



Universidade de Aveiro
2007

Departamento de Electrónica, Telecomunicações e
Informática

Secção Autónoma de Ciências da Saúde

Departamento de Línguas e Culturas

**Maria Helena Borges
Aguiar Vilarinho
Machado Castro**

**Estudo de Voz em Laringectomizados Falantes de
Português Europeu**



Universidade de Aveiro
2007

Departamento de Electrónica, Telecomunicações e
Informática

Secção Autónoma de Ciências da Saúde

Departamento de Línguas e Cultura

**Maria Helena Borges
Aguiar Vilarinho
Machado Castro**

**Estudo de Voz em Laringectomizados Falantes de
Português Europeu**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Ciências da Fala e da Audição, realizada sob a orientação científica do Doutor António Joaquim da Silva Teixeira Professor Auxiliar do Departamento de Electrónica Telecomunicações e Informática da Universidade de Aveiro

O júri

presidente

Doutor Nelson Pacheco da Rocha
Professor Catedrático da Universidade de Aveiro

vogais

Doutor Jorge Eduardo de Freitas Spratley
Professor Auxiliar Convidado da Faculdade de Medicina da Universidade Do
Porto

Doutor António Joaquim da Silva Teixeira
Professor Auxiliar da Universidade de Aveiro (Orientador)

palavras-chave

Laringectomia, Voz esofágica, Voz traqueoesofágica, Análise Acústica.

resumo

Motivada pela prática clínica com utentes laringectomizados, procurou-se estudar os parâmetros acústicos de dois tipos de voz alternativos à voz laríngea. De forma a perceber o comportamento destas vozes no meio ambiente, realizou-se um teste de percepção integrando as vozes já referidas. Os resultados obtidos indicam-nos que há diferenças entre os dois tipos de voz que devem ser analisadas previamente à abordagem de intervenção médica e terapêutica.

keywords

Laryngectomy, Esophageal Voice, Tracheoesophageal Voice, Acoustic Analysis, Perceptual Evaluation.

abstract

Motivated by Clinical Practice with Laryngectomized patients, two types of alternative voice were analyzed in terms of several acoustic parameters. In order to understand the reactions to these voices, a perceptual test was, also, performed. Results indicate the existence of differences between the two types that should be considered before medical intervention.

Índice

Capítulo 1 - Introdução.....	1
1.1 Motivação.....	1
1.2 Objectivos.....	2
1.3 Estrutura da Dissertação.....	2
Capítulo 2 - Voz.....	5
2.1 Produção de voz.....	5
2.2 Vogais do Português Europeu.....	8
2.3 Parâmetros relacionados com a qualidade da voz.....	9
2.3.1 Duração.....	10
2.3.2 Frequência Fundamental (F_0).....	10
2.3.3 <i>Jitter</i>	10
2.3.4 <i>Shimmer</i>	11
2.3.5 Medidas de ruído.....	12
2.3.6 Formantes (F_1 e F_2).....	13
Capítulo 3 - Voz Esofágica e Traqueo-Esofágica.....	15
3.1 Carcinoma Laríngeo.....	15
3.1.1 Etiologia.....	15
3.1.2 Procedimentos de avaliação.....	16
3.1.3 Classificação dos tumores.....	21
3.1.4 Tratamento dos carcinomas laríngeos.....	23
3.1.5 Laringectomia Total.....	24
3.2 Reabilitação Vocal.....	25
3.2.1 Voz esofágica.....	26
3.2.2 Voz Traqueo - Esofágica.....	31
3.2.3 Comparações entre os dois métodos.....	33
Capítulo 4 - Corpus.....	37
4.1 Introdução.....	37
4.2 Definição do Corpus.....	37
4.3 Gravação.....	39
4.3.1 Informantes.....	39
4.4 Anotação.....	40
4.5 Extração de parâmetros.....	42
Capítulo 5 - Análise Acústica.....	45
5.1 Resultados.....	45
5.1.1 Duração.....	46
5.1.2 Frequência fundamental.....	49

5.1.3	Jitter	53
5.1.4	Shimmer	56
5.1.5	Medidas de ruído (Proporção Harmónico-Ruído).....	57
5.1.6	Formantes	61
5.2	Discussão.....	66
Capítulo 6 - Avaliação Perceptiva.....		71
6.1	Introdução.....	71
6.2	Método.....	73
6.3	Resultados	74
6.3.1	Resultados do Teste de Identificação	75
6.3.2	Resultados do Teste de Avaliação da Qualidade Vocal.....	81
Capítulo 7 - Conclusões e Trabalho Futuro		85
7.1	Resumo do trabalho realizado.....	85
7.2	Principais resultados e conclusões	86
7.3	Limitações do estudo.....	87
7.4	Desenvolvimentos Futuros.....	87
Bibliografia.....		88

Capítulo 1 - Introdução

1.1 Motivação

A Laringectomia Total é um procedimento cirúrgico no qual é retirada toda a laringe e estruturas adjacentes sendo, a alteração mais imediata, a presença de um orifício na região anterior do pescoço (traqueostoma) e a perda da capacidade de produção de Voz.

Após a cirurgia e no decorrer do processo de intervenção terapêutica, o utente tem disponíveis alguns métodos de reabilitação que poderão ajudá-lo no seu regresso à vida activa de forma mais ou menos rápida. Os mais amplamente utilizados são a voz esofágica e a reabilitação através da colocação de uma prótese traqueo-esofágica. As diferenças entre estes métodos são várias e o estudo das suas características acústicas - duração das vogais, frequência fundamental (F0), a intensidade, o *jitter*, o *shimmer*, o *Harmonic-to-noise-ratio* e as formantes - é fundamental para percebermos se valerá a pena optar por um ou outro método de aquisição de voz.

Na prática clínica da autora, a intervenção com este tipo de pacientes é diária e, durante anos, no serviço onde exerce a sua actividade apenas era possível a opção pela aquisição da voz esofágica. Entretanto, com o decorrer do tempo, uma nova visão sobre a reabilitação destes pacientes tornou possível, no serviço, a colocação de próteses traqueo-esofágicas em utentes laringectomizados. De uma forma geral, parece haver vantagens na utilização deste método, mas sentiu-se a necessidade de comparar os dois grupos de utilizadores de forma objectiva e perceber se, na realidade, a opção pela prótese será a mais vantajosa.

O estudo das características acústicas da voz esofágica e traqueo-esofágica em utentes submetidos a Laringectomia Total em Portugal é escasso e mesmo os estudos já realizados não estão publicados. Os trabalhos realizados para análise destas características encontram-se publicados nas línguas espanhola e inglesa e, mesmo estes carecem de análises mais aprofundadas que nos permitam saber se há ou não diferenças significativas que possam possibilitar aos utilizadores destes métodos uma inscrição mais efectiva na sua vida social, após a realização de uma cirurgia tão extensa como é a laringectomia total.

1.2 Objectivos

Os objectivos principais deste estudo são: (1) analisar as características acústicas das vogais orais do Português em sujeitos submetidos a Laringectomia Total, utilizadores de Voz Esofágica (VE) e Traqueo-esofágica (VTE) tais como: os parâmetros de duração das vogais, a sua frequência fundamental, os valores de *jitter*, *shimmer*, *Harmonis-to-noise Ratio* e os valores das primeiras formantes (F1 e F2) de cada um dos grupos; (2) comparar os valores acima referidos entre os dois grupos utilizadores de voz esofágica (VE) e traqueo-esofágica (VTE); (3) comparar os resultados obtidos com estudos realizados para outras línguas e verificar a existência de dados comuns.

Para consecução destes objectivos, na segunda parte do trabalho, optou-se pela realização de um teste perceptivo de identificação e avaliação da qualidade vocal dos utilizadores de VE e VTE que pretende: (1) avaliar de forma subjectiva a qualidade vocal de cada um dos utilizadores de VE e VTE; (2) analisar a performance em tarefas de identificação de 2 grupos de ouvintes, um grupo familiarizado com estes tipos de voz e outro grupo não familiarizado; (3) verificar que tipo de voz tem maior preferência entre o grupo de ouvintes; (4) analisar os dados e verificar a existência de correlações entre os dados da análise acústica e os obtidos na avaliação subjectiva.

1.3 Estrutura da Dissertação

Este trabalho está dividido em sete capítulos. Este capítulo inicia-se com a introdução geral ao trabalho realizado, apresentando a motivação e os objectivos. O segundo capítulo refere-se ao enquadramento teórico sobre a produção de voz normal, sendo feita uma breve abordagem aos sistemas que contribuem para a sua formação. Ainda neste capítulo, é feita uma descrição das vogais orais do Português Europeu e dos parâmetros que caracterizam a qualidade vocal dos falantes. No terceiro capítulo descreve-se a patologia oncológica abordada neste trabalho, métodos de diagnóstico e reabilitação. No quarto capítulo está descrito o Corpus, abordando-se os aspectos relacionados com a sua gravação, descrição dos informantes, anotação e extracção dos

parâmetros. No quinto capítulo é feita toda a análise acústica dos parâmetros previamente extraídos, como as medidas de duração, frequência fundamental, *Harmonic-to-noise-ratio*, *Jitter*, *Shimmer*, F1 e F2. O sexto capítulo refere-se à avaliação perceptual com descrição do teste aplicado, dos ouvintes e análise dos respectivos resultados. No capítulo sete apresentam-se as conclusões gerais deste trabalho.

Capítulo 2 - Voz

2.1 Produção de voz

Neste capítulo, pretende-se fazer uma abordagem geral aos mecanismos de produção, aos sistemas necessários à produção e controlo que se consideram essenciais para o conhecimento da Voz.

Para a produção da voz humana existem diversos sistemas que se interligam e em conjunto, são responsáveis pela produção da voz e da fala. O **sistema respiratório** é constituído pelo tracto respiratório superior e pelo tracto respiratório inferior. O tracto respiratório superior é formado por diversos órgãos: cavidade nasal, cavidade oral, faringe e laringe. Estas estruturas têm um papel preponderante na respiração e participam de forma activa nos processos de mastigação, deglutição, fonação, ressonância e articulação (Doyle, 1997; Close et al, 1998; Zemlin, 2000, Behlau, 2001). No tracto respiratório inferior estão incluídas estruturas como a traqueia, brônquios e pulmões. Esta estrutura, também conhecida como árvore traqueobrônquica, tem como principais funções assegurar o suporte respiratório para a sobrevivência e para a produção da voz (Zemlin, 2000; Behlau, 2001; Guimarães, 2007).

O sistema respiratório tem uma função dupla: (1) assegurar o suporte necessário à nossa sobrevivência; (2) fornecer grande parte da energia aerodinâmica necessária aos sistemas responsáveis pela produção de voz que já foram referidos anteriormente.

A respiração para a fala tem características diferentes da respiração vital em diferentes abordagens. O ciclo respiratório apresenta duas fases, separadas por si por um pequeno intervalo: a inspiração e a expiração. Na respiração em repouso, a fase inspiratória é três vezes superior à fase expiratória. Na respiração para a fala torna-se necessária uma maior quantidade de ar, uma velocidade respiratória mais reduzida e uma fase expiratória prolongada (Behlau & Pontes, 1995).

A laringe é um sistema musculo-cartilágneo e neuromuscular com funções complexas que envolvem a protecção das vias aéreas inferiores, a respiração, a deglutição e a fala (fonação e ressonância) (Zemlin, 2000, Behlau, 2001, Guimarães, 2007).

A produção de voz é uma das três funções realizadas pela laringe humana. As funções primárias que asseguram a sobrevivência humana são a respiração e a protecção das vias respiratórias inferiores (encerrando a via respiratória no processo de deglutição e através do reflexo de tosse).

A laringe comunica inferiormente com os pulmões (vias respiratórias inferiores) e, superiormente, com a faringe, cavidade bucal e nariz (vias respiratórias superiores). Assim, sabemos que os órgãos mais intimamente associados à produção de voz são os pulmões, a traqueia, a laringe, as cavidades nasais e a cavidade oral. Estas estruturas formam um sistema versátil e complexo de produção de som (Doyle, 1997; Close et al, 1998; Zemlin, 2000).

A laringe, situada na extremidade superior da traqueia, é uma estrutura ímpar, musculo-cartilaginosa, localizada na linha média, situada na região anterior do pescoço, como mostra a Figura 1. Situa-se entre a traqueia (localizada inferiormente) e o osso hioide (na parte superior), apresentando funções biológicas e “não-biológicas” que serão focadas posteriormente.

Em termos biológicos, a laringe pode ser considerada um componente intrínseco do sistema respiratório, funcionando como um dispositivo protector para as vias aéreas inferiores. Age como uma válvula impedindo que o ar escape dos pulmões, impede a entrada de substâncias estranhas na laringe e expulsa, vigorosamente, as substâncias estranhas que ameacem entrar na laringe ou na traqueia (Zemlin, 2000; Pinho, 1998).

