterem a clara distinção entre flutuações “naturais” e “significativas”.
Stone, R.E. Jr. & Rainey, C.L. (1991) – <i>Intra- and intersubject Variability in Acoustic Measures of Normal Voice.</i>	Ct: 24 (Visi-Pitch)	3x 2 meses	F0 + desvio padrão dBA + desvio padrão Jitter	Coefficiente de correlação	0.13-0.82	A variabilidade apresentada pelos sujeitos ao longo das semanas foi desencorajadora para os investigadores. O

						aperfeiçoamento das ferramentas de análise vocal deve preceder o seu uso na avaliação e intervenção terapêutica.
--	--	--	--	--	--	--

Legenda:	
FO – Frequência Fundamental	DPF – Directional Perturbation Factor
HNR – Harmonic-to-Noise-Ratio	EKG – Electroglotografia
NHR – Noise-to-Harmonic-Ratio	GCQ – Glottal Closed Quotient
RAP – Relative Average Perturbation	TMF – Tempo Máximo de Fonação
JF – Jitter Factor	LTAS – Long-Term Average Spectrum
PPQ – Pitch Perturbation Quotient	dBA – Curva de Ponderação A

Capítulo 2 – CORRELAÇÃO ENTRE MEDIDAS DE AVALIAÇÃO ACÚSTICA E PERCEPTUAL

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, 2004; Guimarães, 2007).

Neste sentido torna-se importante realizar análises multifactoriais 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 disfónico 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; Carding, 2004; Ma & Yiu, 2006; Martens et al., 2007; Behlau, 2010).

Saliente-se, então, que nenhuma forma de avaliação substitui outra – todas são complementares e construtivas no processo terapêutico (Colton e Casper, 1996; Andrada 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 objectivos, os parâmetros acústicos não invalidam a avaliação áudio-perceptual por duas razões fundamentais: as dimensões de apreciação de qualidade de uma voz (quando caracterizada perceptualmente) 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). Por outro lado, a correlação entre os resultados destas duas dimensões não é nem óbvia nem directa, sendo ainda matéria de investigação e debate, o que denota a dificuldade clássica que existe em

expressar a acuidade auditiva humana através de modelos matemáticos (Gregio et al., 2006; Maryn et al., 2009).

A análise dos parâmetros vocais alterados torna-se importante no processo de diagnóstico de uma perturbação da voz e condicionam todo o processo de intervenção terapêutica (Awan & Roy, 2009; Camargo & Madureira, 2010). Aquela pode ser avaliada de forma subjectiva (avaliação perceptivo-auditiva) e/ou objectivamente com o auxílio de equipamentos de análise acústica. Note-se que nos diferentes estudos de correlação entre as medidas subjectivas e instrumentais a percentagem de concordância pode variar entre 49,9% (Wuyts *et al.* 2000) e 86,0% (Yu *et al.* 2001).

Vários estudos investigaram a relação entre as medidas acústicas isoladas e a avaliação perceptual. Os seus resultados foram inconclusivos pois não revelaram correlações estatisticamente significativas (Dejonckere, 1996; Giovanni, 1999; Heman-Ackah, 2002, 2003; Morsomme, 2001; Carding et al., 2004; Rabinov, 1995 *in* Hakkesteegt *et al.*, 2008; Ma e Yiu, 2006; Kreiman & Gerratt, 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; Behlau, 2001; Parsa & Jamieson, 2001; Carmargo & Madureira, 2010). Destaca-se mais uma vez a importância da abordagem integrada, no sentido de considerar múltiplos recursos de avaliação disponíveis (Sader & Hannayama, 2004).

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 tracto vocal não é nem linear, nem passiva (Dajer, 2010). 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 indirectamente o modelo acústico com o auditivo. As medidas ou parâmetros de avaliação propostos pelo método acústico acabam, então, por

contemplar apenas uma fracção da totalidade das apreciações ou julgamentos passíveis ao ouvido humano (Eskenazi & Childers, 1990), o qual assume um carácter multidimensional e consegue aglutinar características áudio-perceptuais (Parsa & Jamieson, 2001; Patel & Shrivastav, 2007; Harnisch et al., 2008).

Correlacionar as medidas acústicas com as medidas perceptivo-auditivas e, ainda, com os mecanismos fisiológicos da produção vocal é uma tarefa complexa (Eskenazi et al., 1990; Kreiman et al., 1993; Dejonckere, 1995; Dejonckere & Lebacqz, 1996; Frolick et al., 2000). Isto porque se crê que uma dimensão ou parâmetro específico da qualidade vocal possui uma ou mais pistas acústicas para a sua percepção.

Actualmente tenta-se que a conjugação entre medidas acústicas possibilite a quantificação de aspectos da produção vocal. É pela combinação de vários parâmetros objectivos que se obtém uma correlação mais forte com a análise perceptual (Wolfe & Martin, 1997; Michaelis, 1997, 1998; Piccirillo, 1998; Klein, 2000; Wuyts, 2000; Yiu *et al.*, 2000; Parsa & Jamieson, 2001; Yu, 2001, 2002; Hartl, 2003; Kent et al., 1994 *in* Murdock, 2005; Patel & Shrivastav, 2007; Schutte, 1983 *in* Hakkesteegt *et al.*, 2008).

Recentemente os autores começaram a propor o estudo de três medidas derivadas de processos de modelação periódica linear (Qi et al., 1999): 1) SNR – *signal-to-noise ratio*; 2) PA – *pitch amplitude*; e, 3) SFR – *spectral flatness ratio*. Tanto o PA como o SFR mostraram a possibilidade de diferenciação entre vozes normais e patológicas, em tarefa de vogal sustentada e conversação espontânea. Hillenbrand et al. (1996) e seguidores notaram que as medidas de cepstrum se correlacionam com as medidas perceptivas de grau geral de alteração vocal e soproidade de forma mais forte face às restantes medidas.

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 perceptuais (Finizia et al., 1999; Bunton et al., 2007; Kreiman et al., 2007; Patel & Shrivastav, 2007).

São 20 os estudos recentes mais significativos nesta temática:

Wolfe et al., 1995, caracterizaram 4 medidas acústicas (frequência fundamental, *jitter*, *shimmer* e HNR), num estudo com 20 controlos (vozes “normais”) e 60 casos de pacientes disfónicos (lesões nodulares, paralisias de corda unilaterais e disfonias funcionais). Através de uma análise de regressão a correlação entre os parâmetros acústicos e a análise áudio-perceptual foi de $r=0.56$. A análise factorial da correlação apresentou $r=0.54$ entre o *shimmer* isolado e a avaliação perceptiva.

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 perceptiva, em 5 níveis – desde o 0=normal a 4=severo. O tratamento estatístico com a análise de uma função de entrada directa discriminante revelou que a combinação das 4 medidas conseguia um grau de concordância com a classificação perceptual de 66,1%. Salientaram que esta concordância foi estatisticamente significativa para todos os níveis, excepto para o grau 1=ligeiro/intermitente de alteração vocal.

Michaelis et al., 1998, construíram um diagrama de rugosidade com base na combinação de factores acústicos, analisado a partir de uma base de dados de 88 amostras eufónicas 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 2 estudos na tentativa de construir um índice multiparamétrico que correspondesse à severidade da disфонia. Assim, através de uma análise de regressão multivariada identificaram 14 medidas básicas capazes de distinguir entre vozes eufónicas e disfónicas. Concluíram que a combinação entre 4 delas – estimativa de pressão subglótica, a gama de variação da frequência, o ratio 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 controles (vozes “normais”) e 319 indivíduos disfônicos, classificados perceptualmente 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*)⁹.

O *Dysphonia Severity Index* é definido como uma combinação ponderada de 4 medidas acústicas seleccionadas por uma análise multivariada de 13 parâmetros. A análise discriminativa da combinação dos 4 parâmetros finais – *jitter*, tempo máximo de fonação do /a/, pico de frequência e mínimo valor de intensidade – revelou uma concordância de 49,9% (193 em 387 sujeitos) com a avaliação perceptiva.

Yu et al., 2001, fizeram a análise multiparamétrica de uma amostra de 63 vozes masculinas disfônicas e 21 controles eufônicos, avaliados perceptualmente através do parâmetro G(rau) da escala GRBAS, por um painel de juízes apoiado num software de apoio à decisão. Caracterizam com o EVA® 10 parâmetros: frequência fundamental, *jitter*, intensidade, HNR, HNR (f>1KHz), coeficientes de Lyapounov, débito de ar oral, extensão vocal e tempos 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 correlação significativa entre os julgamentos perceptivos e seis dos diferentes parâmetros. Especificamente a combinação frequência fundamental, HNR, coeficientes de Lyapounov, pressão sub-glótica, 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, excepto no corpus de análise – um total de 74 vozes femininas, 6 eufônicas e 68 disfônicas. Os resultados revelam uma correlação entre avaliação perceptiva e acústica da seguinte magnitude:
- 64% para a classificação do parâmetros G(rau) com uma escala analógica;

⁹ A fórmula de cálculo do DSI = $0.13 \times \text{TMF} + 0.0053 \times \text{F0-High} - 0.26 \times \text{I-low} - 1.18 \times \text{Jitter}(\%) + 12.4$

- 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 do “Rainbow Passage”) ou vogal sustentada. Usaram dois juízes com experiência que classificaram as amostras com a escala GRBAS. A análise de consistência inter-avaliadores 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) ($r=-.86$, $r^2=.74$, na fala; $r=-.80$, $r^2=.64$, na vogal) e S(Soprosidade) ($r=-.71$, $r^2=.51$, na fala; $r=-.70$, $r^2=.54$, na vogal). O CPPS correlaciona-se inversamente ($r=-.50$, $r^2=.25$) e o sPPQ positivamente ($r=.40$, $r^2=.16$) 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) ($r=.53$, $r^2=.28$; $r=.54$, $r^2=.29$; $r=.60$, $r^2=.36$; $r=.57$, $r^2=.32$, respectivamente) e S(Soprosidade) ($r=.54$, $r^2=.30$; $r=.52$, $r^2=.27$; $r=.54$, $r^2=.30$; $r=.53$, $r^2=.28$, respectivamente). NHR, APQ e RAP não se correlacionaram com a percepção de R(Rugosidade).