Em termos não-biológicos, podemos dizer que a principal função da laringe é produzir som. Como a fala é parte integrante do comportamento humano e é através dela que somos capazes de comunicar com os outros e dar a conhecer as nossas vontades e necessidades, a noção de que esta função não é biológica está sujeita a críticas. Contudo, o importante é ter a noção de que a laringe funciona como um gerador de som, apenas quando não está a realizar as funções biológicas vitais, já anteriormente referidas (Doyle, 1997; Close et al, 1998; Zemlin, 2000).

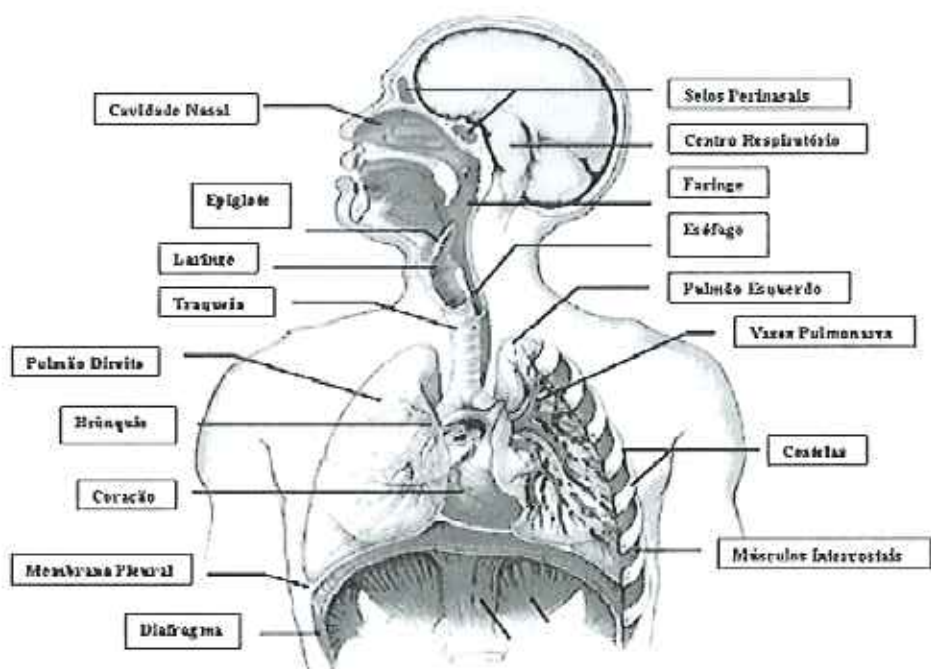


Figura 1 – Esquema representativo dos sistemas respiratório e fonador. Adaptada de <http://www.ama-assn.org>.

A produção de voz é uma função secundária não vital da laringe. Contudo, para o mundo em que vivemos, trata-se de uma função importantíssima uma vez que tem um papel de transmissora de palavras, produtora de musicalidade e é, também, um meio de expressão de emoções agindo como espelho do “eu interior” de cada indivíduo.

Tal como referido anteriormente, a laringe é o órgão principal da produção de voz. Contudo, “...esta soaria extremamente fraca se não houvesse uma influência acústica das estruturas ressoadoras supraglóticas do tracto vocal (cavidade laringea, faríngea, bucal e nasal)” (Guimarães, 2007). São estas estruturas, onde o fluxo de ar laríngeo é modulado e onde é conferido um determinado timbre, que nos permitem distinguir entre sons consonânticos e sons vocálicos (Moutinho, 2000).

O sistema de ressonância é constituído pela cavidade faríngea, pela cavidade bucal e pela cavidade nasal. A cavidade faríngea tem a forma de cone e estende-se desde a base do crânio até ao bordo inferior da cricóide. Serve de trajecto comum entre o tracto respiratório e digestivo e funciona como ressoador supraglótico (Zemlin, 2000; Guimarães, 2007).

A cavidade bucal, contígua à cavidade faríngea, é delimitada anteriormente pelos lábios, lateralmente pelas bochechas, no plano superior é delimitada pelas

bochechas e na parte posterior pelo arco palatal (Zemlin, 2000). Nesta cavidade, situam-se os articuladores. No maxilar inferior encontram-se os órgãos articuladores (órgãos que tocam ou se aproximam de um determinado ponto de articulação) e no maxilar superior (articulador passivo), encontram-se os pontos de articulação (lugares em direcção aos quais se dirigem os órgãos articuladores, o lábio inferior e a língua).

Os traços articulatorios, com que distinguimos os sons da fala, têm na origem esses pontos de articulação: lábio superior, dentes, alvéolos, pré-palato, palato duro, palato mole ou *velum* e úvula (Moutinho, 2000). É importante salientar que o véu palatino, sendo móvel, pode ou não elevar-se consoante o tipo de actividade que está a ser realizada. Na fala e para os sons do Português Europeu, se o véu se eleva e toca na parede posterior da faringe os sons são orais, se não se eleva, o ar passa simultaneamente pela cavidade bucal e nasal, dando origem aos sons nasais (Moutinho, 2000).

A classificação articulatória dos sons da fala consiste na categorização dos sons que se baseia essencialmente nos movimentos dos articuladores (Matcus, Falé & Freitas, 2005 citados por Martins, 2007). Os sons da fala podem ser divididos em vogais e consoantes.

Neste estudo, pretende-se uma análise mais pormenorizada das vogais orais do Português Europeu, pelo que, esta abordagem, será mais abrangente na secção seguinte deste trabalho.

2.2 Vogais do Português Europeu

As vogais são produzidas sem oposição significativa ao fluxo de ar proveniente dos pulmões. Caracterizam-se pela passagem livre do ar e vibração das pregas vocais, pelo que, todas elas, são consideradas vozcadas (Moutinho, 2000). Assim, não havendo obstrução do tracto vocal, não se faz referência a pontos de articulação classificando-se estas em função de diversos parâmetros: altura (posição do dorso da língua), grau de abertura e posição dos lábios ou labialização (Moutinho, 2000). Relativamente à altura, são consideradas 3 posições – alta, média e baixa. Em repouso, a língua está numa posição central e média dentro da cavidade oral (Matcus, Falé & Freitas, 2005).

O grau de abertura depende da altura do dorso da língua e da abertura do maxilar inferior no momento da realização da vogal, podendo então, as vogais, ser classificadas

como: fechadas (ex: [i]) semi-fechadas (ex: [o]), semi-abertas (ex: [E]) e abertas (ex: [a]) (T'cixcira, 2000; Moutinho, 2000).¹

Para o mesmo grau de abertura, o dorso da língua desloca-se no sentido antero-posterior, o que permite uma classificação relativa à zona de articulação: anteriores, avançadas ou palatais (avanço da língua para a posição neutra e palato); médias ou centrais (dorso da língua em posição média); e posteriores, recuadas ou velares (dorso da língua elevada em direcção ao véu palatino) (Martins, 2007).

A labialização indica o papel desempenhado pelos lábios na produção da vogal. A ausência ou presença da projecção dos lábios permite a distinção entre vogais labializadas ou arredondadas e não labializadas ou não arredondadas. No Português Europeu só as vogais posteriores ou velares são arredondadas [o, O, u], todas as outras são não labializadas ou não arredondadas (Moutinho, 2000, T'cixcira, 2000). Para uma melhor compreensão do que atrás foi exposto apresenta-se de seguida uma tabela de classificação das nove vogais orais do Português Europeu [a, ɒ, E, @, e, i, o, O, u].

GRAU DE ABERTURA	ZONA DE ARTICULAÇÃO			ALTURA DA LÍNGUA
	Anteriores	Médias	Posteriores	
Abertas		[a]		Baixa
Semi-abertas	[E]		[O]	Baixa
Semi-fechadas	[e]	[ɒ]	[o]	Média
Fechadas	[i]	[@]	[u]	Alta

Tabela 1 – Classificação das vogais orais do Português Europeu. Fonte: Moutinho (2000)

2.3 Parâmetros relacionados com a qualidade da voz

As vogais são caracterizadas por determinados parâmetros acústicos. Os parâmetros abaixo descritos foram os utilizados, neste estudo, para comparar os grupos de utilizadores de VE e VTE analisados. Os parâmetros que maior referência apresentam na literatura são: duração, frequência fundamental (F_0), *jitter*, *shimmer*, *Harmonic-to-noise ratio*, e F1 e F2 (primeira e segunda formantes).

¹ Neste trabalho optou-se pela utilização do Alfabeto Fonético SAMPA.

2.3.1 Duração

O som é composto por diversos parâmetros, um dos quais a duração que corresponde à quantidade de tempo gasto na realização de um som ou sequência de sons (Pinto Moura, 2006).

2.3.2 Frequência Fundamental (F_0)

Para uma melhor compreensão do estudo deste parâmetro é importante salientar que, no estudo das perturbações da voz, os termos *pitch* e frequência apesar de relacionados, podem ser confundidos e ter consequências clínicas bastante expressivas. Assim, importa salientar que o termo *pitch* se refere à sensação subjectiva de frequência sendo, por isso, uma medida perceptiva que se classifica num contínuo agudo-grave. A frequência, por outro lado, refere-se sempre a um atributo físico de certos sinais: a velocidade de repetição de uma determinada forma de onda (Baken & Orlikoff, 2000). Para se medir F_0 assume-se que o sinal é pelo menos aproximadamente periódico, o que significa que é caracterizado por um tipo de onda que se repete a uma velocidade fixa e invariável. A cada repetição da onda chamamos *ciclo* (Baken & Orlikoff, 2000).

A frequência fundamental, abreviadamente conhecida como F_0 , é descrita como o número de vezes que uma onda se repete numa determinada unidade de tempo. É medida em *hertz* (Hz) e 1 Hz é igual a um ciclo por segundo (Baken & Orlikoff, 2000).

2.3.3 Jitter

O *jitter* é uma medida de curto termo (ciclo a ciclo), de variabilidade não voluntária de F_0 , que permite determinar o grau de estabilidade do sistema fonatório (Baken & Orlikoff, 2000; Guimarães, 2007).

A extração dos valores de *jitter* pode ser realizada através de medidas relativas ou absolutas. As medidas absolutas não entram em linha de conta com os valores de F_0 produzidos pelo indivíduo. São medidas conseguidas através do factor de perturbação ou através do factor de perturbação direccional. O factor de perturbação é definido como a percentagem de ciclos numa vogal sustentada que tem uma variação igual ou maior a 0,5 milissegundos, e o factor de perturbação direccional considera apenas o

número de vezes que F_0 muda de sinal. Os resultados são apresentados em segundos, milissegundos ou microsegundos (Baken & Orlikoff, 2000).

As medidas relativas têm em conta a média de F_0 . As frequências mais elevadas apresentam, habitualmente, um menor índice de perturbação. Existem diversas possibilidades para analisar o *jitter*. O programa *Praat* utilizado neste estudo, analisa as medidas acústicas utilizando várias medições: *jitter (local)*, *jitter (local absoluto)*, *jitter rap*, *jitter ppq5* e *jitter ddp*, (Boersma & Weenink, 1992-2005; Casper & Colton, 1998).

O *jitter (local)* é a diferença média absoluta entre dois períodos consecutivos, dividida pela média do período; o *jitter (local absoluto)* é a diferença média absoluta entre dois períodos consecutivos; o *jitter (RAP – Relative Average Perturbation)* é a diferença média absoluta entre um período e a sua média e dos seus dois períodos consecutivos, dividida pela média do período; o *jitter ppq5* é a diferença média absoluta entre um período e a sua média e a dos seus quatro períodos vizinhos, dividida pela média do período; o *jitter ddp* é a diferença média absoluta entre diferenças consecutivas de períodos consecutivos, dividida pela média do período. Este valor é três vezes o valor de RAP (Boersma & Weenink, 1992-2005).

A variabilidade do *jitter* nas diferentes vogais, no adulto, continua a não encontrar consenso. Alguns investigadores concordam que a magnitude do *jitter* é influenciado pelo tipo de vogal e, outros, entre os quais Orlikoff, referem que não existem razões fisiológicas para que isso aconteça (Guimarães, 2007).

A literatura refere valores de *jitter* mais elevados para a vogal [a], que é uma vogal baixa, do que para as vogais [i] e [u] que são vogais mais altas. Contudo, outros autores referem precisamente o inverso.

2.3.4 Shimmer

A perturbação da amplitude (*shimmer*), outro índice de estabilidade da fonte de som, é uma medida de amplitude da vibração, que quantifica as alterações mínimas da amplitude do sinal, com base em cada ciclo fonatório (Casper & Colton, 1998; Guimarães, 2007).

Os valores de *shimmer* parecem ser tão importantes como os de *jitter* e parecem contribuir, de forma significativa, para a percepção da rouquidão. (Baken & Orlikoff, 2000). A relação entre a amplitude da perturbação e algumas anomalias da função

glótica não estão devidamente clarificadas, mas aquilo que parece evidente é que, se a patologia laringea for eliminada, os valores de *shimmer* tendem a normalizar (Hirano et al, 1988)

Tal como o *jitter*, o *shimmer* tende a ser ligeiramente elevado em perturbações de voz, apresentando valores mais altos nas frequências graves e de baixa intensidade (Behlau, 2001; Guimarães, 2007).

Hoje em dia, a maioria dos sistemas computadorizados providencia medidas de variabilidade da intensidade embora, os diferentes sistemas, apresentem diferentes valores de *shimmer* para o mesmo sinal. Assim, os utilizadores devem estar bastante seguros e entender o método através do qual foram obtidos os valores de *shimmer* e compará-los com os dados normativos. (Bakcn & Orlikoff, 2000).

No programa de voz utilizado neste estudo (*Praat*) existem diferentes medidas para analisar o *shimmer*: *shimmer (local)*, *shimmer (local, dB)*, *shimmer (apq3)* e *shimmer (apq5)*.

O *shimmer (local)* é a média absoluta das diferenças entre as amplitudes de períodos consecutivos dividida pela média da amplitude. O *shimmer (local, dB)* mede a diferença absoluta do logaritmo de base 10, da diferença entre as amplitudes de períodos consecutivos multiplicado por 20. O *shimmer apq3* mede o quociente de perturbação da amplitude em três pontos, ou seja, a média absoluta da diferença entre a amplitude de um período e a média das amplitudes dos seus vizinhos, dividida pela média da amplitude. O *shimmer apq5* mede o quociente de perturbação da amplitude em cinco pontos, ou seja, a média absoluta da diferença entre a amplitude de um período e a média das amplitudes dos quatro vizinhos, dividida pela média da amplitude.

2.3.5 Medidas de ruído

A proporção harmónico ruído, conhecida em inglês como *Harmonic-to-noise-ratio* (HNR), é uma medida que relaciona a componente harmónica versus a componente de ruído da onda acústica. Isto é, se o ruído espectral (inter-harmónico) aumenta, parece lógico que a energia da frequência dos harmónicos baixe. (Guimarães, 2007). Por esta razão, criou-se o cálculo matemático designado por Índice sinal-ruído, que é a divisão entre a média da amplitude do sinal e a medida da amplitude do componente de ruído da onda. A unidade usada é o dB (Guimarães, 2007).

Yumoto et al. (citado por Baken & Orlikoff, 2000) consideram que, para a voz normal, a média de HNR normal é de 11,9 dB e Behlau (1997) refere que os valores médios de normalidade para os homens são de 11,8 dB (no registo modal), considerando patológicos os valores inferiores a 7 dB.

2.3.6 Formantes (F1 e F2)

Tal como já foi referido anteriormente, as vogais são sons produzidos com a vibração das pregas vocais e sem constricções no tracto vocal, sendo a sua representação visual bastante expressiva. As vogais são facilmente reconhecidas no espectograma porque exibem uma estrutura formancial bem definida. (Mateus, Falé & Freitas, 2005).

A análise acústica das vogais pode efectuar-se com base na extracção de parâmetros dos respectivos espectros sonoros, onde se encontram representados os valores da fonte de vibração (F0), os seus múltiplos inteiros (harmónicos) e os picos de ressonância (formantes). Os valores da frequência dos formantes estão relacionados com a influência da configuração do tracto vocal perante a passagem da onda sonora. Cada vogal tem uma gama de valores definidos para a frequência dos formantes o que as permite distinguir entre si (Rosique, 2003).

Os formantes que caracterizam as diferentes vogais resultam das diferentes formas do tracto vocal. Qualquer corpo de ar, tal como o tracto vocal ou uma garrafa, vibra de uma forma que depende do seu tamanho e da sua forma. Se, por exemplo, soprarmos pelo gargalo de uma garrafa vazia, produzimos um som de baixa frequência, mas, se enchermos parcialmente essa garrafa com água, como o volume de ar diminui, conseguimos produzir tons de frequência mais elevada (Ladefoged, 2001)

Se no caso das consoantes, o tracto vocal realiza diferentes configurações (obstrução parcial ou total), a configuração que permite a realização das vogais, caracteriza-se pela passagem livre do ar pelas cavidades de ressonância acima da região glótica (Moutinho, 2000). As diferentes formas e tamanhos que assumem as cavidades referidas, no momento da articulação de cada vogal, definem ressoadores com características próprias para certas frequências, chamadas formantes. Estes são, então, zonas de frequência intensificadas pelas diferentes zonas de ressonância (Moutinho, 2000; Mateus, Falé & Freitas, 2005).

Apesar de as vogais apresentarem um maior número de formantes, os três primeiros são os mais importantes e, habitualmente, os dois primeiros contêm informação suficiente para a sua identificação (Mateus, Falé & Freitas, 2005). O primeiro formante (F1) está relacionado com o parâmetro de altura do dorso da língua na produção das vogais: um F1 baixo caracteriza as vogais altas e um valor de F1 elevado caracteriza as vogais mais baixas (Mateus, Falé & Freitas, 2005). Assim, não será de surpreender que a vogal [a] tenha um valor de F1 elevado, e que as vogais [i] e [u] tenham um valor de F1 mais baixo (Moutinho, 2000).

Os parâmetros de ponto de articulação, recuo/avanço do dorso da língua e o grau de arredondamento dos lábios são representados pelo segundo formante (F2): a um F2 baixo, equivale uma vogal posterior e produzida com arredondamento dos lábios e, a um F2 de valor elevado, corresponde uma vogal anterior sem arredondamento dos lábios. Então, [a] e [i] apresentam valores de F2 superiores aos valores encontrados para a vogal [u]. Na realidade, quanto mais recuado estiver o dorso da língua, mais baixo o valor de F2 (Moutinho, 2000; Mateus, Falé & Freitas, 2005).

Na figura 2 apresenta-se o triângulo acústico das vogais tónicas do Português Europeu, proposto por Delgado Martins (1988):

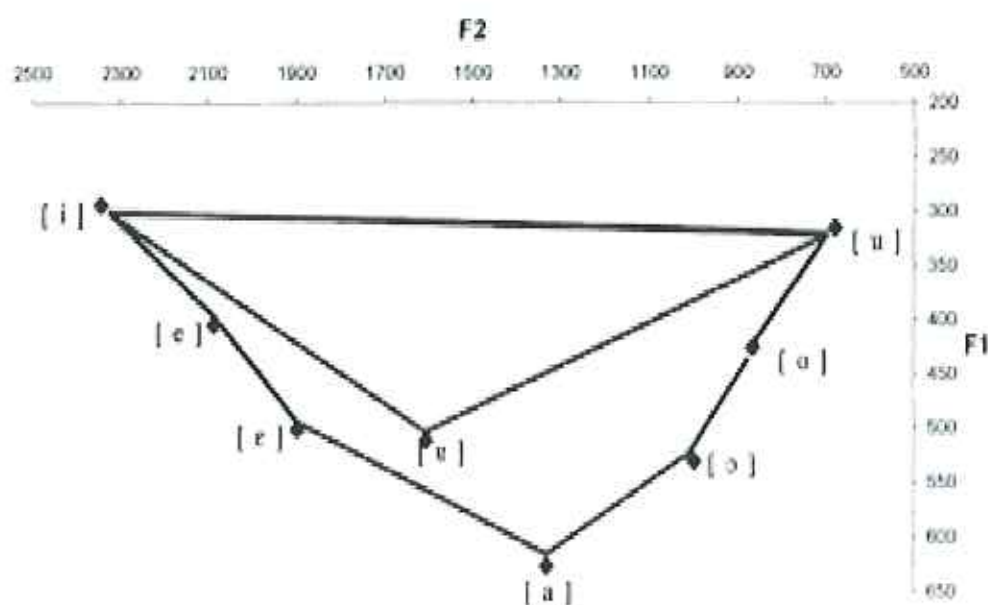


Figura 2 - Triângulo Acústico das vogais orais tónicas do PE padrão. Fonte: Delgado Martins (1988)

Capítulo 3 - Voz Esofágica e Traqueo-Esofágica

3.1 Carcinoma Laríngeo

O carcinoma laríngeo é o mais comum dos cancros da região da cabeça e pescoço.

De acordo com a EUCAN (Associação Europeia de Cancro), Portugal é o segundo país europeu com maior índice de mortalidade por cancro na laringe. O cancro da laringe é muito frequente no nosso País e sabe-se que, anualmente, surgem cerca de 1000 casos/ano e morrem mais de 400 pessoas com esta doença (Coutinho, 2005).

O carcinoma laríngeo é uma patologia cada vez mais frequente e com maior taxa de ocorrência em indivíduos do sexo masculino, com idades compreendidas entre os 50 e os 70 anos (Coutinho, 2005).

Os tumores laríngeos, quando diagnosticados precocemente, apresentam um bom prognóstico em termos de tratamento conservador, podendo ser tratados com radioterapia ou cirurgia laser. Contudo, consoante a fase em que a patologia é detectada e a sua extensão, o tratamento pode passar por uma intervenção cirúrgica mais radical - a remoção da totalidade da laringe (Corinna, 2001).

3.1.1 Etiologia

O consumo de tabaco é um factor etiológico significativo para o desenvolvimento do carcinoma da laringe em homens e mulheres. Está directamente relacionado com medidas de uso tais como a quantidade e duração do hábito. Estudos revelam que dez, anos após o abandono do consumo de tabaco, os riscos de mortalidade são os mesmos do que para um não-fumador. Parece haver um efeito potenciador entre o consumo de álcool e de tabaco que aumenta o nível de risco de carcinoma laríngeo (Casper & Colton, 1998).

O indivíduo típico, diagnosticado com carcinoma laríngeo, é do sexo masculino, tem cerca de 60 anos, hábitos tabágicos marcados, sendo um consumidor moderado ou

excessivo de álcool. Contudo, nos últimos anos, a incidência desta patologia tem vindo a aumentar nas mulheres (Casper & Colton, 1998).

Apesar de encontrarmos algumas famílias com mais do que um elemento diagnosticado com este carcinoma, não existem estudos que revelem, como factores causais, alterações de ordem genética.

3.1.2 Procedimentos de avaliação

Os cuidados médicos apropriados requerem, sempre, um exame prévio à realização do diagnóstico e ao início do tratamento.

Os otorrinolaringologistas utilizam um grande número de procedimentos para examinarem a laringe. A grande maioria dos procedimentos permite ao examinador a observação da laringe sem anestesia, ou outro tipo de manobras cirúrgicas. Outros são medidas indirectas que permitem a avaliação da função vocal.

O sintoma mais comum do carcinoma laríngeo é a rouquidão, pelo que se torna fundamental a avaliação da função fonatória. Para além disso, é necessário também entender os parâmetros acústicos que estão alterados e reunir todos os dados de modo a fazer o diagnóstico da forma mais correcta possível (Casper & Colton, 1998; Behlau, 2001).

O procedimento mais clássico que visa a avaliação da laringe é a **laringoscopia indirecta**, realizada com o espelho circular de Garcia, desenvolvida pelo professor espanhol de canto Manuel Garcia, Pai da Laringologia Moderna. Este exame realiza-se com auxílio de uma fonte de luz artificial - espelho frontal - colocada em diferentes posições, de tal forma que a imagem é reflectida através do espelho posicionado na faringe do paciente de forma a iluminar as cordas vocais e outras estruturas da laringe, como se pode observar nas Figuras 3a e 3b.



Figura 3a – Vista frontal. Fonte: Netter (1964)

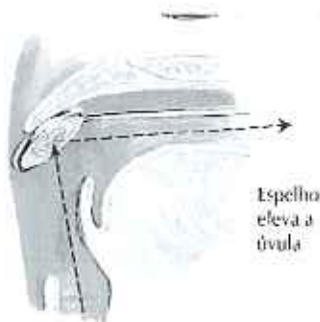


Figura 3b – Vista de perfil. Fonte: Netter (1964)

Este exame tem a vantagem de ser rápido e relativamente fácil de realizar, iniciando-se com a colocação do espelho de Garcia na região posterior da boca entre os pilares, enquanto se direcciona a fonte de luz (espelho frontal) para a região faríngea. Habitualmente não é necessária anestesia e, apesar da observação ser simples, é fundamental um conhecimento profundo de toda a anatomia, fisiologia e patologia da laringe de modo a que o diagnóstico seja o mais correcto possível. A imagem só pode ser observada por breves segundos, uma vez que o paciente só pode sustentar a vogal pedida por pouco tempo. Não há gravação do exame e a fonação limita-se apenas à produção da vogal [i]. Este som é usado por ser a vogal mais anterior, permitindo a anteriorização da epiglote, dilatando a hipofaringe e o vestíbulo laríngeo, permitindo uma óptima visualização das pregas vocais (Casper & Colton, 1998; Pinho, 1998; Behlau, 2001).

Tem como desvantagens não poder ser gravado (já referido), o tempo de fonação da vogal sustentada pedida ao paciente ser de curta duração e haver, em alguns casos, fraca colaboração do utente quer por reflexo de vômito acentuado, quer por dificuldades de amplitude na abertura da boca (Casper & Colton, 1998).