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 perceptualmente 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 de G foi 0.43, para o R foi 0.14, para o B foi 0.32, para A 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 & Baylor, 2005, estudaram a correlação 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-perceptual do grau geral de perturbação vocal e “agradibilidade” (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 disfônicos e 6 eufônicos. 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 “agradibilidade”. A medida acústica isolada com a associação mais forte à avaliação perceptual 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 disфонia.

Eadie & Baylor, 2006, analisaram os resultados da classificação perceptual através da escala visual analógica (VAS) – parâmetros grau, rugosidade e soproidade – de 36 vozes (21 disfônicas 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 softwares (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 soproidade 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 correlação (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 de rouquidão 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 soproidade e rugosidade. Concluíram que o tempo máximo de fonação, o pico de pressão intra-oral na consoante-vogal /pi/, a área do fonetograma e o *jitter* prediziam correctamente 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%, respectivamente) em comparação com as vozes ligeira a moderadamente disfónicas (67,9 e 36,0%, respectivamente).

Martens et al., 2007, estudaram a classificação audio-perceptual de 6 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 (fonetograma) de apoio à avaliação áudio-perceptual. 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 perceptivos 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 (0.58) e o R-Rugosidade (0.57). Neste estudo o HNR não se relacionou com nenhum dos parâmetros perceptuais estudados.

Awan & Roy, 2009, analisaram 176 vozes (porção central de vogal sustentada /α/) 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 actual da CPP (*cepstral peak prominence*) face à amplitude esperada (CPP/EXP) – que faziam parte da equação do *Predicted Dysphonia Severity*¹⁰ (PDS, Awan & Roy, 2006)); e, com um programa computadorizado específico de classificação perceptual da severidade da alteração vocal, 10 alunas de mestrado em Terapia da Fala quantificavam numa escala com 7 níveis de EAI as amostras, divididas por duas sessões (88 vozes cada), 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=.906, R²=.82, p<.0001). Isoladamente, cada medida da PDS também apresentou uma correlação entre a percepção de severidade de alteração da voz forte e inversa com INVSQRTSIG (r=-.819, r²=.67) e CPP/EXT (r=-.754, r²=.80), e forte e positiva com o LOGSHIM (r=.814, r²=.77). A correlação com o DFTR foi moderada (r=-.511, r²=.82). Note-se que a medida apresentada por estes investigadores revelou-se robusta na distinção entre vozes ligeira e moderadamente alteradas (níveis intermédios da escla perceptual usada), assim como a analisar as amostras vocais extremamente alteradas.

Maryn et al. (2009) levaram a cabo uma meta-análise com a revisão da literatura publicada acerca da relação entre as medidas acústicas e classificação áudio-perceptual da qualidade vocal. Assim, analisaram um corpus de 25 estudos (cerca de 21 com

¹⁰ A fórmula de cálculo do PDS= 6.923+1.394 (LOGSHIM) - 1.002 (INVSQRTSIG) – 0.030 (DFTR) – 0.19 (CPP/EXP)

medidas de vogal sustentada e 7 de fala encadeada; 3 apresentavam os dois tipos de amostra vocálica) os quais apresentavam um total de 87 marcadores acústicos (os estudos com vogal sustentada identificavam individualmente 69 medidas; os artigos com fala encadeada associavam-se a 26 parâmetros acústicos). A medida de corte para a determinação do grau de correlação foi o valor de coeficiente de correlação $r_w > 0.60$. Deste modo, a associação entre percepção de alteração vocal associava-se: a) em amostras de vogal sustentada com – r de Pearson na *autocorrelation peak*, amplitude do *pitch*, *spectral flatness* do sinal residual e SCPP – *smoothed cepstral peak prominence*; b) em tarefas de avaliação de fala encadeada com – *signal-to-noise ratio from Qi*, *cepstral peak prominence* e SCPP - *smoothed cepstral peak prominence*.

AVALIAÇÕES ACÚSTICAS E PERCEPTUAIS DE PATOLOGIAS/TAREFAS VOCAIS ESPECÍFICAS (alguns SEM CORRELAÇÃO):

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: 1) os valores da medida acústica em análise nas vozes normais se encontra entre 7.0 e 17.7dB; 2) 15 das 18 amostras de qualidade vocal alterada no pré-operatório – 84.6% - estavam fora deste intervalo; 3) 95.7% das vozes no pós-operatório incluíam-se na gama de normalidade prevista, o que indicava a eficácia 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 de sujeitos – 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 perceptual 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 ($r=0.59$), 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 corticoesteróides e fizeram o seu emparelhamento quanto ao género com a mesma quantidade de controlos. Analisaram perceptualmente a voz através da escala GRB classificada por 4 juizes 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 sopradas e roucas dos asmáticos. 62,5% dos casos foram classificados no grau moderado a severo de disфонia.

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 da Kay Elemetrics. Concluíram que todos os pacientes apresentavam algum grau de disфонia (G=Grade), com alteração mais marcada da rugosidade e tensão. A soproside 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 um 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 cricoaritenóideu lateral com Electromiografia).

Após a revisão e análise da bibliografia descrita neste capítulo reforça-se a percepção de que há muita dificuldade em associar directamente as medidas acústicas às classificações áudio-perceptuais da voz patológica. Apesar de ser cada vez mais fácil

aceder a softwares de análise acústica, nem todas as medidas oferecidas têm uma representatividade clínica eficiente e robusta (Maryn et al., 2009). Note-se, ainda, que a capacidade de avaliação áudio-perceptual incorre em muitos viés, os quais também condicionam negativamente o resultado final apresentado (Awan & Lawson, 2009).

Capítulo 3 – PLANO DE TRABALHOS DA DISSERTAÇÃO DE DOUTORAMENTO

3.1. PARTICIPANTES E MÉTODOS DE RECOLHA DOS DADOS PARA A TESE

O estudo de campo a desenvolver ao longo da elaboração da dissertação inclui 3 etapas distintas que serão descritas de seguida:

- 1) Estudo retrospectivo, observacional, descritivo, do tipo série de casos – para recolher as 14 vozes-treino (“âncora”) e as 36 vozes-teste de pacientes de uma base de dados da doutoranda (total de 45 vozes + 10% repetição). Pretendem-se amostras com diferentes graus de severidade de alteração vocal: nas vozes-treino os três parâmetros terão, cada um, quatro representantes vocais de severidade distintas e 10% de repetição; nas vozes-teste este número aumenta, por forma a que o tamanho amostral permita um maior poder estatístico do teste. As gravações das vozes das bases de dados de treino e de teste foram feitas com uma frequência de amostragem de 44100Hz, e uma resolução de 16bits, para um computador com um Processador Pentium IV e utilizando um microfone de mesa da marca Philips SBC ME 400, unidireccional (cardióide), numa sala com um nível de ruído inferior a 40dB SPL. A distância do microfone à boca foi fixa a 10 cm, tendo sido pedido ao paciente para produzir de forma sustentada e confortável a vogal / α / (Heman-Ackah et al., 2002; Awan & Lawson, 2009; Awan & Roy, 2009), durante pelo menos 5 segundos (Smits et al., 2005; Awan & Lawson, 2009; Awan & Roy, 2009), em duas tentativas. A última foi aquela usada para estudo. Foi analisado a porção de sinal a partir do 2º segundo da amostra (Ackah et al., 2002; Smits et al., 2005; Awan & Roy, 2009).

A avaliação perceptual usada foi adaptada pela doutoranda das escalas revistas no capítulo anterior e tendo em consideração a reprodutibilidade dos parâmetros nos

estudos enumerados, assumindo o nome de GRB: G=grau, R=rugosidade, B=soprosidade (Comitê de Foniatria da Sociedade Europeia de Laringologia (ELS) – Dejonckere et al., 2001). A descrição de cada parâmetro é a seguinte:

Parâmetros	Definição
<i>G – Grau (Grade)</i>	Grau de alteração vocal – impressão global da voz, identificação do grau de alteração vocal, como um todo.
<i>R – Rugosidade (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. A voz é percebida com ruídos inesperados produzidos a baixa frequência.
<i>B – Soprosidade (Breathiness)</i>	Presença de turbulência ou ruído audível, escape de ar importante através da glote, sensação de ar na voz.

Estas definições serão facultadas aos juizes e avaliadores e, então, consideradas “âncoras” ou pistas escritas à classificação.

A avaliação áudio-perceptual para a classificação e construção das bases de dados – 50 vozes (de treino e de teste) – serão apresentadas segundo uma ordem aleatória, escutadas e analisadas por um painel de quatro peritos com mais de 10 anos de experiência e treino neste tipo de avaliação. Foi considerada a cotação dada por pelo menos 3 dos avaliadores (minimizando o erro de consistência inter-avaliadores), com base no teste de proporções.

Cada estímulo será apresentado repetidamente, com um intervalo de 2 segundos, para proceder à classificação dos três parâmetros perceptivos, numa escala analógica visual (VAS – Visual Analogue Scale), assinalando com uma cruz sobre uma linha 10cm, sendo que quanto mais à direita, mais alterada será a qualidade vocal. Quando todos os estímulos tiverem sido classificados, o avaliador pode corrigir qualquer das respostas dadas. A resposta final fica registada

automaticamente numa base de dados construída para o efeito. A ordem de apresentação dos estímulos foi determinada de forma aleatória, para evitar os efeitos de familiaridade. Existem 45 vozes diferentes e 5 (10%) repetidas.

2) Estudo transversal, quasi-experimental, descritivo:

- a. Após seleccionar um grupo ilustrativo de vozes patológicas portuguesas para criar uma base de dados de referência – seleccionada e classificada por um painel de peritos (quatro) – para a ilustração do tipo e severidade de cada um dos três parâmetros perceptuais da escala GRB, classificados numa escala analógica visual (VAS – Visual Analogue Scale) de 100mm. Tendo em vista que estas serão classificadas por sujeitos/avaliadores e pretendemos evitar o cansaço dos mesmos, não ultrapassaremos as 14 vozes em cada momento de treino e 36 nos de teste.