A partir da década de 60, os instrumentos construídos, com o auxílio de fibras ópticas, têm vindo a substituir o espelho circular que, entretanto, continua a prestar um incalçável contributo ao diagnóstico de doenças da laringe (Behlau, 2001).

O princípio desses laringoscópios modernos consiste na visão das estruturas em estudo, mediante a iluminação através de fibras que captam a luz de uma fonte de alta potência; a imagem é trazida até ao olho do observador por um segundo canal de fibras ou por prismas reflectores (Behlau, 2001).

Conservando este princípio, inúmeras modificações e adaptações podem ser introduzidas. Assim, temos o laringoscópio rígido e o flexível cujos diâmetros variam conforme a conveniência. Os laringoscópios rígidos são introduzidos através da boca do paciente, com um tipo de manobra semelhante à laringoscopia indirecta, não havendo grande incómodo para o utente excepto as já referidas para o exame acima descrito (Casper & Colton, 1998; Behlau, 2001).

Os laringoscópios flexíveis, de menor diâmetro, são geralmente introduzidos por via nasal e chegam mais próximo da laringe. Propiciam, não só o estudo funcional da produção de voz em condições praticamente idênticas às situações habituais de fala e canto, mas também das outras funções da laringe - tosse, deglutição e acção esfinctérica.

Causa algum desconforto que pode ser minorado com a aplicação de uma anestesia tópica que relaxe o paciente durante a realização do exame (Behlau, 2001).

A avaliação endoscópica rígida e flexível tem vantagens e desvantagens inerentes a cada um dos procedimentos que serão apresentadas no Quadro 1.

	Endoscopia Flexível	Endoscopia Rígida
Vantagens	Requer pouca colaboração do paciente Imagem não distorcida Trato vocal em posição habitual de emissão Avalia fala encadeada Permite acesso e avaliação das estruturas do tracto vocal Avalia função velofaríngea e aspectos faringolaríngeos da deglutição	Imagem ampliada e com luminosidade Oferece detalhes relativos à mucosa Define aspectos gradativos de coloração da mucosa Favorece o diagnóstico diferencial das lesões laringeas
Desvantagens	Imagem reduzida e com pouca luminosidade Não avalia detalhes da mucosa nem de coloração Não favorece o diagnóstico diferencial das lesões de massa	Requer colaboração do paciente Pode provocar reflexo de vômito Tracto vocal em posição não natural Não permite articulação dos sons da fala – centrada da laringe Restringe movimentos da deglutição

Quadro 1 – Vantagens e desvantagens da endoscopia flexível e rígida (adaptado de Behlau, 2001)

Na Figura 4, apresenta-se uma imagem representativa de uma endoscopia flexível e uma representação da imagem que se observa:

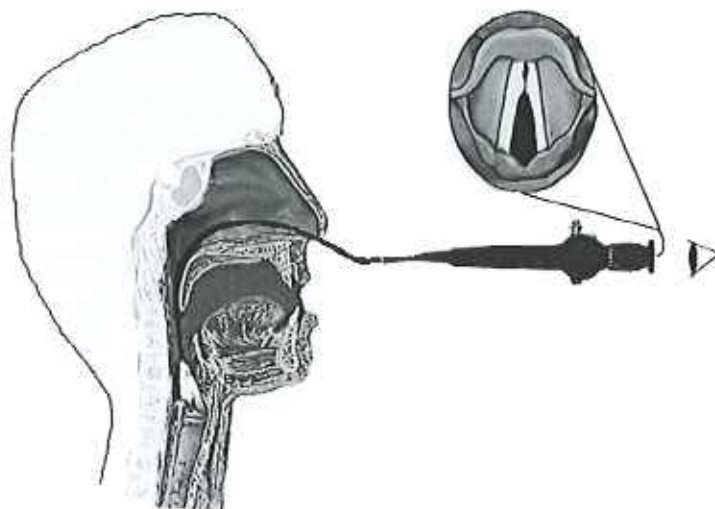


Figura 4 – Imagem representativa de endoscopia flexível onde se observa a introdução do fibroscópio por via nasal até à laringe (acima das pregas vocais).

Na Figura 5, observa-se uma imagem representativa de endoscopia rígida onde, é notória a diferença de realização do exame e o tipo de imagem que se obtém:

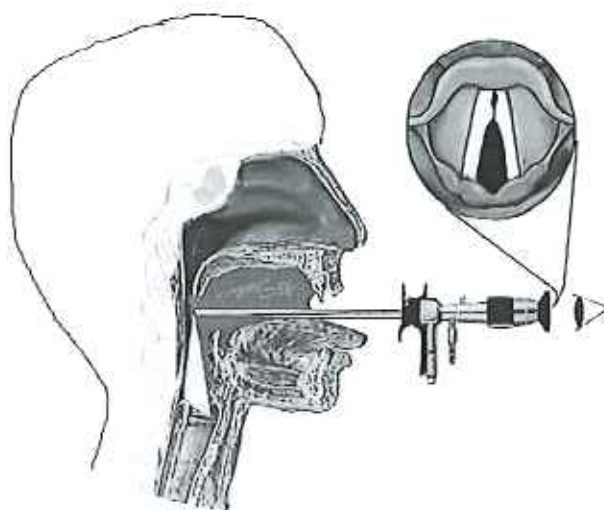


Figura 5 – Imagem representativa de endoscopia rígida onde se observa a colocação do endoscópio na cavidade oral até à parede posterior da faringe.

Actualmente, sabe-se que a qualidade vocal é determinada pelas características da vibração da mucosa das pregas vocais, sendo importantíssima a existência de instrumentos semiológicos que permitam a sua avaliação (Behlau, 2001).

Ainda dentro dos exames de avaliação da laringe, temos a laringoscopia directa, que permite a avaliação das estruturas laríngeas, mediante a introdução de tubos rígidos por via oral até à laringe. Este exame é de extrema importância para o diagnóstico dos tumores da laringe uma vez que, durante a sua realização, é possível efectuar uma biopsia, o que permitirá o estudo mais detalhado do carcinoma. A desvantagem do exame relaciona-se com o facto de não permitir avaliação das estruturas laríngeas em movimento, uma vez que requer anestesia geral ou sedação profunda. A grande vantagem é a possibilidade de ampliação da imagem pelo acoplamento de um microscópio e a manipulação directa das pregas vocais por meio de instrumentos específicos (Behlau, 2001; Pinho, 1998).

Na figura 6 podemos observar a imagem da realização de uma laringoscopia directa, exame de grande relevância para o diagnóstico de tumores da laringe.

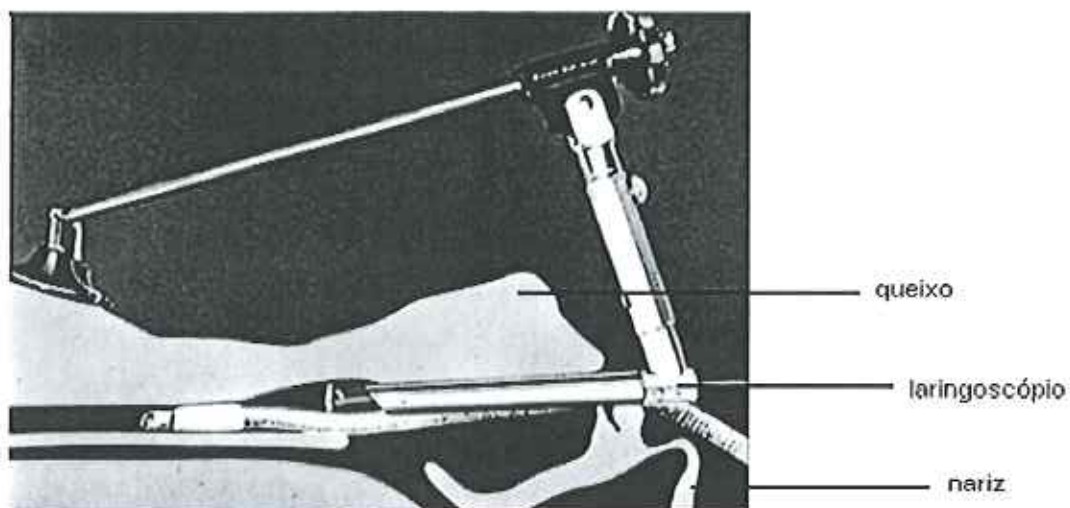


Figura 6 – Imagem representativa de Laringoscopia Directa onde se observa a colocação da cabeça em hiperextensão, com a introdução do laringoscópio a ser realizada por via oral.

3.1.3 Classificação dos tumores

Para que se possa orientar devidamente o tratamento dos tumores da laringe e para que seja possível comparar os resultados obtidos com os diferentes tipos de tratamento, é fundamental que haja uniformidade na sua descrição. Para isso, utiliza-se um sistema denominado TNM - *staging system*, proposto pela União Internacional Contra o Cancro (UICC, 1997) que estadia os tumores de acordo com o local da origem: supraglote (epiglote, pregas ventriculares, aritnoides e pregas ariepiglóticas), glote (pregas vocais, comissura anterior e posterior) e infraglote (desde as pregas vocais até ao primeiro anel traqueal). Nesta classificação o T corresponde à localização primária do tumor, o N à presença de metástases cervicais linfonodais e o M à presença de metástases à distância. No quadro 2, baseado na UICC, observa-se a classificação para estadiamento dos tumores da laringe.

T-stage	Descrição
Tx	Tumor primário sem possibilidade de ser avaliado
T0	Sem evidência de tumor primário
T1s	Carcinoma <i>in situ</i>
T1	Tumor limitado à prega vocal (podendo envolver a comissura anterior e posterior, com mobilidade preservada da prega vocal)
T1a	Tumor limitado a uma prega vocal
T1b	Tumor envolvendo as duas pregas vocais
T2	Tumor que invade a supraglote e/ou infraglote com alteração da mobilidade da prega vocal
T3	Tumor limitado à laringe com fixação da prega vocal
T4	Tumor que invade a cartilagem tiróide com extensão para outros tecidos além da laringe (por exemplo: orofaringe, hipofaringe e tecidos moles do pescoço)
N –stage	
Nx	Impossibilidade de avaliação dos nódulos linfáticos
N0	Ausência de metástases em nódulos linfáticos regionais
N1	Metástase em nódulo linfático ipsilateral de 3 cm até à sua maior dimensão
N2	Metástase num único nódulo linfático, maior que 3 cm até 6 cm na sua maior dimensão; ou múltiplos nódulos ipsilaterais nenhum maior que 6 cm na sua maior dimensão; ou nódulos bilaterais ou contra-laterais nenhum maior que 6 cm na sua maior dimensão
N3	Metástase em nódulo linfático maior do que 6 cm na sua maior dimensão
M –stage	
Mx	Impossibilidade de avaliação de metástases à distância
M0	Ausência de metástases à distância
M1	Presença de metástases à distância

Quadro 2- TNM – Estadiamento dos tumores da laringe (baseado na UICC- adaptado de Close et al, 1998)

Como já foi referido anteriormente, o sintoma mais comum de carcinoma laríngeo é a rouquidão. Nos tumores glóticos este é considerado um sintoma precoce, uma vez que interfere com a vibração das pregas vocais. Nos tumores supra ou infraglóticos pode ser um sintoma tardio que revela a invasão das pregas vocais pelo tumor. Nos tumores de grande dimensão os sinais e sintomas mais referidos são desconforto e dor de garganta, dor irradiada ao ouvido, disfagia, halitose, dispnca, *neck swelling*, sensação de corpo estranho e infecção do pescoço (Doyle, 1994; Casper & Colton, 1998; Close et al, 1998).

3.1.4 Tratamento dos carcinomas laríngeos

De uma forma geral, os tumores laríngeos de pequenas dimensões (T1, T2) são tratados com radioterapia ou, em alguns casos, com cirurgia laser. Os tumores T3 e T4 de pequenas dimensões podem ser tratados com radioterapia ou com um tratamento combinado de radioterapia e quimioterapia. A laringectomia total é mais adequada em tumores T4 mais extensos e também, nos casos de recidiva, após o tratamento com radioterapia. Para tumores que invadam os tecidos exteriores à laringe (faringe e esófago), pode ser necessária a realização de cirurgias mais extensas e radicais que exijam reconstruções bastante complexas (Close et al, 1998; Casper & Colton, 1998).

Em casos de presença de metástases nos nódulos linfáticos, a laringectomia total pode ser realizada juntamente com a dissecação cervical uni ou bilateral (dependendo da localização dos nódulos linfáticos). Se o tratamento primário for a laringectomia total, pode ser necessário a realização de radioterapia pós-operatória, dependendo dos aspectos histológicos característicos do tumor retirado (Close et al, 1998; Casper & Colton, 1998).