Os resultados da avaliação perceptiva – das vozes-treino e das vozes teste – por um grupo de quatro juizes com especialização em voz e/ou pelo menos 10 anos de experiência nesta área de actuação profissional, serão estudadas em termos estatísticos, servindo como pré-teste para a análise e interpretação dos dados obtidos na etapa 2) b. da dissertação. Será encontrada a pontuação média de cada parâmetro, para cada amostra de voz e para cada perito, encarada como a variável dependente deste estudo. A taxa de semelhança entre avaliações dos peritos será medida pelo coeficiente de Coehn's kappa Fleiss ou o α de Krippendorff. Para determinar a medida final de cada parâmetro será necessária a mesma avaliação áudio-perceptual por, pelo menos, três dos avaliadores. As vozes-treino serão avaliadas, descritas e classificadas – S.D. F0, *Jitter*, *Shimmer*, HNR – por quatro softwares de análise acústica (*opensource* e comercializados), o que também servirá de pré-teste para a análise e interpretação dos dados obtidos na etapa 2) c. da dissertação.

b. Esta amostra de 14 vozes-treino (“âncora” auditiva) e 36 vozes-teste será cotada – através da escala GRB (Dejonckere et al., 2001) – por uma amostra significativa e válida de juízes de 2 tipos. O primeiro grupo é constituído por 30 alunos de Terapia da Fala/Fonoaudiologia (falantes nativo do Português; com pelo menos um semestre da unidade curricular de “Voz”; sem historial de perturbações auditiva ou alterações da acuidade auditiva à data do estudo; falantes nativos do Português Europeu ou Brasileiro); o segundo grupo é constituído por 30 Terapeutas da Fala/Fonoaudiólogos com especialização/experiência em voz (sem historial de perturbações auditiva ou alterações da acuidade auditiva à data do estudo; falantes nativos do Português Europeu ou Brasileiro). Nos dois grupos realizar-se-ão os seguintes procedimentos de investigação:

Sessão 1: Apresentação do estudo e descrição teórica da escala perceptual usada, ministrada pela doutoranda. Escuta e classificação das 36 vozes-teste no final da sessão, num documento impresso construído para o efeito (com uma tabela de dupla entrada), que inclui pequeno resumo descritivo da escala usada e respectiva forma de cotação (“âncora” visual).

Sessão 2: (7 dias após a sessão1) Descrição teórica da escala perceptual usada, ministrada pela doutoranda. Escuta, análise e classificação de vozes-treino – 14 vozes pré-seleccionadas, 7 masculinas e 7 femininas – apresentadas sucessivamente, de uma só vez, com pequeno intervalo temporal (2seg) entre amostras para possibilitar a classificação. Escuta e classificação das 36 vozes-teste, no final da sessão, num documento impresso construído para o efeito (com uma tabela de dupla entrada), que inclui pequeno resumo descritivo da escala usada e respectiva forma de cotação (“âncora” visual).

Sessão 3: (15 dias após a sessão 2) Escuta, análise e classificação de vozes-treino – 14 vozes pré-seleccionadas, 7 masculinas e 7 femininas – apresentadas sucessivamente, de uma só vez, com pequeno intervalo temporal (2 seg) entre amostras para possibilitar a classificação. Escuta

e classificação das 36 vozes-teste no final da sessão, num documento impresso construído para o efeito (com uma tabela de dupla entrada), que inclui pequeno resumo descritivo da escala usada e respectiva forma de cotação (“âncora” visual).

A classificação das vozes-treino e teste será feita para cada voz com um intervalo fixo de 2 segundos entre os estímulos. No final da classificação dos três parâmetros de uma dada amostra há a possibilidade de ajuste da mesma. Assim, o número de apresentações de um dado estímulo é, à partida, ilimitado. Os juízes têm a informação permanente – escrita na folha de registo – do significado teórico de cada um dos parâmetros de qualidade vocal a avaliar (“âncora” visual). Podem também realizar perguntas ao longo do teste. Espera-se que para as 14 vozes-treino e 36 vozes-teste sejam necessários cerca de 30 minutos de atenção auditiva e classificação.

- c. A amostra de 36 vozes-teste será avaliada, descrita e classificada – S.D. F0, *Jitter*, *Shimmer*, HNR – por quatro softwares de análise acústica (*opensource* (PRAAT) e comercializados (VoiceStudio da Seegnal; Voxmetria da CTS Informática; CSL da Kay Elemetrics)). Será levado a cabo o estudo da consistência e correlação inter-medidas dos diferentes parâmetros acústicos, pelos quatro programas.
- d. Por último pretendemos descrever e discutir os resultados da correlação (regressão logística) entre a avaliação acústica e perceptual da base de dados de vozes-teste, com base na análise de regressão multivariada entre as quatro medidas acústicas e os três parâmetros áudio-perceptuais, medindo o grau de significância para os primeiros predizerem os segundos e vice-versa.

O objectivo deste segundo momento é verificar se:

- Existe consistência intra-avaliadores áudio-perceptuais (teste-reteste), num grupo de juízes treinados mas sem experiência vs um grupo de juízes com experiência e treinados, em avaliações realizadas em três momentos temporais distintos?
- Existe consistência inter-avaliadores áudio-perceptuais, num grupo de juízes treinados mas sem experiência vs um grupo de juízes com experiência e treinados?
- Existe consistência inter-avaliadores das medidas acústicas, entre cada um dos softwares de análise utilizados?
- Existe correlação entre as medidas perceptivas e acústicas?

3) O resultado prático final, para além da dissertação, pressupõe construir material de suporte pedagógico ao treino de avaliação perceptual (*Voice-PE – Voice – Perceptual Evaluation*), que se colocará à disposição de futuros e actuais terapeutas que, deste modo, terão acesso a mais um material de treino (com “âncoras” auditivas e visuais) e preparação para uma mais eficaz avaliação e intervenção terapêutica com casos de patologia vocal. Esse material implica a colaboração entre a Doutoranda e um Engenheiro de Software, que consiga criar esta ferramenta informática que facilite a interacção com o programa.

3.2. TRATAMENTO ESTATÍSTICO

Para o tratamento estatístico dos dados da dissertação serão usados: o Índice de Concordância Kappa, os testes de Mann-Whitney, de Kruskal-Wallis, Igualdade de Duas Proporções, Qui-Quadrado, Correlação de Spearman e de Pearson, e o Teste de Correlação através da análise de Regressão Logística. O nível de significância adotado será de 0,05 (5%). Os intervalos de confiança serão construídos com 95% de confiança estatística.

Na avaliação áudio-perceptual – tanto do painel de peritos como dos grupos de estudo (treino e teste) – as concordâncias intra e interavaliadores serão medidas pelo Índice de Concordância Coehn`s kappa Fleiss ou o α de Krippendorff, que é uma análise utilizada para medir o grau de concordância entre duas variáveis qualitativas (Kappa <20% desprezível; 21 a 40% mínimo; 41 a 60% regular; 61 a 80% bom; acima de 81% ótimo), permitindo uma variabilidade de acordo de +/- 10 pontos na classificação EAI da escala perceptual (Chan & Yiu, 2002; Awan & Lawson, 2009).

Os testes não paramétricos serão usados se as suposições iniciais para o seu uso – nomeadamente a normalidade da distribuição e homogeneidade de variâncias – não forem consideradas satisfatórias.

Assim, a consistência intra-avaliadores – medição da manutenção da cotação perceptiva nos 10% de repetições da amostra-teste – será medida através do cálculo do coeficiente de correlação de Pearson, permitindo uma variabilidade de acordo de +/- 10 pontos na classificação EAI da escala perceptual (Chan & Yiu, 2002; Awan & Lawson, 2009).

A consistência inter-avaliadores, para cada um dos grupos em estudo – teste-piloto com os peritos; alunos; profissionais – será analisado através do alfa de Chronbach permitindo uma variabilidade de acordo de +/- 10 pontos na classificação EAI da escala perceptual (Chan & Yiu, 2002; Awan & Lawson, 2009).

Na comparação entre os dois grupos que participarão na investigação – sem treino (alunos) vs com experiência (terapeutas) – em relação aos três parâmetros de avaliação perceptual será usado o teste de Igualdade de Duas Proporções. Este é um teste não paramétrico que compara se a proporção de respostas de duas determinadas variáveis e/ou dos seus níveis são estatisticamente significativas.

Na tentativa de estabelecer uma relação e/ou associação entre o grau de mestria (treino vs experiência) e as classificações perceptivas utilizar-se-á o teste de Qui-Quadrado para a Independência. É um teste não paramétrico utilizado para verificar se duas variáveis e os seus níveis possuem ou não uma dependência e/ou associação estatística.

Na análise das medidas acústicas dos diferentes softwares de análise será usado o teste de Mann-Whitney. Este teste não paramétrico é usado em amostras independentes, na comparação dos pares de variáveis. A consistência entre as medidas de análise acústica dos quatro softwares (consistência inter-avaliadores) será determinada pela medição do RMS (Root Mean Square) – erro padrão da média – que determina a distância que existe entre as medidas, para os diferentes programas de avaliação.

Na análise das medidas acústicas (para os quatro softwares) e os três tipos de vozes (grau geral, rugosidade e sopro) será usado o teste de Kruskal-Wallis, por existirem mais de duas variáveis simultaneamente. A comparação das vozes aos pares, para determinar a diferença entre elas, será realizada através do teste de Mann-Whitney.

A correlação de Spearman será usada para medir a relação entre as variáveis nos dois modos de avaliação estudados – áudio-perceptual e através do método acústico. A correlação de Spearman baseia-se na ordenação de duas variáveis sem qualquer restrição quanto à distribuição de valores, ou seja, é mais usada para dados não paramétricos. Será produzida uma matriz de regressão logística com a determinação

dos sinais de correlação (positivo ou negativo) – pela análise dos resíduos – 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). Serão consideradas e analisadas apenas as correlações com qualidade boa e ótima, isto é, acima de 61%. O Teste de Correlação será usado para validar a correlação e testar o coeficiente entre as variáveis, com um nível de significância $p < .05$ (intervalo de confiança de 95%) as diferenças são consideradas significativas.

3.3. CRONOGRAMA

	BRASIL – UNICAMP	PORTUGAL – FEUP
Contrução da Base de Dados	Outubro/2010	
Avaliação áudio-perceptual pelo grupo de peritos	Outubro/2010	
Avaliação Audio-perceptual pelo grupo de alunos	Novembro/2010	Fevereiro/2011
Avaliação Audio-perceptual pelo grupo de profissionais	Novembro/2010	Fevereiro/2011
Avaliação Acústica	Dezembro/2010	Fevereiro/2010
Anaálise de dados e Redacção da Tese	Outubro/2010  Junho/2011	
Entrega da Tese		Julho/2011

Capítulo 4 – CONCLUSÃO

O conhecimento e a adopção na prática dos modelos teóricos que sustentam as representações e medidas perceptuais e acústicas da fala e da voz – com vista ao estabelecimento de correlações entre a percepção e restantes níveis de descrição do sinal sonoro – são de extrema importância, e foram referidos ao longo do primeiro e segundo capítulo deste estado da arte.

Por último, foi feita a descrição da metodologia para a construção de conhecimento que se reflectirá na dissertação, a qual procurará acrescentar dados práticos e pedagógicos sobre a aplicabilidade do método acústico pelo Terapeuta da Fala.

Conclui-se que o estudo do sinal da fala ganhou um novo impulso com a possibilidade de correlação entre os seus componentes acústicos, perceptivos e fisiológicos. A incorporação de recursos instrumentais, especialmente a análise acústica, representou uma nova perspectiva na compreensão dos ajustes subjacentes à produção da fala e da voz, com e sem alterações, o que se repercute no processo terapêutico – desde o diagnóstico até à intervenção do Terapeuta da Fala propriamente dita.

O aprofundar destes conhecimentos permite uma multiplicidade de actuações terapêuticas, algumas ainda pouco exploradas pelo Terapeuta da Fala em Portugal. O mais comum é o nosso grupo profissional fazer uso da análise acústica para a avaliação e tratamento das alterações da fala e da voz, contudo, podemos 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 (SAAC));

- Construção e implementação de interfaces de comunicação (por exemplo, sistemas de comunicação por telefone, comunicação virtual, localização por satélite, sistemas de auto-atendimento);
- 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. Deve incluir a avaliação perceptivo-auditiva (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.

Ainda existe um amplo campo de investigação nesta área. São úteis estudos futuros que visem confirmar resultados e desenvolver ferramentas de uso clínico, que consigam contornar as desvantagens dos métodos existentes, propiciando aos pesquisadores, profissionais da saúde e aos próprios falantes, dados mais precisos e confiáveis.

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 actuação clínica e podendo, inclusive, fazer incursões noutros campos inovadores ao actuar com as mais variadas manifestações de fala e voz.

Ao chegar ao final da dissertação, para além de deixar um material de treino e teste em termos de avaliação áudio-perceptual, para actuais e futuros Terapeutas da Fala portugueses, prevê-se a possibilidade de criar uma medida de

quantificação/classificação da qualidade vocal que conjugue os parâmetros da escala GRB e os valores mais relevantes estatisticamente da análise acústica levada a cabo.

BIBLIOGRAFIA

1. Abitbol, J.; Abitbol, P.; Abitbol, B. – “Sex Hormones and the Female Voice”. *J Voice*. 1999; 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. – “Effect of caffeine on the vocal folds: a pilot study”. *J Laryngol Otol*. 1999; 113: 341-345.
4. Albertini, G.; Giaquinto, S.; Mignamo, M. – “Spectral analysis of the human voice: a potentially useful tool in rehabilitation.” *Eur J Phys Rehabil Med*. 2009; 45(4): 537-45.
5. Allali, A.; LeHuche, F.; “*A Voz – Patologia de Origen Funcional*”. 2ª Edição, Porto Alegre: Artmed Editora, 2005.
6. Anders, L.; Hollien, H.; Hurme, P.; Sonninen, A.; Wendler, J. – “Perceptual evaluation of hoarseness by several classes of listeners.” *Folia Phoniatr Logop*. 1988; 40: 91-100.
7. Andrada e Silva, M.A., Duprat, A.C. “Voz Cantada”. In: Ferreira, L.P.; Beffi-Lopes, D.M.; Limongi, S.C.O. (2004) – *Tratado de Fonoaudiologia*. São Paulo: Roca.
8. Andrews, M.L. (1995) – *Manual of Voice Treatment, Pediatrics Through Geriatrics*. San Diego: Singular.
9. Araújo, S.A. (1999) – “*Análise acústica da voz normal*.” Dissertação de Mestrado em otorrinolaringologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo.
10. Araújo, S.A.; Grellet, M.; Pereira, J.C.; Rosa, M.O. – “Normatização de medidas acústicas da voz normal.” *Rev Bras Otorrinolaringol*. 2002; 68: 540-544.
11. Aronson (1990) – *Clinical Voice Disorders*. 3rd Edition. New York: Thieme Medical Publishers, Inc.

12. Awan, S.N.; Lawson, L.L. – “The Effect of Anchor Modality on the Reliability of Vocal Severity Ratings.” *J Voice*. 2009; 23(3): 341-352.
13. Awan, S.N.; Roy, N. – “Acoustic prediction of voice type in women with functional dysphonia.” *J Voice*. 2005; 19: 268-282.
14. Awan, S.N.; Roy, N. – “Toward the development of an objective index of dysphonia severity: A four-factor model.” *Clin Ling & Phon*. 2006; 20: 35-49.
15. Awan, S.N.; Roy, N. – “Outcomes Measurement in Voice Disorders: Application of a Acoustic Index of Dysphonia Severity.” *J Speech Lang Hear Res*. 2009; 52: 482-499.
16. Baken, R.J.; Orlikoff, R.J. (2000) – “*Clinical Measurement of Speech and Voice*.” 2nd Edition. San Diego: Singular Publishing.
17. Barros, A.P.B.; Carrara De-Angelis, E.C. – “Análise Acústica da Voz”. In: Dedivitis, R.A.; Barros, A.P.B. (2002) – *Métodos de avaliação e diagnóstico de laringe e voz*. São Paulo: Lovise Editores.
18. Behlau, Mara (Org). (2001) – *Voz. O Livro do Especialista*. Rio de Janeiro: Livraria e Editora Revinter, Lda. Volume 1.
19. Behlau, M. – “Técnicas Vocais”. In: Fernandes, F.D.M. et al. (2010) – *Tratado de Fonoaudiologia*. 2^a Edição. São Paulo: Roca.
20. Behlau, M. – “Consensus auditory-perceptual evaluation of voice (CAPE-V), ASHA 2003”. *Rev Soc Bras Fonoaudiologia*. 2004; 9(3):187-9.
21. Behram, A. – “Common Practices of Voice Therapists in the Evaluation of Patients”. *J Voice*. 2005; 19(3): 454-469.
22. Bele, I.V. – “Reliability in Perceptual Analysis of Voice Quality.” *J Voice*. 2005; 19(4): 555-573.
23. Benninger, MS; Jacobson, BH; Johnson, AF (Edit.). (1994) – *Vocal Arts Medicine: The Care and Prevention of Professional Voice Disorders*. New York: Thieme Medical Publishers, Inc.
24. Bielamowicz, S.; Kreiman, J.; Gerratt, B.R.; DAuer, M.S.; Berke, G.S. – “Comparison of voicr analysis systems for perturbation measurement.” *J Voice*. 1996; 39: 126-34.

25. Bonatto, M.T.R.L.; Andrada e Silva, M.A.; Costa, O.H. – “A relação entre respiração e sistema sensorio-motor oral em crianças disfônicas.” *Ver CEFAC*. 2004; 6(1): 58-66.
26. Bhuta, T.; Patrick, L.; Garnett, J.D. “Perceptual Evaluation of voice quality and its correlation with acoustic measurements”. *J Voice*. 2004; 18: 299-304.
27. Bistrisky, Y.; Frank, Y. – “Efficacy of voice and speech training of prospective elementary school teachers”. *Israeli J Speech Hear*. 1981; 10: 16-32.
28. Boone, D.R.; McFarlane, S.C. (2000) – *The Voice and Voice Therapy*. (6th Edition). Boston: Allyn & Bacon.
29. Boucher, V.J. – “Acoustic Correlates of Fatigue in Laryngeal Muscles: Findings for a Criterion-Based Prevention of Acquired Voice Pathologies.” *J Speech, Lang, Hear Res*. 2008; 51: 1161-1170.
30. Brasolotto, A.G.; Rehder, M.I. (2011) – “Diagnóstico Vocal Fonoaudiológico.” *In*: Rehder, M.I. & Branco, A. *Disfonia e Disfagia*. Rio de Janeiro: Livraria e Editora Revinter, Ltda.
31. Brindle, B.R.; Morris, H.L. – “Prevalence of voice quality deviations in the normal adult populations”. *J Commun Disord*. 1979; 12: 439-45.
32. Broaddus-Lawrence, P.L. ; Treole, K. ; McCabe, R.B. ; Allen, R.L. ; Toppin, L. – “The Effects of Preventive Vocal Hygiene Education on the Vocal Hygiene Habits and Perceptual Vocal Characteristics of Training Singers”. *J Voice*. 2000; 14(1): 58-71.
33. Bunton, K. ; Kent, R.D. ; Duffy, J.R. ; Rosenbeck, J.C. ; Kent, J.F. – “Listener Agreement for Auditory-Perceptual Ratings of Dysarthria.” *J Speech Lang Hear Res*. 2007; 50: 1481-1495.
34. Calas M.; Verhulst J.; Lecoq M.; Dalleas B.; Seilhean M. – “La Phatologie Vocale chez L'Enseignant”. *Rev Laryngol Otol Rhinol Bord*. 1989; 110: 397-406.
35. Camargo, Z.A. – *Avaliação Objectiva da Voz*. *In* : Carrara-de-Angelis, E. et al. – “A actuação fonoaudiológica no cancer de cabeça e pescoço.” São Paulo: Lovise; 2000.