3.1.5 Laringectomia Total

Em 1873 foi realizada a primeira laringectomia total para tratamento de um caso de carcinoma laríngeo, realizado por Billroth e descrito pelo seu colega Gussenbauer (1874). Já nesta cirurgia foi focada a importância da reabilitação vocal e pulmonar do paciente em questão e foi referida a criação de uma laringe artificial que permitiria ao paciente usar o ar expirado para falar. Terá sido esta a primeira tentativa de criação de um instrumento semelhante a uma prótese fonatória que foi sendo desenvolvida e aperfeiçoada até aos dias de hoje (Corina, 2001).

Deste procedimento cirúrgico, resultam alterações bastante significativas de carácter funcional, principalmente pela mudança da condição respiratória (traqueostomia definitiva), e de carácter estético, pelas sequelas de esvaziamento cervical, comprometendo uma região bastante visível e valorizada do nosso organismo, a região da cabeça e pescoço (Pinho, 2003).

A laringectomia total está indicada para os casos em que o tumor primário é extenso (T3 ou T4), em casos de extensão à infraglotte ou casos em que o tratamento com radioterapia não resultou (Casper & Colton, 1998; Close et al, 1998).

Numa laringectomia total toda a laringe é retirada, incluindo o osso hióide. O tamanho e a extensão da exérese até a região da faringe, depende, obviamente da extensão do tumor.

As complicações mais imediatas desta cirurgia são a exposição da artéria carótida, a formação de fistulas e o aparecimento de edema. O olfacto e o gosto sofrem algumas alterações, sendo que o primeiro se mantém alterado durante um período mais longo de tempo.

A laringectomia tem, como consequência imediata, a perda temporária da voz e, apesar das alternativas à voz laríngea existirem e estarem disponíveis, requerem sempre uma forma de aprendizagem mais ou menos prolongada. Até estas alternativas estarem estabelecidas, o indivíduo laringectomizado está dependente da escrita, gestos ou voz ciciada para comunicar e estas formas de comunicação, além de limitadas e lentas, podem tornar-se frustrantes. (Casper & Colton, 1998; Pinho, 2003)

3.2 Reabilitação Vocal

As possibilidades de recuperação da capacidade fonatória incluem tentativas de utilização de segmentos anatómicos remanescentes (voz esofágica), meios protéticos (Laringe electrónica), cirúrgicos (shunts traqueoesofágicos) ou protético-cirúrgicos (fístulas traqueoesofágicas), numa abordagem de substituição da função fonatória alterada. (Doyle, 1994)

Em cada uma das alternativas referidas, observam-se vantagens e desvantagens e, a opção deve ser definida pela expressão de autonomia do indivíduo, devendo o esclarecimento de cada uma das alternativas ser dado pela equipa multidisciplinar que acompanha o doente. É necessário ter em conta as condições socioeconómicas e culturais do indivíduo e as possibilidades de procedimento da equipa que o assiste.

É importante referir que a fala requer não apenas a laringe, mas também um bom controlo do tracto vocal. Em algumas cirurgias oncológicas da laringe, são também retiradas algumas porções do tracto vocal. A perda da língua, ou de outras partes do sistema articulatorio pode afectar, severamente, a capacidade de produzir determinados sons. No entanto, não considerando estes casos, existem algumas alternativas funcionais para substituir a fala com voz laríngea.

A **fala murmurada** é produzida na ausência de qualquer vibração sonora e resulta do aproveitamento do ar nos espaços de ressonância, modulado pelos articuladores (Pinho, 1998). Acontece no momento pós-operatório imediato e só deverá ser mantida se o doente rejeitar a laringe electrónica, não conseguir produzir voz esofágica ou se não tiver indicação para prótese.

A **laringe electrónica** é um aparelho mecânico que, quando accionado por contacto na região do pescoço, desencadencia ondas vibratórias numa das suas extremidades e o som é transmitido através dos tecidos para a cavidade oral ou faríngea. Os mecanismos articulatorios normais são mantidos. Este recurso tem como vantagem principal a aquisição rápida e funcional de um meio alternativo de comunicação, embora apresente como desvantagens, a presença uma voz metálica e aplanada, sem variação prosódica de intensidade e frequência. Além disso, o facto de uma das mãos estar sempre ocupada, inviabiliza a comunicação em determinados contextos (Pinho, 1998; Doyle, 1994).

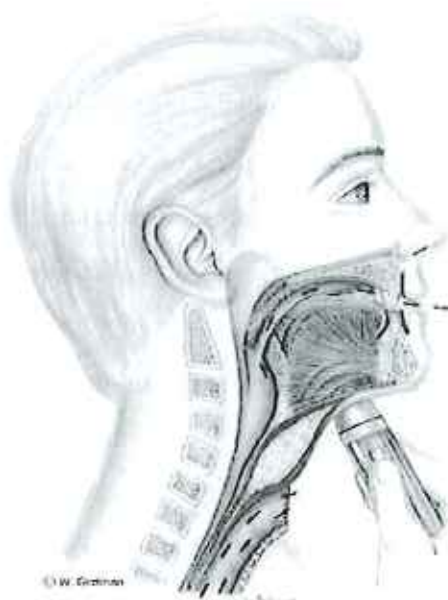


Figura 7 – Esquema representativo da utilização de uma laringe electrónica por utente submetido a laringectomia total. Retirado de www.orl.nl.

3.2.1 Voz esofágica

A **voz esofágica (VE)** resulta da vibração do segmento faringo-esofágico provocada pela passagem de uma corrente de ar vinda do estômago/esófago. Em 1968, Diedrich usou pela primeira vez o termo faringo-esofágico para descrever uma região anatómica usada como fonte geradora de som, para utentes submetidos a laringectomia total (Doyle, 1994).

A reconstrução necessária para o reposicionamento anterior da traquéia, criando o traqueostoma e as suturas dos tecidos musculares que fazem parte desta região, resultam em alterações anatómicas e de funcionamento evidentes. A remoção cirúrgica da laringe requer a reconstrução de tecidos remanescentes da faringe e da parte superior do esófago de forma muito individualizada. A literatura clínica parece dar suporte ao facto dos indivíduos submetidos a laringectomia total, não apresentarem a mesma integridade estrutural e funcional apesar de, aparentemente, a cirurgia ser a mesma. Este dado é bastante importante do ponto de vista clínico, uma vez que está relacionado com a reabilitação terapêutica e com a aquisição da VE (Doyle, 1994).

A aquisição da VF está dependente do funcionamento “quase normal” do esfíncter esofágico superior e da capacidade individual de cada paciente para que utilize uma estrutura anatómica que não é passível de ser controlada voluntariamente e que varia bastante de pessoa para pessoa (Doyle, 1994; Casper & Colton, 1998).

Tendo em conta esta variabilidade individual e a capacidade que cada paciente apresenta, existem diferentes métodos utilizados para a aquisição de VE.

Método de Deglutição de Ar

É uma manobra de deglutição do ar, seguida da sua expulsão, com tentativa de sonorização que se inicia com a emissão de uma vogal aberta como: /a/, /E/ e /O/. É útil, pois estabelece uma associação com uma manobra que pode fazer parte do quotidiano do doente - a cructação.

Este método tem algumas desvantagens das quais se destacam, a presença de ruído, devido à introdução muito rápida ou de muito ar e, também resultantes de impedimentos de base anatómica (estenose ou hipertonia do esófago).

Método de Inalação ou Aspiração de Ar

É uma manobra de introdução de ar nas cavidades superiores do tracto digestivo, criando uma diferença de pressão e uma força de sucção que leva o ar em direcção ao esófago. Para a realização deste método o indivíduo deve deixar a mandíbula e língua baixas e “puxar” o ar pela boca, devolvendo o ar, emitindo uma vogal. Este método tem como principais problemas o ruído do traqueostoma e o facto de, ao devolver o ar sonorizado pela boca, o laringectomizado tender a expirar com vigor.

Método de Injecção de Ar ou Método Holandês

É uma manobra que se traduz pela uma pressão da base da língua contra a parede posterior da faringe, realizada com pressão na emissão de determinados fonemas. De entre os 3 métodos, o de Injecção de Ar parece ser o que revela melhores resultados, mas não há dados na literatura que permitam confirmar este facto. O ensino deste método inicia-se com a produção dos sons /p/, /t/ e /k/, promovendo o aumento da pressão intraoral. Numa fase posterior, aumentam-se as produções, acrescentando

vogais em forma silábica, depois, passa-se para a produção de palavras monossilábicas com os sons treinados de forma isolada e sequencial, injectando ar entre cada uma, posteriormente, procede-se à utilização de palavras dissilábicas e polissilábicas, expressões coloquiais, contagens até à utilização da voz em todos os contextos linguísticos.

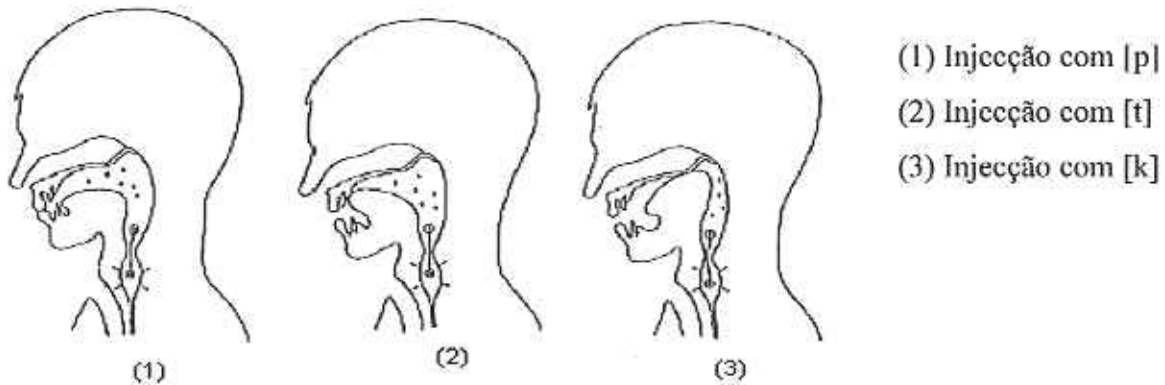


Figura 8 – Esquema representativo da utilização do Método de Injecção de Ar ou Método Holandês. Fonte: Iglesia et al (2006). Nesta imagem observa-se que, à medida de que se altera a consoante a produzir a língua altera a sua posição. Em todas elas se observa a vibração do segmento faringo-esofágico que serve como fonte de vibração para a produção de som.

Os vários métodos, atrás referenciados, podem ser usados de forma complementar, dependendo da fase do tratamento e da capacidade individual de cada utente, potencial utilizador deste tipo de voz (Doyle, 1994).

Na figura 8, observa-se um esquema representativo da produção de VE num utente submetido a laringectomia total.

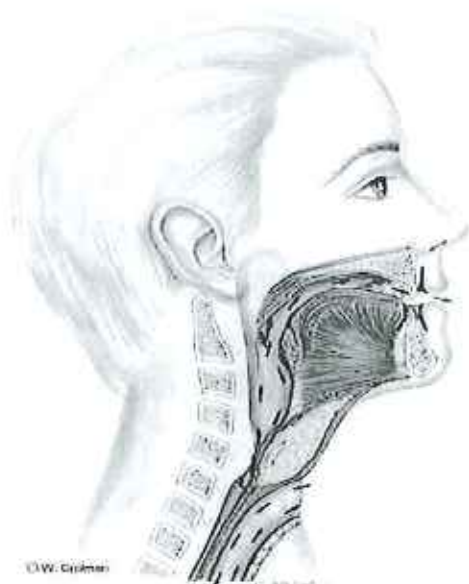


Figura 9 - Esquema representativo da produção de Voz Esofágica (em geral) em utente submetido a laringectomia total. Retirado de www.orl.nl. Nesta imagem observa-se a diferenciação entre a via respiratória e a via oral.

3.2.1.1 Características da Voz Esofágica

A voz esofágica tem como características subjectivas uma qualidade rouca e bastante grave e uma entoação monótona, embora seja perfeitamente perceptível por qualquer ouvinte, desde que bem treinada. Tem a vantagem de ser uma adaptação natural do próprio organismo, não necessitando de qualquer dispositivo externo, estando as duas mãos livres durante a produção do discurso. Contudo, em locais com maior ruído de fundo, este tipo de voz é de difícil percepção para os ouvintes (Casper & Colton, 1998; Pinho, 1998; Behlau, 2005).

Por outro lado, a voz esofágica exige uma aprendizagem média de seis meses e não há forma de garantir o sucesso deste processo. O indivíduo necessita de algum tempo para conseguir as primeiras emissões em voz esofágica e, para que o discurso seja inteligível, este deve ter uma boa capacidade articulatória. Os pacientes têm dificuldade em ser ouvidos quando há ruído de fundo e, por vezes, este tipo de voz é considerado pelos ouvintes como menos agradável, em comparação com outros métodos, embora esta preferência seja alterada quando comparado com a laringe electrónica (Casper & Colton, 1998; Close et al.1998).