36. Camargo, Z.A. ; Madureira, S. ; Tsuji, D.H. – *Analysis of dysphonic voices based on the interpretation of acoustic, physiological and perceptual data.* In: "6th International Seminar on Speech Production". Sydney; 2003.
37. Camargo, Z.; Madureira, S. – Análise Acústica: Aplicações na Fonoaudiologia. In Fernandes, F.D.M. et al., (2010) – *Tratado de Fonoaudiologia*. 2ª Edição. São Paulo: Roca.
38. Campisi, O.; Tewfik, T.L.; Pelland-Blais, E.; Hussein, M.; Sadegui, N. – "Multidimensional Voice Program analysis in children with vocal cord nodules." *J Otolaryngol*. 2000; 29(5): 302-8.
39. Cannito, M.P.; Woodson, G.E.; Murry, T.; Bender, D. – "Perceptual Analyses od Spasmodic Dysphonia Before and After Treatment." *Arch Otolaryngol Head Neck Surg*. 2004; 130: 1393-1399.
40. Carding, P.; Carlson, E.; Epstein, R.; Mathieson, L.; Shewell, C. – "Formal perceptual evaluation of voice quality in United Kingdom." *Log Phon Vocol*. 2000; 25: 133-138.
41. Chan, R.W.K. – "Does the voice improve with vocal hygiene education? A study of some instrumental voice measures in a group of kindergarden teachers". *J Voice*. 1994; 8: 279-91.
42. Chan, K.M.K.; Yiu, E.M-L. – "A Comparison of Two Perceptual Voice Evaluation Training Programs for Naive Listeners." *J Voice*. 2006; 20(2): 229-241.
43. Chen, S.H.; Chang, S.Y.; Fue, S.W.; Chang, P. – "A Survey on Voice Disorders on Junior High School Teachers in Taipei". Apresentação no XXth Congress of the International Association of Logopedics and Phoniatics. Tokyo. 1996.
44. Childers, D.G.; Skynner, D.P.; Kemerait, R. – "The Cepstrum: a guide to processing." *Proceedings of the IEEE*. Annals of the IEEE. 1977; 65(10): 1428-1442.
45. Colton, R.H.; Casper, J.; *"Compreendendo os problemas da voz: uma perspectiva fisiológica ao diagnóstico e ao tratamento"*, Porto Alegre: Editora Artes Médicas, 1996.
46. Comins, R. – "Vocal Tuition for professional voice users: a tutor`s account". *J Brit Voice Ass*. 1995; 4: 32-43.

47. Costa, Henrique O.; Duprat, André; Eckley, Cláudia; Silva, Marta A.A. e. – “Caracterização do Profissional da Voz para o Laringologista”. *Revista Brasileira de Otorrinolaringologia*. 2000; 66 (2) Parte 1: 129-134.
48. Cummings, Louise; “*Clinical Linguistics*”. Edinburgh: Edinburgh University Press, 2008.
49. Dajer, M.E.; “Análise de Sinais de Voz por Padrões Visuais de Dinâmica Vocal”. São Paulo: Escola de Engenharia de São Carlos (Tese de Doutorado), 2010.
50. Davis, S.B. – “Acoustic characteristics of normal and pathological voices.” In: Lass, N.J. (1979) – *Speech and language: advances in basic research and practice*. New York: Academic Press.
51. De Krom, G. – “Consistency and Reliability of Voice Quality Ratings for Different Types of Speech Fragments.” *J Speech Her Res*. 1994; 37: 985-1000.
52. De Bodt, M.S.; Van de Heyning, P.H.; Wuyts, F.L.; Lambrechts, L. – “The perceptual evaluation of voice disorders.” *Acta Otorhynolaryngol Belg*. 1996; 50: 283-91.
53. De Bodt, M.S.; Wuyts, F.L.; Van de Heyning, P.H.; Croux, C. – “Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual ratings of voice quality.” *J Voice*. 1997; 1: 74-80.
54. Dedivitis, R.A.; Barros, A.P.B.; Queija, D.S.; Alexandre, J.C.; Rezende, W.T.; Corazza, V.R., et al. – “Interobserver perceptual analysis of smokers voice.” *Clin Otolaryngol Allied Sci*. 2004; 29: 124-7.
55. Dedivitis, R.A.; Barros, A.P.B.; Queija, D.S.; Pfuetzenreiter Jr., E.G.; Bohn, N.P. – “Achados perceptivo-auditivos e acústicos em pacientes submetidos à laringectomia fronto-lateral.” *Rev Bras Cir Cabeça Pescoço*. 2008; 37(3): 163-165.
56. Dejonckere, P.H. – “Principal components in voice pathology.” *J Voice*. 1995; 4: 96-105.
57. Dejonckere, P.H. (Edit) (2001) – *Occupational Voice – care and cure*. San Diego: Singular Publishing Group.

58. Dejonckere, P.H.; Bradley, P.; Clemente, P. et al. – “A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques.” *Eur Arch Otorhinolaryngol*. 2001; 258: 77-82.
59. Dejonckere, P.H.; Lebacqz, J. – “Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology.” *ORL J Otorhinolaryngol Relat Spec*. 1996; 58(6): 326-32.
60. Dejonckere, P.H.; Obbens, C.; Demoor, G.M.; Wienecke, G.H. – “Perceptual evaluation of dysphonia: Reliability and relevance.” *Folia Phoniat Logopaed*. 1993; 45: 76-83.
61. Deliyski, D. – “Acoustic model and evaluation of pathological voice production”. *Proceedings of the 3rd Conference on Speech Communication and Technology*. 1993; 1969-1972.
62. Eadie, T.L.; Doyle, F. “Classification of Dysphonic Voice: Acoustic and Auditory-Perceptual Measures”. *J Voice*. 2005; 19: 1-14.
63. Eadie, T.L.; Baylor, C.R. “The Effect of Perceptual Training on Inexperienced Listeners` Judgements of Dysphonic Voice”. *J Voice*. 2006; 20(4): 527-544.
64. Eskenazi, L.; Childers, D.G.; Hicks, D.M. – “Acoustics correlates of vocal quality.” *J Speech Hear Res*. 1990; 33: 298-306.
65. Espanhol, D.R. – *Determinadores de Pitch*. Universidade de São Paulo: Departamento de Engenharia Elétrica. Tese de Mestrado em Engenharia Elétrica. 2004.
66. Fant, G. (1970) – *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*. The Hague: Mouton.
67. Finizia, C.; Dotevall, H.; Lundström, E.; Lindström, J. – “Acoustic and Perceptual Evaluation of Voice and Speech Quality.” *Arch Otolaryngol Head Neck Surg*. 1999; 125: 157-163.
68. Fioravanti, Carlos – “Prova de Resistência”. *Pesquisa FAPESP. Ciência e Tecnologia no Brasil*. 2003; 89: 32-37.
69. Freeman, M.; Fawcus, M. (2004) – “*Distúrbios da Voz e seu Tratamento*”, 3^a Edição, São Paulo: Livraria Santos Editora.

70. Freitas, D. – “Complicações extra-esofágicas da DRGE”. *Endonews*. 2003; 6: 21.
71. Fritzell, Björn – “Voice Disorders and Occupations”. *Log Phon Vocol*. 1996; 21:7-12.
72. Fröhlich, M.; Michaelis, D.; Strube, H.W.; Kruse, E. – “Acoustic voice quality description: case studies for different regions of the hoarseness diagram”. In: Wittenberg T, Mergell P, Tigges M, Eysholdt U (ed). *Advances in quantitative laryngoscopy*, 2nd Round Table; 1997; Erlangen. p.143-150.
73. Fröhlich, M.; Michaelis, D.; Strube, H.W.; Kruse, E. – “Acoustic voice analysis by means of the hoarseness diagram”. *J Speech Hear Res*. 2000; 43: 706-720.
74. Garrett, C.G. – “Hoarseness”. *Med Clin North Am*. 1999; 83: 115-123.
75. Gerratt, B.R.; Kreiman, J. – “Theoretical and methodological development in study of pathological voice quality.” *J Phonetics*. 2000; 28: 335-42.
76. Gerratt, B.R.; Kreiman, J.; Antonnanzaz-Barroso, N.; Berke, G.S. – “Comparing internal and external standards in voice quality judgments”. *J Speech Hear Res*. 1993; 36: 14-20.
77. Giovanni, A.; Robert, D.; Estublier, N.; Teston, B.; Zanaret, M.; Cannoni, M. – “Objective evaluation of dysphonia: Preliminary results of a device allowing simultaneous acoustic and aerodynamic measures.” *Folia Phoniatr Logop*. 1996; 48: 175-185.
78. Gliklich, R.E.; Glosky, R.M.; Montgomery, W.W. – “Validation of a Voice Outcome Survey for Unilateral Vocal Cord Paralysis”. *Otolaryngol Head Neck Surg*. 1999; 120: 153-8.
79. Gotaas, C.; Starr, C.D. – “Vocal Fatigue among Teachers”. *Folia Phoniatr*. 1993; 45: 120-9.
80. Guimarães, I. (2002) – *An electrolaryngographic study of dysphonic Portuguese speakers*. University of London. Tese de Doutoramento.
81. Guimarães, I. (2007) – “*A Ciência e a Arte da Voz Humana*”, Alcoitão: ESSA – Escola Superior de Saúde de Alcoitão.

82. Granqvist, S. – “The visual sort and and rate method for perceptual evaluation in listening tests”. *Logoped Phoniatr Vocol.* 2003; 28: 109-116.
83. Hakkesteegt, M.M.; Brocaar, Michael M.P.; Wieringa, M.H.; Feenstra, Louw – “The Relationship Between Perceptual Evaluation and Objective Multiparametric Evaluation of Dysphonia Severity”. *J. Voice.* 2008; 2: 138-145
84. Hammarberg, B. (1986) – *Perceptual and acoustic analysis of dysphonia.* Department of Logopedics and Phoniatics, Huddinge University Hospital, Stockolm. Tese de Doutoramento.
85. Hammarberg, B. – “Voice Research and Clinical Needs”. *Folia Phoniatr Logop.* 2000; 52: 93-102.
86. Harvey, P.L. – “The three ages of voice – the young adult patient.” *J Voice.* 1997; 11: 144-52.
87. Harnisch, W.; Brosch, S.; Schmidt, M.; Hagen, R. – “Breathing and voice quality after surgical treatment for bilateral vocal cord paralysis.” *Arch Otolaryngol Head Neck Surg.* 2008; 134: 278-84.
88. Heidel S.E.; Torgerson J.K. – “Vocal problems among aerobic instructors and aerobic participants”. *J Commun Disord.* 1993; 26:179-191.
89. Hemler, R.J.B.; Wieneke, G.H.; Dejonckere, P.H. – “The Effect of Relative Humidity of Inhaled Air on Acoustic Parameters of Voice in Normal Subjects”. *J Voice.* 1997; 11(3):295-300
90. Hillenbrand, J.; Cleveland, R.A.; Erickson, R.L. – “Acoustic correlates of breathy vocal quality”. *J Speech Hear Res.* 1994; 37:769-778.
91. Hillenbrand, J.; Houde, R.A. – “Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech.” *J Speech Hear Res.* 1996; 39: 311-321.
92. Hirano, M.; Bless, D.M. (1997) – *Exame Videoestroboscópico da laringe.* Porto Alegre: Artes Médicas.
93. Hiroto, I. – “Hoarseness: view points of voice physiology.” *Jap J Logoped Phoniatr.* 1967; 8: 1-9.
94. Hogikyan, N.D.; Sethuraman, G. – “Validation of an instrument to measure voice-related quality of life (V-RQOL)”. *J Voice.* 1999; 13 :557-569.

95. Jacobson, BH; Johnson, A; Grywalski C.; Silbergleit A; Jacobson, G; Benninger, M.S. – “The Voice Handicap Index (VHI): Development and Validation”. *Am J Speech Lang Pathol.* 1997; 6(3): 66-70.
96. Johnson, K. (2003) – *Acoustics and Auditory Phonetics*. 2nd Edition. Malden: Blackwell.
97. Jones, K.; Sigmon, J.; Hock L.; Nelson, E. – “Prevalence and Risk Factors for Voice Problems among Telemarketers”. *Arch Otolaryngol Head Neck Surg.* 2002; 128: 571-77.
98. Jones, P.M. – “Striking the Right Chord”. *Royal College of Speech and Language Therapists Bulletin.* 1999; 565: 13-4.
99. Jónsdóttir, V.I. – “Cordless amplifying system in classrooms. A descriptive study of teachers and student`s options”. *Log Phon Vocol.* 2002; 27: 29-36.
100. Jónsdóttir, V.I.; Boyle, B.E.; Martin, P.J.; Sigurdardóttir, G. – “A Comparison of the occurrence and nature of vocal symptoms in two groups of Icelandic teachers”. *Log Phons Vocol.* 2002; 27:98-105.
101. Jónsdóttir, V.I.; Laukkanen; A-M; Vilkmán, E. – “Changes in Teachers Speech during a Working Day with and without Electric Sound Amplification”. *Folia Phoniatr Logop.* 2002; 54:282-287.
102. Karnell, M.P.; Melton, S.D.; Childes, J.M.; Coleman, T.C.; Dailey, S.A.; Hoffman, H.T. – “Reliability of Clinician-Based (GRBAS and CAPE-V) and Patient-Based (V-RQOL and IPVI) Documentation of Voice Disorders.” *J Voice.* 2007; 21(5): 576-590.
103. Kasuya, H.; Ogawa, S.; Mashima, K.; Ebihara, S. – “Normalized noise energy as an acoustic measure to evaluate pathologic voice”. *J Acoust Soc America.* 1986; 80(5): 1329-1334.
104. Kempster, G.B.; Gerratt, B.R.; Verdolini, A.K.; Barkmeier-Kraemer, J.; Hillman, R.E. – “Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol.” *Am J Speech Lang Patholog.* 2009; 18: 124-132.
105. Kent, R.D.; Read, C. (1992) – *The Acoustics Analysis of Speech*. San Diego: Singular Publishing.
106. Kent, R.D. – “Vocal tract acoustics.” *J Voice.* 1993; 7(2): 97-117.

107. Kent, R.D. – “Hearing and believing: some limits to the auditory-perceptual assessment in voice quality perception.” *Am J Speech Lang Path.* 1996; 5: 7-23.
108. Kent, R.D. (1997) – *The Speech Sciences*. San Diego: Singular Publishing.
109. Kent, R.D. (1997) – *The MIT Encyclopedia of Communication Disorders*. Massachusetts: Massachusetts Institute of Technology.
110. Klingholtz, F. – “Acoustic recognition of voice disorders: a comparative study of running speech versus sustained vowels.” *J Acoust Soc Am.* 1990; 87: 2218-2224.
111. Kooijman, P.G.; de Jong, F.I.; Oudes, M.J.; Huinck, W.; van Acht, H.; Graamans, K. – “Muscular tension and body posture in relation to voice handicap quality in teachers with persistent voice complaints.” *Folia Phoniatr Logop.* 2005; 57: 134-147.
112. Koschkee, D.C. (1993) – *Voice Disability Index*. Madison: University of Wisconsin Hospital and Clinics.
113. Koufmann J. A.; Isaacson, G. (Edit) – “Voice Disorders”. *Otol Clin N.A.* 1991; 24(5).
114. Koufmann J.; Blalock, D. – “Vocal Fatigue and Dysphonia in The Professional Voice User: Bogart-Bacall Syndrome”. *Laryngoscope.* 1998; 98:493-8.
115. Kreiman, J.; Gerratt, B. – “Sources of listener disagreement in voice quality assessment.” *J Acoust Soc Am.* 2000; 108: 1867-1876.
116. Kreiman, J.; Gerratt, B.R.; Kempster, G.B.; Ernan, A.; Berke, G.S. – “Perceptual evaluation of voice quality: review, tutorial and a framework for future research”. *J Speech Hear Res.* 1993; 36: 21-40.
117. Kreiman, J.; Gerratt, B.R.; Precoda, K. – “Listener experience and perception of voice quality.” *J Speech Hear Res.* 1990; 33:103-15.
118. Kreiman, J.; Gerratt, B.R.; Precoda, K.; Berke, G.S. – “Individual differences in voice quality perception”. *J Speech Hear Res.* 1992; 35:512-520.
119. Ladefoged, P.; “*A course in Phonetics*”. 5ª Edição. Boston: Thomson Wadsworth, 2006.

120. Languaite J.K. – “Adult Voice Screening”. *J Speech Hear Disord.* 1972; 37: 147-51.
121. Lee, C.K.; Childers, D.G. – “Some acoustical, perceptual and physiological aspects of vocal quality.” *In: Gauffin, J.; Hammarberg, B. (Edit.) – Vocal Fold Physiology – acoustic, perceptual and physiological aspects of voice mechanisms.* San Diego: Singular; 1991.
122. Laver, J. (1980) – *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.
123. Lee, C.F.; Carding, P.N.; Fletcher, M. – “The nature and severity of voice disorders in lung cancer patients.” *Logopedics Phoniatrics Vocology.* 2008; 33: 93-103.
124. Long J.; Williford H.N.; Olson, M.S.; Wolfe, V.– “Voice problems and risk factors among aerobic instructors”. *J Voice.* 1998; 12: 197-207.
125. Lopes, L.; Vaz-Freitas, S.; Sousa, R.; Matos, J.; Abreu, F.; Ferreira, A. – “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 de 2008.
126. Ma, E.P-M.; Yiu, E.M-L. – “Voice Activity and Participation Profile: Assessing the Impact of Voice Disorders on Daily Activities”. *J Speech Lang Hear Res.* 2001; 44(3): 511-524.
127. Ma, E.P.; Yiu, E.M.L. – “Suitability of acoustic perturbation measures in analysing periodic and nearly periodic voice signals.” *Folia Phoniatr Logop.* 2005; 57(1): 38-47.
128. Ma, E.P-M.; Yiu, E.M.L. – “Multiparametric Evaluation of Dysphonic Severity”. *J Voice.* 2006; 20: 380-390.
129. Maccallum, J.K.; Cai, L.; Zhou, L.; Zhang, Y.; Jiang, J.J. – “Acoustic analysis of aperiodic voice: perturbation and nonlinear dynamic properties in esophageal phonation.” *J Voice.* 2009; 23(3): 283-90.
130. Madazio, G.; Behlau, M. – “Atualização em Voz”, *in Berretin-Félix et al. – (Re)Habilitação Fonoaudiológica. Avaliação da Eficácia.* São Paulo: Pulso Editorial Lda., 2009.

131. 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.
132. Martin, S. – “Voice care and development for teachers: survey report”. *J Brit Voice Ass.* 1994; 3: 92-8.
133. Martin, D.P.; Fitch, J.; Wolfe, V.I. – “Pathological voice type and the acoustic prediction of severity.” *J Speech Hear Res.* 1995; 38: 765-771.
134. Martin, D.P.; Wolfe, V.I. – “Effects of perceptual training on ability to use the vocal profile analysis scheme.” *Percept Motor Skills.* 1996; 33: 1291-1298.
135. Maryn, Y.; Roy, N.; DeBodt, M.; Van Cauwenberge, P.; Corthals, P. – “Acoustic measurement of overall voice quality: a meta-analysis.” *J Acoust Soc Am.* 2009; 126(5): 2619-34.
136. Maryn, Y.; Corthals, P.; Van Cauwenberge, P.; Roy, N.; DeBodt, M. – “Toward Improved Ecological Validity in the Acoustic Measurement of Overall Voice Quality: Combining Continuous Speech and Sustained Vowels.” *J Voice.* 2010; 24(5): 540-55.
137. Mattiske, J.A.; Oates, J.M.; Greenwood, K.M. – “Vocal Problems among Teachers: A Review of Prevalence, Causes, Prevention and Treatment”. *J Voice.* 1998; 12 (4): 489-499.
138. Mehta, D.D.; Hillman, R.E. – “Voice assessment: updates on perceptual, acoustic, aerodynamic and endoscopic imaging methods.” *Curr Opin Otolaryngol Head and Neck Surg.* 2008; 16: 211-15.
139. Mendes, A.; Rothman, H.; Sapienza, C. & Brown, W. “Effects of Vocal Training on the Acoustic Parameters of the Singing Voice”. *J Voice.* 2003; 17: 529-543.
140. Miller, M.K.; Verdolini, K – “Frequency and risk factors for voice problems in teachers of singing and control subjects”. *J Voice.* 1995; 9: 348-62.
141. Michaelis, D.; Gramss, T.; Strube, H.W. – “Glottal-to-noise excitation ratio – a new measure for describing pathological voices”. *Acustica – Acta Acustica.* 1997; 83: 700-706.