O som resultante da voz esofágica, é perceptivamente semelhante ao som produzido pelas cordas vocais embora, normalmente, com uma frequência fundamental

mais baixa. Segundo Sniddecor (1975) e Martin (1979), o valor de frequência fundamental para a voz esofágica é aproximadamente 64 Hz (Behlau, 2004). Mara Behlau refere um valor de 60 Hz para indivíduos do sexo masculino. Os indivíduos mais competentes na utilização da voz esofágica, tendem a produzir frequências fundamentais superiores aos que revelam maior dificuldade na utilização deste tipo de voz (Behlau, 2004).

Slavin, em 1995, realizou um estudo com 26 utilizadores de voz esofágica com boa inteligibilidade e constatou que, muitos dos utilizadores tiveram dificuldade em controlar a frequência fundamental, durante o discurso dinâmico. Como consequência apresentaram uma grande variabilidade em relação aos utilizadores de voz normal. Alguns autores admitem que o valor da frequência fundamental resultante da voz esofágica depende da exacta localização do segmento vibrante, embora existam poucas evidências para suportar esta hipótese (Casper & Colton, 1998).

No que concerne à intensidade, a voz esofágica é cerca de 5 a 8 dB inferior à voz laríngea. A capacidade de extensão da intensidade apresenta um potencial de variação máximo de 20dB, enquanto na voz laríngea se verifica um valor de aproximadamente 45 dB (Close et al, 1998).

Relativamente ao valor de *jitter* (perturbação da frequência), a voz esofágica é mais instável do que a voz laríngea normal (Close et al, 1998). A tendência dos utilizadores de voz esofágica oscilar acima e abaixo de uma determinada frequência, é mais ou menos a mesma do que os utilizadores de voz normal. No entanto, o grau de oscilação difere, sendo maior nos utilizadores de voz normal. Outro indicador de estabilidade da fonte sonora, o *shimmer*, que mede a perturbação da amplitude de vibração, é mais elevado na voz esofágica do que na voz normal (Close et al, 1998).

Foram realizados uma série de estudos (Sisty and Weinberg, 1972; Kytta, 1964; Robbins, Fisher, Bloom e Singer, 1984 citados por Cervera et al, 2001) com medidas de duração relacionadas com as diferentes formas de utilização da voz pós laringectomia total, incluindo palavras por minuto, sílabas por segundo, duração da leitura e tempo máximo de fonação durante a produção de uma vogal.

A duração total da vogal ou tempo máximo de fonação de uma vogal, reflecte a capacidade do falante para controlar o volume de saída de ar. No caso da voz esofágica, reflecte também a competência de, rapidamente, renovar o esófago com a quantidade de ar suficiente. O utilizador de voz esofágica está em desvantagem relativamente aos utilizadores de outros tipos de voz, devido às pequenas quantidades de ar presentes no

esófago, enquanto que os falantes com VTE têm a vantagem de possuírem o suporte de ar pulmonar no seu máximo. De uma forma geral, os utilizadores de voz esofágica têm mais lentamente do que os utilizadores de voz normal, e os anormais períodos de silêncio que apresentam reflectem a maior necessidade que têm de repor o ar. Os utilizadores mais competentes de voz esofágica têm períodos de silêncio muito mais curtos, o que indica que são mais rápidos a ingerir ar e, por isso, apresentam um discurso mais fluente. A fonação sustentada dos utilizadores de voz esofágica é mais curta (normalmente inferior 6 segundos) do que os utilizadores de voz normal, (15-20 segundos), o que comprova que os utilizadores de voz esofágica têm pouco volume de ar (Casper & Colton, 1998).

3.2.2 Voz Traqueo - Esofágica

Mais de vinte anos após a colocação da primeira prótese fonatória, em 1980, por Blom e Singer, a VTE tornou-se o método de reabilitação vocal pós-laringectomia total mais utilizado no mundo Ocidental (Corinna, 2001).

Esta técnica envolve a realização de uma pequena fistula cirúrgica entre a traqueia e a neofaringe. Este procedimento pode ser realizado simultaneamente à laringectomia total, ou alguns meses após a sua realização.

Tal como na VE, a fonte sonora é situada ao nível do segmento faringo-esofágico. Contudo, na VE, tal como na voz laríngea (VL), o fluxo de ar pulmonar é utilizado para a produção da voz. Quando o traqueostoma é encerrado, o fluxo de ar expirado é desviado através da prótese para o esófago, onde a neoglote inicia a sua vibração (Corinna, 2001).

A prótese é composta por uma válvula unidireccional que pode ser colocada durante a laringectomia (fistuloplastia fonatória primária) ou numa cirurgia posterior (fistuloplastia fonatória secundária), onde se realiza uma fistula traqueoesofágica. Esta fistula consiste numa ligação entre a traqueia e o esófago (Casper & Colton, 1998; Behlau, 2004).

A criação da fistula permite que o ar pulmonar produza uma vibração do segmento faringoesofágico de forma a criar som. O som dirige-se então para a faringe, depois para as cavidades oral e nasal onde é articulado e amplificado de forma a produzir fala.

As complicações que podem surgir após a colocação de próteses vocais incluem, entre outros, formação de granuloma, estenose do traqueostoma e/ou da fístula, dificuldades de deglutição, fistulas faringocutâneas e esofagotraqueais, necrose dos tecidos circundantes do traqueostoma, hemorragias, infecções. Contudo, salienta-se que estas lesões são pouco frequentes e têm implicações reduzidas para o doente (Behlau, 2004; Dikeman, 2003). Pelas suas características, a aquisição da VTE é fácil e rápida (Behlau, 2004).

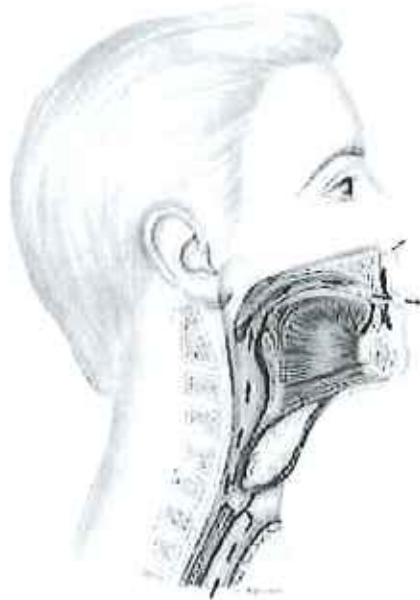


Figura 10 - Esquema representativo da utilização de voz traqueo-esofágica em utente submetido a laringectomia total. Retirada de www.orl.nl.

3.2.2.1 Características da Voz Traqueo-Esofágica (VTE)

A fonte de energia que suporta a VTE é a fonte de ar pulmonar, o que torna este método mais próximo do padrão vocal laríngeo. Esta característica possibilita um maior tempo de emissão, melhor inteligibilidade e maior encadeamento da fala, comparativamente com os outros métodos referidos (Jorge et al, 2004, Behlau, 2001).

Os utilizadores de VTE tendem a produzir valores de F0 muito próximos dos indivíduos com voz laríngeo. A variabilidade de F0 parece ser mais pequena do que nos falantes com voz esofágica (Corinna, 2001).

A intensidade da VTE parece ser ligeiramente mais baixa do que a produzida pelos utilizadores de voz laríngea. A variação da intensidade pode, inclusive, ser maior do que nos falantes normais (Corinna, 2001; Behlau, 2001).

Os registos das características de *jitter* dos falantes com VTE não são muito claros. Um estudo realizado em 1984, por Robbins et al, com 35 indivíduos, 15 utilizadores de VTE, 15 utilizadores de voz esofágica e 15 falantes normais, revelou valores muito semelhantes entre os utilizadores de VTE e os normofalantes, enquanto o outro revelou valores superiores ao normal. Seria de esperar valores de *jitter* similares entre os utilizadores de VTE e de voz esofágica, uma vez que utilizam o mesmo sistema anatómico de vibração, o segmento faringo-esofágico. Outro indicador de estabilidade da fonte sonora, o *shimmer*, que mede a perturbação da amplitude de vibração, é maior na VTE do que na voz normal. (Casper & Colton, 1998)

3.2.3 Comparações entre os dois métodos

A taxa de sucesso de aquisição da VTE é superior à da VE, com valores na ordem dos 90% (Corina, 2001). Alguns estudos comparativos das duas vozes indicam que a VTE é superior à VE em inúmeros aspectos. Pindzola e Cain (1988), citados por Corina (2001) referem que, no estudo perceptivo de avaliação da fluência, *pitch*/qualidade, velocidade de fala, inflexão e aceitação, a VTE obteve melhores resultados. Outro estudo perceptivo revelou que a qualidade da VTE é superior à da VE no que diz respeito à velocidade da fala e à fluência. (Corina, 2001)

Nos diferentes estudos acústicos realizados por diversos autores relativamente aos dois tipos de voz, parece haver ser unanimidade quanto aos resultados obtidos. Todos eles consideram que, para os diferentes parâmetros analisados, a VTE apresenta valores mais semelhantes à VI do que a VE.

Estudos mais recentes, onde foram utilizados parâmetros acústicos como medidas de duração relacionadas com as diferentes formas de utilização da voz pós laringectomia total, incluindo palavras por minuto, sílabas por segundo, duração da leitura e tempo máximo de fonação durante a produção de uma vogal, também obtiveram os resultados já focados anteriormente (Cervera et al, 2001).

A duração total da vogal ou tempo máximo de fonação de uma vogal, reflecte a capacidade do falante para controlar o volume de saída de ar. No caso da voz esofágica, reflecte também a competência de rapidamente renovar o esófago com a quantidade de

ar suficiente. Como já foi referido, o utilizador de voz esofágica está em desvantagem relativamente aos utilizadores de outros tipos de voz, devido às pequenas quantidades de ar presentes no esófago, enquanto os falantes com VTE têm o suporte de ar pulmonar no seu máximo (Corinna, 2001).

De uma forma geral, os utilizadores de voz esofágica têm mais lentamente do que os utilizadores de voz normal, e os anormais períodos de silêncio que apresentam, reflectem a maior necessidade que têm em repor o ar. Os utilizadores mais competentes de voz esofágica têm períodos de silêncio muito mais curtos, o que indica que são mais rápidos a ingerir ar e, por isso, apresentam um discurso mais fluente. A fonação sustentada dos utilizadores de voz esofágica é mais curta (normalmente inferior 6 segundos) do que os utilizadores de voz normal, (15-20 segundos), o que comprova que os utilizadores de voz esofágica têm pouco volume de ar (Casper & Colton, 1998).

Os falantes com VTE têm com uma velocidade inferior relativamente aos falantes normais, mas apresentam valores superiores aos utilizadores de voz esofágica (VF). Esta diferença entre os utilizadores de VTE, de VE e de voz laríngea (VL) reflecte a dificuldade em realizarem um bom controlo do segmento faringo-esofágico e a necessidade de articularem com grande precisão. Estas diferenças foram posteriormente confirmadas por Christensen, and Kempster (1986), citados por Cervera, Miralles e Alvarez (2001) que demonstraram que a duração das vogais dos utilizadores de VTE é superior à VL. Estes autores sugerem que, apesar das alterações na duração das vogais, a inteligibilidade mantém-se elevada, sendo que o seu valor se situa na ordem dos 90%, segundo um estudo perceptivo anterior realizado por Miralles e Cervera (1995), citado por Cervera, Miralles e Alvarez (2001).

Segundo Cervera et al (2001), na voz esofágica, a duração das palavras é significativamente mais curta, atribuindo este facto à redução da segunda sílaba das palavras, o que consideram ser uma consequência da limitação existente a nível do reservatório de ar que os utilizadores de voz esofágica usam para a fonação. Desta forma, verifica-se uma menor duração da segunda sílaba relativamente à primeira, aumentando a duração da primeira vogal. Refere ainda que, o aumento da duração da vogal poderá estar relacionado com uma taxa de vibração mais lenta do segmento faringo-esofágico, em comparação com as vibrações laríngeas da voz normal. No entanto, a formulação deste tipo de hipóteses deve ser feita com prudência, enquanto não se aprofundarem os conhecimentos sobre as propriedades mioelásticas do esfíncter esofágico (Cervera et al, 2001).

Segundo Christensen e Weinberg (1976), na VE e na VTE, a frequência das formantes é superior à da voz laríngea normal, provavelmente devido ao encurtamento do tracto vocal. Os utilizadores destes dois tipos de voz podem tentar compensar este encurtamento do tracto vocal através da protrusão labial ou alterando a posição da língua (Casper & Colton, 1998). Segundo Cervera et al, a estrutura das formantes pode variar de acordo com o contexto de consoantes em que a vogal se encontra. Os autores consideram ainda que, o factor mais importante para determinar a posição das frequências das formantes é o comprimento do tracto vocal, pelo que tractos vocais mais curtos apresentam frequências de formantes mais altas, tal como foi verificado por Peterson & Barney (1952) onde os valores das formantes foram avaliados em homens, mulheres e crianças. Também referem que, ao relacionar os parâmetros acústicos F1 e F2 com as dimensões fonéticas alto/baixo e anterior/posterior, os utilizadores de VE e de VTE articulam as vogais anteriormente, com posição alta de língua, relativamente à voz normal (Cervera et al, 2001).

Capítulo 4 - Corpus

4.1 Introdução

Este capítulo tem como principal objectivo descrever a criação do Corpus escolhido e a respectiva gravação, reflectindo sobre o material utilizado e algumas das dificuldades encontradas ao longo de todo o processo. Pretende, ainda, caracterizar os informantes que participaram neste estudo, descrever as anotações realizadas e descrever a extracção dos parâmetros para posterior análise.

4.2 Definição do Corpus

Procedeu-se à criação do *corpus*, tendo como base os mesmos critérios que os utilizados no estudo de Cervera et, al publicado em 2001, no qual foram comparadas as características acústicas de dois grupos de utentes submetidos a laringectomia total, com um grupo de falantes normais. No estudo realizado para os falantes da Língua Espanhola, as vogais foram produzidas num contexto CVCV (C – consoante, V- vogal) pelos dois grupos, um constituído por utentes utilizadores de voz esofágica e outro por utilizadores de voz traquico-esofágica, tendo sido gravadas 24 palavras posteriormente analisadas pelo CSI. (Kay Elemetrics).

No nosso estudo, relativo ao Português Europeu, foi construído um *corpus* constituído por 34 palavras dissilábicas em contexto CVCV, contendo as 9 vogais orais do Português-Europeu na primeira sílaba. O *corpus* foi dividido em duas partes, uma principal, apresentada na tabela 2 e uma alternativa, que não será utilizada neste estudo.

As vogais [6] e [ə] foram excluídas do *corpus*, pela inexistência de palavras dissilábicas acentuadas na primeira sílaba contendo estas vogais (átonas). Pretendia-se que as palavras pertencentes ao *corpus* fossem de uso comum, para facilitar a produção e gravação, no entanto, tal nem sempre foi possível e optou-se pela sua inclusão na lista final de palavras.

Os contextos consonânticos variaram entre Oclusiva-vogal-Oclusiva-Vogal e Fricativa-Vogal-Oclusiva-Vogal. Evitou-se a utilização de consoantes nasais, vibrantes e laterais, uma vez que as mesmas possuem formantes, o que iria dificultar a tarefa de

anotação na delimitação da vogal. Como esta situação poderia alterar os resultados, tomou-se esta medida para permitir uma maior fiabilidade da análise.

Foi apresentada uma lista inicial de palavras, posteriormente revista e adaptada. Este *corpus* foi utilizado em dois artigos realizados para a disciplina de Projecto de Investigação no ano lectivo 2005/2006, por alunas do 4º ano de Terapia da Fala da Escola Superior de Saúde da Universidade de Aveiro.

Os trabalhos realizados foram: “Análise Acústica das Vogais Oraís do Português-Europeu na Voz Esofágica” elaborado por Ana Sousa e Bárbara Macedo e “Análise das Características das Vogais em Pacientes Submetidos a Laringectomia Total, Utilizadores de Voz Traqueoesofágica” elaborado por Susana Dias e Alberta Machado.

O primeiro trabalho referido teve como objectivos “... caracterizar os parâmetros acústicos da voz esofágica e analisar os parâmetros duração, frequência fundamental (F0) e frequência das duas primeiras formantes (F1 e F2), das vogais orais do Português-Europeu da voz esofágica, e comparar os resultados obtidos na amostra de falantes do Português-Europeu com os resultados apresentados no estudo realizado por Cervera et al para falantes da língua espanhola utilizadores do mesmo tipo de voz” (Sousa & Macedo, 2006).

O segundo trabalho referido teve como objectivos “... contribuir para uma melhor caracterização da voz traqueoesofágica, através da realização de análises objectivas e realização da análise acústica das amostras de voz, relativa aos parâmetros F0, F1, F2 e duração..” (Dias & Machado, 2006).

[a]	[i]	[u]	[ɛ]	[o]	[o]	[e]
capa	pipa	cuco	peca	copo	coco	peta
[kap6]	[pip6]	[kuku]	[pEk6]	[KOpu]	[Koku]	[pet6]
pato	pico	puto	pepe	toque	topo	pêca
[patu]	[piku]	[putu]	[pEp@]	[tOk@]	[topu]	[pck6]
taco	tipo	tuta	teto	tota	colo	
[taku]	[tipu]	[tut6]	[tɛtu]	[tOt6]	[kotu]	-
sapo	lita	suco	seca	foca	soco	seco
[sapu]	[fit6]	[suku]	[sEk6]	[fOk6]	[soku]	[scku]
faça	fica	chuto	seta	soca	sopa	cete
[fak6]	[fik6]	[Sutu]	[sɛt6]	[sOk6]	[sop6]	[set@]

Tabela 2 – Corpus utilizado no estudo (as primeiras três linhas da tabela constituem a parte principal do *corpus*, as restantes a parte secundária)

4.3 Gravação

A gravação decorreu, individualmente, numa sala onde não foi possível assegurar as condições ideais de insonorização.

Foi ainda realizado um pequeno questionário, com o objectivo de obter os dados pessoais dos informantes e outras informações relativas à cirurgia e ao processo de reabilitação dos mesmos, cujos dados serão apresentados, posteriormente, neste capítulo.

O processo de gravação iniciou-se com uma breve explicação do procedimento, nomeadamente, o facto de algumas palavras conterem acentuação, para facilitar o processo de leitura e garantir a produção correcta dos fonemas. As palavras foram ordenadas aleatoriamente e exibidas 3 vezes cada uma. Foi apresentada uma palavra a cada 3 segundos. No final das 34 que compõem o *corpus*, efectuou-se uma pausa até o referente afirmar que estava disponível para avançar na gravação. Para tal, utilizou-se uma frequência de amostragem de 20kHz.

Optou-se por não utilizar a mudança automática das palavras, para adaptar a velocidade de transição ao ritmo de cada um dos informantes.

No decorrer do processo de gravação, ocorreram alguns problemas dos quais se destacam: a produção incorrecta de algumas palavras do *corpus*, principalmente as que não são de uso comum, a presença de algum ruído de fundo durante a gravação e, ainda, o facto de alguns dos informantes terem manifestado a necessidade de tossir com alguma frequência, contribuindo também para a existência de ruído.

Para o processo de gravação foi utilizado um computador portátil correndo o programa ProRec, uma placa de som CSL 4400 da Kay Elemetrics Corp e recorreu-se à utilização de um microfone com suporte de cabeça modelo C 420 da AKG, colocado a cerca de 3 cm da boca dos informantes. Este microfone é do tipo condensador e omnidireccional no que respeita à direccionalidade. Após a gravação foram criados ficheiros individuais correspondentes a cada palavra integrante do *corpus*, usando a informação guardada durante a gravação pelo programa ProRec.

4.3.1 Informantes

Foram seleccionados 9 informantes, todos pertencentes à Associação Portuguesa dos Limitados da Voz (APLV), 6 dos quais utilizadores funcionais de VE e os restantes utilizadores de VTF, previamente avaliados e orientados por Terapeutas da Fala que os

acompanharam durante o seu processo de intervenção. Não se procedeu à gravação de um grupo de controlo utilizador de voz laríngea normal, uma vez que se recorreu a um estudo recente contendo as informações necessárias relativas aos dados normativos de informantes do mesmo sexo, da mesma faixa etária e localização geográfica.

Os informantes eram todos do sexo masculino, com idades compreendidas entre os 55 e os 71 anos, com uma média de idades de 64,2 anos, submetidos a Laringectomia Total. Destes elementos, sete tiveram acompanhamento em Terapia da Fala e um dos referentes efectuou a cirurgia fora do país. Todos foram previamente informados dos objectivos do estudo em questão e aceitaram participar voluntariamente.

Nome	Idade	Data Intervenção	Radioterapia (pós-cirurgia)	Terapia da Fala	Opção de Comunicação
J.M.	55	15/11/2000	Sim	Sim	Voz Esofágica
J.L.	57	22/04/2001	Sim	Sim	Voz Esofágica
J. P.	55	19/08/2001	Sim	Não	Voz Esofágica
A.M.	71	24/01/2000	Não	Sim	Voz Esofágica
M. R.	70	17/06/1991	Sim	Não	Voz Esofágica
L.M.	71	17/03/1999	Sim	Sim	Voz Esofágica
F.F.	64	24/04/2000	Sim	Não	Voz Traqueo-esofágica
D. R.	64	01/05/2002	Sim	Não	Voz Traqueo-esofágica
J. F.	71	28/05/2005	Não	Sim	Voz Traqueo-esofágica

Tabela 3 – Caracterização dos Informantes

4.4 Anotação

Para a realização da anotação das palavras e dos fonos do *Corpus* previamente elaborado, foi utilizado o programa *Speech Filling System (SFS)*, desenvolvido em 2001 por Mark Huckvale. Esta anotação teve como base a observação do sinal acústico, dos espectogramas *narrow band e wide band*, da frequência fundamental e das formantes das palavras previamente segmentadas. A anotação foi realizada manualmente marcando-se, numa primeira fase, o início e o final da palavra. De seguida, foi realizada a anotação manual dos fonos, identificando os segmentos consonânticos que antecedem

e seguem a vogal da primeira sílaba, ficando esta delimitada. Na anotação foi utilizado o Alfabeto SAMPA para o Português.

Após a recolha das gravações, foi decidido que apenas se faria a anotação de três informantes de cada um dos grupos. A escolha dos informantes utilizadores de VE foi aleatória. Apresentam-se nas figuras 11 e 12 dois exemplos de arquivos anotados, para cada uma das vozes.

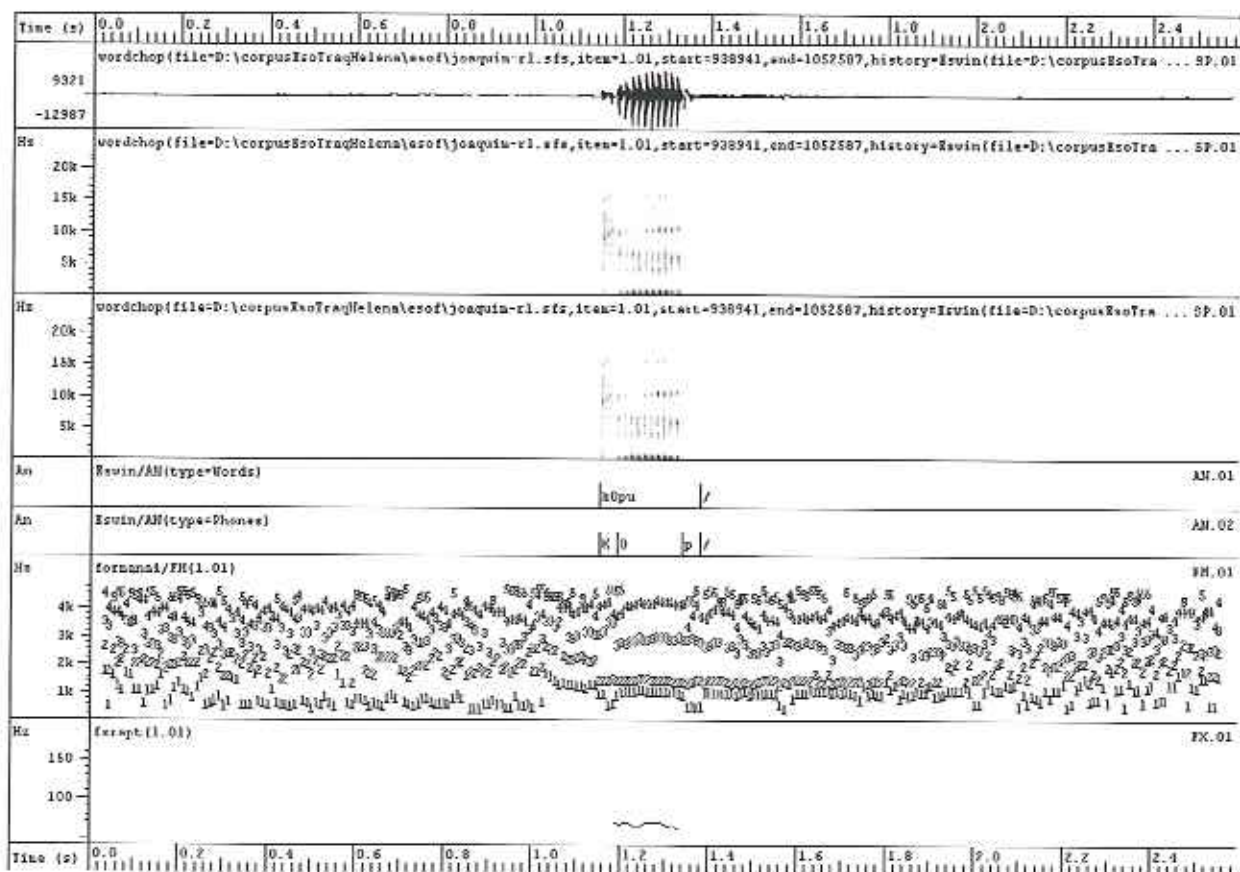


Figura 11 – Exemplo de anotação da palavra /kOpu/ para um Informante utilizador do VE. De cima para baixo podem ser vistos o sinal de voz, os dois espectrogramas(*wide band* e *narrow band*), a anotação da palavra e a anotação dos fones, as formantes e F0.