142. Michaelis, D.; Frölich, M.; Strube, H.W. – “Selection and combination of acoustic features for the description of pathologic voices.” *J Acoust Soc Am*. 1998; 103: 1628-1638.
143. Mitev, P.; Hadjitodorov, S. – “A method for turbulent noise estimation in voiced signals.” *J Med Biol Eng Comput*. 2000; 38: 625-631.
144. Mjaavatn, P.E. (1980) – “Voice Difficulties among Teachers”. Apresentação do XVIII Congress of the International Association of Logopedics and Phoniatics. Washington.
145. Moran, R.J.; Reilly, R.B.; De Chazal, P., P.D.; Lacy, P.D. – “telephony-based voice pathology assessment using automated speech analysis”. *IEEE Transactions on Biomedical Engineering*. 2006; 53(3): 468-477.
146. Morely, D.E. – “A ten-year survey of speech disorders among university students”. *J Speech Hear Disord*. 1952: 25-31.
147. Morton V. (1995) – *Educating Teachers*. Trabalho apresentado no 3rd Voice Symposium of Australia. Sydney.
148. Morton, V.; Watson, D.R. – “The Teaching Voice: Problems and Perceptions”. *Logop Phoniatr Vocol*. 1998; 23: 133-139.
149. Murphy, P.J. – “Periodicity estimation in synthesized phonation signals using cepstral harmonic peaks.” *Speech Communic*. 2006; 48: 1704-1713.
150. Murray, L.; Johnston, B.; Lane, A.; Harveu, I.; Donovan, J.; Nair, P.; Harvey, R. – “Relationship between body mass and gastro-oesophageal reflux symptoms: The Bristol Helicobacter Project”. *Int J Epid*. 2003; 32:645-650
151. Murdock, B. – *Disartria. Uma abordagem Fisiológica para Avaliação e Tratamento*. São Paulo: Editora Lovise, Lta., 2005.
152. Nawka, T.; Anders, L.C.; Wendler, J. – “Die auditive Beurteilung heiserer Stimmen nach dem RGH-System.” *Sprache-Stimme-Gehir*. 1994; 18: 130-33.
153. Nemr, K.; Lehn, C. – “Voz em Câncer de Cabeça e Pescoço”. In Fernandes, F.D.M. et al. (2010) – *Tratado de Fonoaudiologia*. 2^a Edição. São Paulo: Roca.

154. Nikolov, Z.; Deliyski, D.; Drumeva, L.; Boyanov, B. – “Computer system for diagnostics of pathological voices”. *Proceedings of the 21st Congress International of Logopedics and Phoniatics*. 1989; 1: 973-976.
155. Oates, J. – “Auditory-perceptual evaluation of disordered vocal quality – pros, cons and future directions.” *Folia Phoniatica Logop*. 2009; 61 (1): 49-56.
156. Oliveira, I.B. – “Qualidade Vocal: Correlatos aústicos, perceptivo-auditivos e fisiológicos.” *Rev Soc Bras Fonoaudiol*. 2004; Suplemento Especial.
157. Oliveira, I.B. – “Avaliação Fonoaudiológica da Voz: Reflexões sobre Condutas com Enfoques à Voz Profissional”. In Fernandes, F.D.M et al. (2010) – *Tratado de Fonoaudiologia*. 2^a Edição. São Paulo: Roca.
158. Orlikoff, R.F.; Dejonckere, P.H.; Dembowski, J.; Fitch, J.; Gelfer, M.P.; Gerratt, B.R. et al. – “The perceived role of voice perception on clinical practice.” *Phonoscope*. 1999; 2: 89-106.
159. Orr, R.; de Jong, F.; Cranen, B. - “Some objective measures indicate of perceived voice robustness in student teachers”. *Log Phon Vocol*. 2002; 27: 106-117.
160. Pabon, J.P.H. – “Objective acoustic voice-quality parameters in the computer phonetogram.” *J Voice*. 1991; 5: 203-216.
161. Pabon, J.P.H. – “Automatic phonetogram recording supplemented with acoustical voice quality parameters.” *J Speech Hear Res*. 1988; 31: 710-722.
162. Patel, S.; Shrivastav, R. – “Perception of dysphonic vocal quality: some thoughts and research update. Perspectives on voice and voice disorders.” July, 2007; 17:3-6 (ASHA SID-3).
163. Parsa, V.; Jamieson, D.G. – “Acoustic discrimination of pathological voice: sustained vowels versus continuous speech.” *J Speech Lan Hear Res*. 2001; 44: 327-339.
164. Pekkarinen, E.; Himberg, L.; Pentti, T. – “Prevalence of Vocal symptoms among Teachers compared with nurses: a questionnaire study”. *Scandinavian J Logoped Phoniatic*. 1992; 17:113-7.

165. Pereira, J.C.; Montagnoli, A.N. (1999) – *Análise Acústica dos Sinais de Voz*. São Carlos: EESC/USP. Sebenta.
166. Piccirillo, J.F.; Painter, C.; Fuller, D.; Fredrickson, J.M. – “Multivariate analyses of objective vocal function.” *Ann Otol Rhinol Laryngol*. 1998; 107: 107-112.
167. Piccirillo, J.F.; Painter, C.; Fuller, D.; Haiduk, A.; Fredrickson, J.M. – “Assessment of two objective voice function indices.” *Ann Otol Rhinol Laryngol*. 1998; 107: 175-185.
168. Pinho, S.M.R.; Ponto, P. “Escala de Avaliação perceptiva da fonte glótica: RASAT.” *Vox Brasilis*. 2002; 3: 11-13.
169. Pinho, S.M.R. – *Fundamentos em Fonoaudiologia. Tratando os Distúrbios da Voz*, 2ª Edição. Rio de Janeiro: Editora Guanabara Koogan, 2003.
170. Pinho, S.M.R.; Pontes, P. – *Músculos intrínsecos da Laringe e Dinâmica Vocal*. (Série Desvendando os Segredos da Voz), Volume 1, Rio de Janeiro: Revinter, 2008.
171. Pinho, S.; Tsuji, D.; Bohadana, S. – *Fundamentos em Laringologia e Voz*. Rio de Janeiro: Editora Guanabara Koogan, S.A., 2006.
172. Pouchoulin, Gilles (2008) – *Approche Statistique pour L'Analyse Objective et la Caractérisation de la Voix Dysphonique*. Université d'Avignon et des Pays de Vaucluse. Tese de Doutorado.
173. Probst, R.; Grevers, G.; Iro, H. (2004) – *Basic Othorinolaryngology. A Step-by-Step Learning Guide*. New York: Thieme Publishing.
174. Qi, Y.; Milstein, C.; Hillman, R.E. – “The estimation fo signal-to-noise ratio in continuous speech for disordered voices.” *J Acoust Soc Am*. 1999; 105: 2532-2535.
175. Roy, N.; Gray, S.D.; Simon, M.; Dove, H.; Corbin-Lewis, K.; Stemple, J.C. – “An Evaluation of the Effects of Two Treatment Approaches for Teachers With Voice Disorders: A Prospective Randomized Clinical Trial”. *J Speech Lang Hear Res*. 2001; 44: 286-96.
176. Roy, N.; Weinrich, B.; Gray, S.D.; Tanner, K.; Toledo, S.W.; Dove, H.; Corbin-Lewis, K.; Stemple, J.C. – “Voice amplification Versus Vocal Hygiene

- Instruction for Teachers With Voice Disorders: A Treatment Outcomes Study". *J Speech Lang Hear Res.* 2002; 45: 625-38.
177. Russell A.; Oates J.; Greenwood K.M. – "Prevalence of voice problems in teachers". *J Voice.* 1998; 12:467-479.
178. Sader, R.C.M.; Hanayama, E.M. – "Considerações teóricas sobre a abordagem acústica da voz infantil." *Rev CEFAC.* 2004; 6(3): 312-8.
179. Sala, Eeva; Laine, Anneli; Simberg, Susana; Pentti, Jaana; Suonpää, Juoko. – "The Prevalence of Voice Disorders Among Day Care Center Teachers Compared with Nurses: A Questionnaire and Clinical Study". *J Voice.* 2001; 15(3): 413-423.
180. Sapienza, C.M.; Crandell, C.C.; Curtis, B. – "Effects of Sound-Field Frequency Modulation Amplification on reducing Teachers Sound Pressure Level in the Classroom". *J Voice;* 13(3): 375-381.
181. Sapir, S. – "Vocal Attrition in Voice Students: Survey Findings". *J Voice.* 1993; 7: 69-74.
182. Sapir, S.; Keidar, A.; Mathers-Schmidt, B. – "Vocal Attrition in Teachers: Survey Findings". *Eur J Disord Commun.* 1993; 28: 177-85.
183. Sapir, S.; Mathers-Schmidt B.; Larson G.W. – "Singer`s and non-singer`s vocal health. Vocal behaviours, and attitudes towards voice and singing: indirect findings from a questionnaire". *Eur J Disord Commun.* 1996; 31: 193-209.
184. Sarfati, J. – "Readaptation vocale des enseignants". *Rev Laryngol.* 1989; 110: 393-5.
185. Schiavetti, N.; Metz, D.E. (2002) – *Evaluating Research in Communicative Disorders.* 4th Edition. Boston: Allyn & Bacon.
186. Sederholm, E.; McAllister, A.; Sundber, J.; Dalkvist, J. – "Perceptual Analysis of Child Hoarsness using Cintinuous Scales". *STL-QPSR.* 1992; 1: 99-113.
187. Shrivastav, R. – "Multidimensional scaling of breath voice quality: individual diferences in perception". *J Voice.* 2006; 20: 211-222.

188. Shrivastav, R.; Sapienza, C.M.; Nandur, V. – “Application of psychometric theory to the measurement of voice quality using rating scales”. *J Speech Lang Hear Res.* 2005; 48: 323-335.
189. Simberg, S.; Laine, A.; Sala, E.; Rönnemaa, A. – “Prevalence of Voice Disorders Among Future Teachers”. *J Voice.* 2000; 14(2): 231-5.
190. Simberg, S.; Sala, E.; Laine, A.; Rönnemaa, A. – “A fast and easy method for voice disorders among teachers students”. *Log Phon Vocol.* 2001; 26: 10-6.
191. Smith E.; Gray, S.D.; Dove, H.; Kirchner, L.; Heras, H. – “Frequency and Effects of Teachers Voice Problems”. *J Voice.* 1997; 11(1): 81-7.
192. Smith, E.; Kirchner, H.L.; Taylor, M.; Hoffman, H.; Lemke, J.H. – “Voice Problems Among Teachers: Differences by Gender and Teaching Characteristics”. *J Voice.* 1998; 12 (3): 328-334.
193. Smith, E.; Lemke, J.H.; Taylor, M.; Kirchner, H.L.; Hoffman, H. – “Frequency of Voice Problems Among Teachers and Other Occupations”. *J Voice.* 1998; 12 (4): 480-88.
194. Smith, E.; Verdolini, K.; Gray, S. *et al.* – “Effects of Voice Disorders on Quality of Life”. *J Speech Lang Pathol.* 1996; 4: 223-44.
195. Smits, I.; Ceuppens, P.; De Bodt, M. – “A Comparative Study of Acoustic Voice Measurements by Means of Dr. Speech and Computerized Speech Lab.” *J Voice.* 2005; 19(2): 187-196.
196. Sousa, R.; Vaz-Freitas, S.; Ferreira, A. – “Práticas Profissionais dos Terapeutas da Fala Portugueses nas Perturbações da Voz”. *Rev Soc Bras Fonoaudiol.* No prelo.
197. Souza, L.B.R. (2010) – *Atuação Fonoaudiológica em Voz.* Rio de Janeiro: Livraria e Editora Revinter, Lda.
198. Stemple, J.C.; Glaze, L.E.; Ferdeman, B.K. (1995) – *Clinical Voice Pathology: Theory and Managment.* 2nd Edition. San Diego: Singular.
199. Takahashi, H.; Koike, Y. – “Some perceptual dimensions and acoustical correlates of pathological voices”. *Acta Oto-Laryngol.* 1976; 338: 2-22.
200. Tavares, E.L.M.; Martins, R.H.G. – “Vocal Evaluation in Techers With or Without Symptoms.” *J Voice.* 2007; 21(4): 407-414.

201. Timmermans, B.; De Bodt, M.S.; Wuyts, F.L. et al. – “Poor quality on future elite vocal performers and professional voice users.” *J Voice*. 2002; 16: 372-382.
202. Timmermans, B.; De Bodt, M.S.; Wuyts, F.L.; Van de Heyning, P. – “Voice quality changes in future professional voice users after 18 months of voice training.” *Folia Phoniatr Logop*. 2004; 261: 1-5.
203. Timmermans, B.; De Bodt, M.S.; Wuyts, F.L.; Van de Heyning, P. – “Voice quality changes in future professional voice users after 9 months of voice training.” *Eur Arch Otorhinolaryngol*. 2005; 19: 511-518.
204. Titze, I. – “Standards in acoustic analysis of voice”. *J Voice*. 1994; 8(1): 1-7.
205. Titze, I.R. (1994) – *Principles of Voice Production*. Englewood Cliffs: Prentice-Hall.
206. Titze, I.R. (1995) – *Summary Statement. Workshop on Acoustic Voice Analysis*. Denver, C.O.: The Nacional Center for Voice and Speech.
207. Titze, I.R.; Lemke, J; Montequin, D. – “Population in the US workforce who rely on voice as a primary tool of trade: a preliminary report”. *J Voice*. 1997; 11: 254:259.
208. Titze, I.R. ; Bergan, C.C.; Hunter, E.J.; Story, B. – “Source and filter adjustments affecting the perception of the vocal qualities twang and yawn.” *Logoped Phoniatr Vocol*. 2003; 28: 147-155.
209. Umapathy, K.; Krisshan, S.; Parsa, V.; Jamieson, D.G. – “Discrimination of pathological voices using a time-frequency approach.” *IEEE. Transactions on Biomedical Enginneering*. Vol. 52, n° 3, pp: 421-430. 2005
210. Urrutikoetxea, A.; Ispizua, A.; Matellanes, F.; Aurrekoetxea, J. “Prevalence of Vocal Nodules in Female Teachers”. Apresentação em Vídeo no 1º Congresso Mundial da Voz. Porto. 1995.
211. Van Lierde, K.M.; Vinck, B.M.; Baudonck, N.; De Vel, E.; Dhooge, I. – “Comparison of the overall intelligibility, articulation, resonance, and voice characteristics between children using cochlear implants and those using bilateral hearing aids: a pilot study.” *Int J Audiol*. 2005; 44: 452-465. (a)

212. Van Lierde, K.M.; Vinck, B.; De Ley, S.; Clement, G; Van Cauwenberge, P. – “Genetics of vocal quality characteristics in monozygotic twins: a multiparameter approach.” *J Voice*. 2005; 19: 511-518. (b)
213. Van Lierde, K.M.; Claeys, S.; De Bodt, M.; Van Cauwenberge, P. – “Vocal quality characteristics on children with cleft palate: a multiparameter approach.” *J Voice*. 2004; 18: 254-362. (a)
214. Van Lierde, K.M.; De Ley, S.; Clement, G; De Bodt, M.; Van Cauwenberge, P. – “Outcome of laryngeal manual therapy in four Dutch adults with persistent moderato-to-severe vocal hyperfunction: a pilot study.” *J Voice*. 2004; 18: 467-474. (b)
215. Van Lierde, K.M.; De Ley, S.; De Bodt, M.; Van Cauwenberge, P. – “Long-Term Outcome of Hyperfunctional Voice Disorders Based on a Multiparameter Approach .” *J Voice*. 2007; 21(2): 179-188.
216. Vasiliakis, M.; Stylianau, Y. – “Voice pathology detection based on short-term *jitter* estimations in running speech”. *Folia Phoniatr Logop*. 2009; 61(3): 153-170.
217. Vaziri, G.; Almassganj, F.; Behroozmand, R. – “Pathological assessment of patients`speech signals using nonlinear dynamical analysis”. *Comput Biol Med*. 2010; 40(1): 54-63.
218. Verdolini, K.; Ramig, L. O. – “Review: Occupational Risks for Voice Problems”. *Log Phon Vocol*. 2001; 26: 37-46.
219. Verdolini-Marston, K.; Titze, I.R.; Druker, D.G. – “Changes in phonation threshold pressure whith indiced conditions of hydration”. *J Voice*. 1990; 4: 142-51.
220. Verdolini, K.; Rosen, C.A.; Branski, R. (2006) – *Classification manual for voice disorders – I. SID 3, Voice and Voice Disorders*. ASHA.
221. Vieira, M.N.; Maran, A.G.; McInnes, F.G.; Jack, M.A. – “Detecting arytoids cartilage misplacement through acoustic and electroglottographic *jitter* analysis.” *In: 4th International Conference on Spoken Language Processing*. Philadelphia; 1996.

222. Vieira, M.N.; McInnes, F.R.; Jack, M.A. – “On the influence of laryngeal pathologies on acoustic and electroglottographic *jitter* measures.” *J Acoust Soc Am.* 2002; 111(2): 1045-55.
223. Vieira, V.P.; De Biase, N.; Pontes, P. – “Análise Acústica e Perceptivo-Auditiva vs Coaptação Glótica em Alteração Estrutural Mínima.” *Acta ORL.* 2005; 6-12.
224. Vilkmán, E. – “Voice Problems at work: a challenge for occupational safety and health arrangement”. *Folia Phoniatr Logop.* 2000; 52: 120-125.
225. Webb, A.L.; Carding, P.N.; Deary, I.J.; Markenzie, K.; Steen, N.; Wilson, J.A. – “The reliability of three perceptual evaluation scales for dysphonia.” *Eur Arch Otorhinolarygol.* 2004; 261: 429-34.
226. Weber, S.A.T (2002) – *Estudo das relações laríngeas e do grau de disfonia em pacientes com paracoccidiodomicose.* Dissertação (Mestrado) – Universidade do Estado de São Paulo. Botucatu.
227. Wirz, S.; Beck, J.M. (1995) – *Assessment of Voice Quality: The vocal profile analysis scheme.* Whurr: London.
228. Wolfe, V.; Steinfatt, T. – “Prediction of vocal severity within and across voice types.” *J Speech Language Hear Res.* 1987; 30: 230-40.
229. Wolfe, V.; Martin, D. – “Acoustic correlates of dysphonia: type and severity.” *J Commun Disord.* 1997; 30: 403-416.
230. Wolfe, V.; Martin, D.; Pamer, C. – “Perception of dysphonic voice quality by naive listeners.” *J Speech, Lang & Hear Res.* 2000; 43: 697-705.
231. Wuyts, F.L.; De Bodt, M.S.; Bruckers, L.; Molenberghs, G. – “Normative data stem.” *ACTA Othorhino-laryngologica BELGICA.* 1996; 50: 331-341.
232. Wuyts, F.L.; De Bodt, M.S.; Molenberghs, G.; Remacle, M. Heyler, L.; Millett, B. – “The Dysphonic Severity Index: an objective measure of vocal quality based on a multiparameter approach”. *J Speech, Language and Hearing Research.* 2000; 43: 796-809.
233. Wuyts, F.L.; De Bodt, M.S.; Van de Heyning, P.H. – “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.” *J Voice.* 1999; 13(4): 508-517.

234. Xue, S.A.; Deliyski, D. – “Effects of aging on selected acoustic voice parameters: preliminary normative data and educational implications.” *Educ Geront.* 2002; 22(2): 159-168.
235. Yamasaki, R.; Leão, S.H.S; Madazio, G.; Padovani, M.; Azevedo, R. – *Análise perceptivo-auditiva de vozes normais e alteradas: Escala analógica visual.* In: XV Congresso Brasileiro de Fonoaudiologia e VII Congresso Internacional de Fonoaudiologia. Out 2007; Gramado – RS.
236. Yamasaki, R.; Leão, S.H.S; Madazio, G.; Padovani, M.; Azevedo, R.; Behlau, M. – *Correspondência entre Escala Analógico-Visual e a Escala Numérica na Avaliação Perceptivo-Auditiva de Vozes.* In: XVI Congresso Brasileiro de Fonoaudiologia. 2008; Campos de Jordão – SP.
237. Yiu, E.; Worrall, L.; Longland, J.; Mitchell, C. “Analysing vocal quality of connected speech using Kay’s computerized speech lab: a preliminary finding”. *Clinical Linguistics & Phonetics.* 2000; 14 (4): 295-305.
238. Yiu, E.M. – “Impact and Prevention of Voice Problems in the Teaching Profession: Embracing the Consumer’s View”. *J Voice.* 2002; 16(2): 215-28.
239. Yu, P.; Ouaknine, M.; Giovanni, A. “Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements”. *J Voice.* 2001; 15: 529-542.
240. Yu, P.; Revis, J.; Wuyts, F.L.; Zanaret, M.; Giovanni, A. – “Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analogue scale.” *Folia Phoniatr Logop.* 2002; 54: 271-281.
241. Yumoto, E.; Sasaki, Y.; Okamura, H. – “Harmonics-to-noise ratio and psychological measurement of the degree of harshness.” *J Speech Hear Res.* 1984; 27: 2-6.
242. Zhang, Y.; Jiang, J.J. “Acoustic Analyses of Sustained and Running Voices of Patients with Laryngeal Pathologies”. *J Voice.* 2008; 22: 1-9.
243. Zraick, R.R.; Wenderl, K.; Smith-Olinde, L. – “The effect of speaking task on perceptual judgement of the severity of dysphonic voice.” *J Voice.* 2005; 19: 574-581.