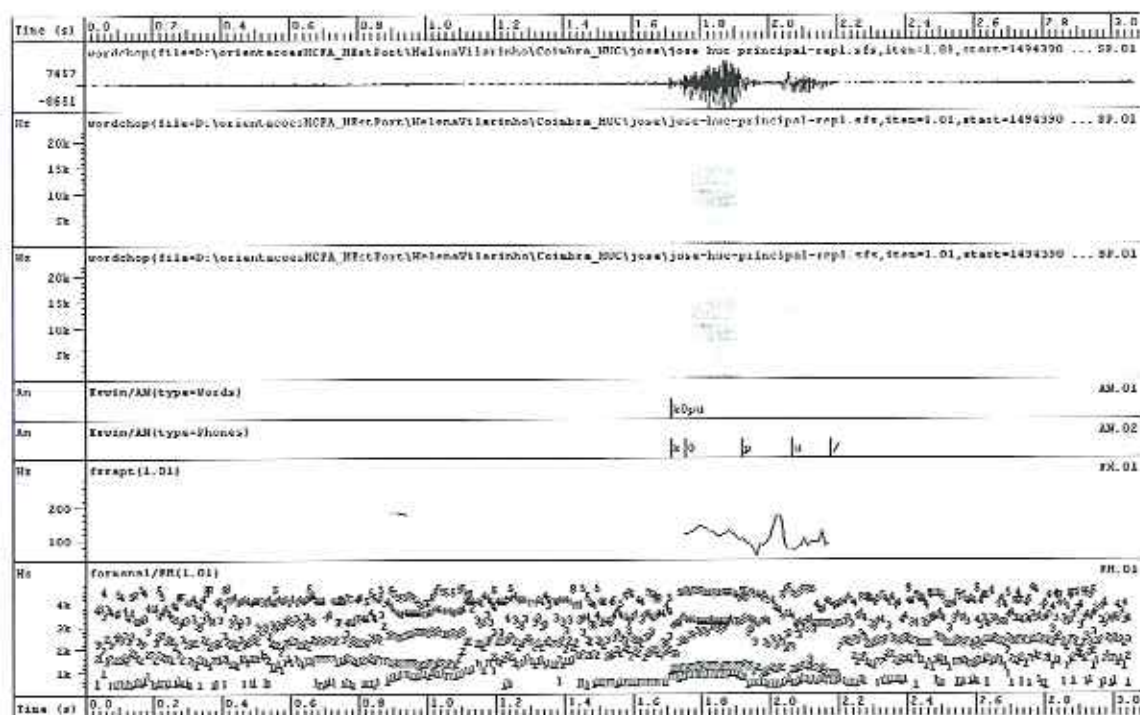


Figura 12 – Exemplo de anotação da palavra /kOpu/ para um informante utilizador de VTE. De cima para baixo podem ser vistos o sinal de voz, os dois espectrogramas (wide band e narrow band), a anotação da palavra e a anotação dos fonemas, as formantes e F0.

Na anotação das palavras do corpus foram sentidas algumas dificuldades, relacionadas sobretudo com o início da vogal da primeira sílaba. A literatura consultada é dúbia relativamente a este parâmetro, pelo que se optou pelo início da anotação da vogal, no primeiro pico negativo do sinal. Fazendo uma análise das características do sinal recolhido, escolheu-se o primeiro pico inferior logo que o sinal começa a apresentar uma maior regularidade (o que nos indica a presença de um sinal periódico característico da vogal). Contudo, para uma maior fiabilidade, utilizou-se também uma observação exaustiva dos espectrogramas, onde é possível a observação mais característica dos traços dos formantes.

4.5 Extração de parâmetros

Tal como já foi referido anteriormente, as anotações foram realizadas no *SI/S*, por haver uma maior familiaridade com este programa. Posteriormente, foi criado um programa para converter os ficheiros analisados de forma a poderem ser lidos pelo *Praat* (programa de análise acústica criada por Paul Boersma e David Weenink em 1992), que apresenta alguns comandos que permitem uma leitura mais rápida e

funcional dos parâmetros que pretendíamos analisar. No programa *Praat*, reuniram-se os ficheiros e procedeu-se à extracção do segmento anotado como vogal tendo, a partir daí, sido calculado, usando as facilidades do *Voice Report*, a duração da vogal, F_0 médio, medidas de *jitter*, *shimmer*, *Harmonic-to-Noise-Ratio* e F1 e F2.

Capítulo 5 - Análise Acústica

Uma vez que em Portugal não há nenhum estudo relativo ao tema abordado neste trabalho, usaram-se alguns estudos publicados em Espanha, Inglaterra e Holanda nomeadamente o de Cervera et al (2001), realizado para a Língua Espanhola, e que tinha como objectivos: (a) descrever as características acústicas das vogais da língua Espanhola em utentes submetidos a laringectomia total; (b) comparar os resultados com os obtidos num grupo de controlo de indivíduos normofalantes. Os resultados foram comparados com um estudo realizado para a língua inglesa (Sisty e Weinberg, 1972 citados por Cervera et al, 2001).

Os resultados obtidos para os falantes da Língua Espanhola demonstraram que, os valores de duração das vogais, de F1 e F2, diferem significativamente dos valores normais. Neste estudo ficou demonstrado que os utilizadores de VE e VTE apresentam valores de F1 e F2 mais elevados e durações superiores que os falantes normais. Foi ainda realizado um teste perceptual de forma a relacionar as características acústicas das vogais com os dados perceptivos obtidos.

Com este teste ficou provado que as vogais produzidas, de uma forma geral, são bem percebidas em contexto de palavras.

5.1 Resultados

Para um estudo efectivo das vogais é necessária a análise de determinados parâmetros que as caracterizam. Para isso, neste capítulo, irá ser feita a observação e descrição dos dados obtidos nos parâmetros de duração, frequência fundamental (F_0), *jitter*, *shimmer*, *Harmonic-to-Noise Ratio (HNR)*, primeira e segunda formantes (F1 e F2) das vogais colocadas na primeira sílaba das palavras do corpus previamente descrito.

5.1.1 Duração

A duração das vogais é um parâmetro relevante na caracterização da voz dos utentes submetidos a Laringectomia Total, uma vez que reflecte a capacidade do indivíduo em manter a fonação.

Tal como já foi referido anteriormente, alguns estudos realizados revelaram que os indivíduos com VTE e VE apresentavam uma duração média das vogais superior aos falantes normais e que não existiam diferenças significativas entre os dois grupos de indivíduos laringectomizados. Como hipótese para explicar estas diferenças temporais, propôs-se o facto dos indivíduos laringectomizados apresentarem um fraco controlo motor da neoglote durante a fonação (Casper & Colton, 1998). O pouco controlo da resistência oferecida pelo segmento faringo-esofágico poderia ser responsável pela dificuldade em iniciar e terminar o vozeamento das vogais nestes pacientes (Casper & Colton, 1998).

Na Figura 13 observam-se os valores médios e respectivos intervalos de confiança a 95% da duração das vogais obtidas neste estudo, por grupo de informantes.

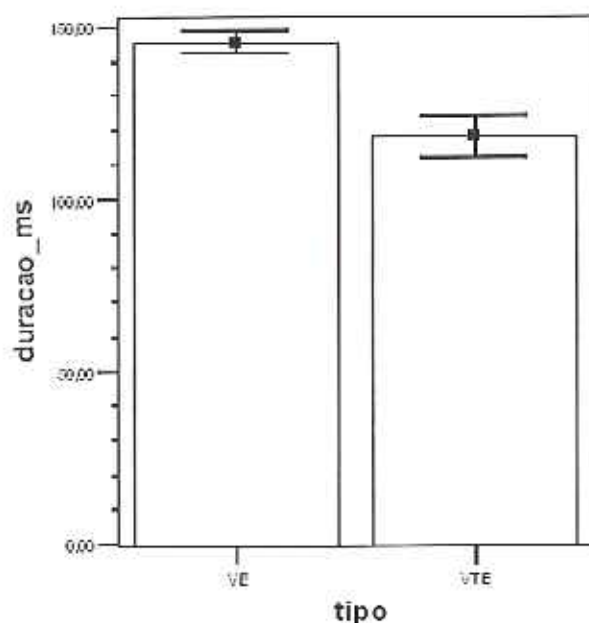


Figura 13 – Durações médias para os dois tipos.

A figura 12 mostra que a média da duração das vogais orais do Português Europeu é superior nos indivíduos utilizadores de VE, cujo valor médio é de cerca de 150ms e para o grupo de VTE é cerca de 120ms. Pode observar-se que a diferença entre os intervalos de confiança obtidos para a VE e a VTE foi significativa (os intervalos de confiança não se sobrepõem).

Comparando estes dados com os dos indivíduos utilizadores de voz laríngea, obtidos num estudo realizado para o Português Europeu (Teixeira e al, 2004), conclui-se que os valores médios de duração da VTE são os que mais se aproximam da normalidade, enquanto os valores obtidos para a VE são muito superiores.

Para que seja possível uma análise mais detalhada destes valores, apresentam-se, na Figura 14, os dados relativos aos valores médios de duração das sete vogais estudadas para o Português Europeu.

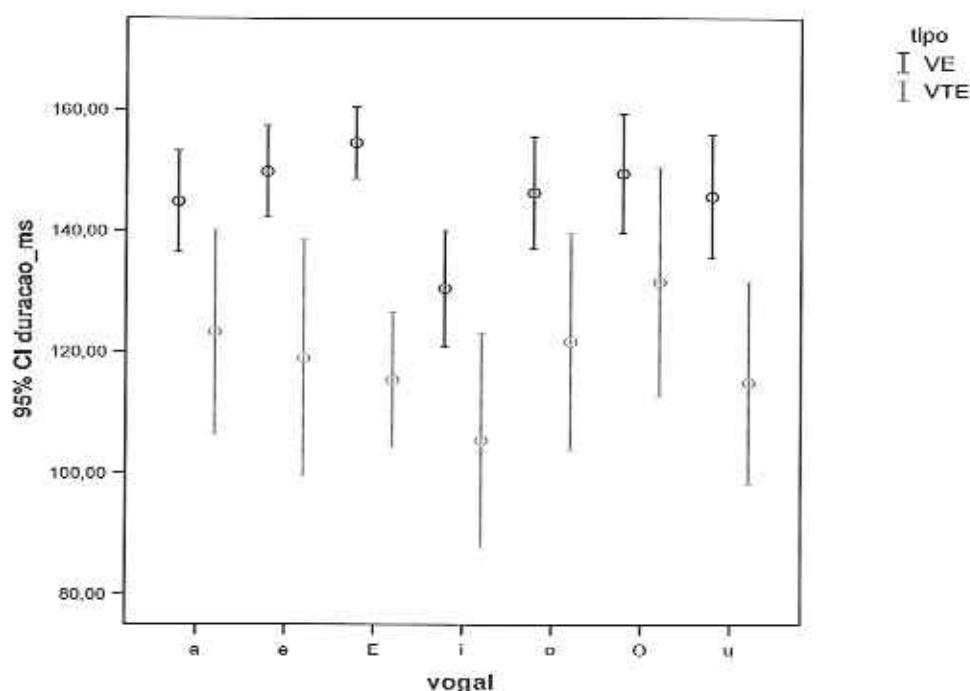


Figura 14 – Duração média de cada uma das vogais para os dois grupos

Analisando a figura 14, podemos verificar que, a duração média de todas as vogais é superior para os indivíduos com VE sendo que, a diferença entre os intervalos de confiança obtidos para a VTE e VE, é significativa (os intervalos de confiança não se sobrepõem) para as vogais [e], [E] e [u] e potencialmente não significativa, (os intervalos de confiança sobrepõem-se) para as vogais [a], [i], [o] e [O].

Para os utilizadores de VE, a vogal com maior duração foi a vogal [E] e a que apresentou menor duração foi a vogal [i]. Para os utilizadores de VTE, a vogal com maior duração foi a vogal [O] e com menor duração a vogal [i]. Dentro deste último grupo, podemos constatar que a variabilidade é maior do que no primeiro grupo. Comparando estes valores com os dados normativos obtidos no estudo realizado para análise das vogais em falantes utilizadores de voz laríngea, verificou-se que a vogal que apresentou menor duração foi [i] e que a de maior duração, a seria vogal [u]. (Teixeira et al, 2004).

De modo a perceber se esta variabilidade individual ocorre neste tipo de falantes, analisa-se na, Figura 15, o valor médio e respectivo intervalo de confiança a 95% de cada uma das vogais estudadas, por informante.

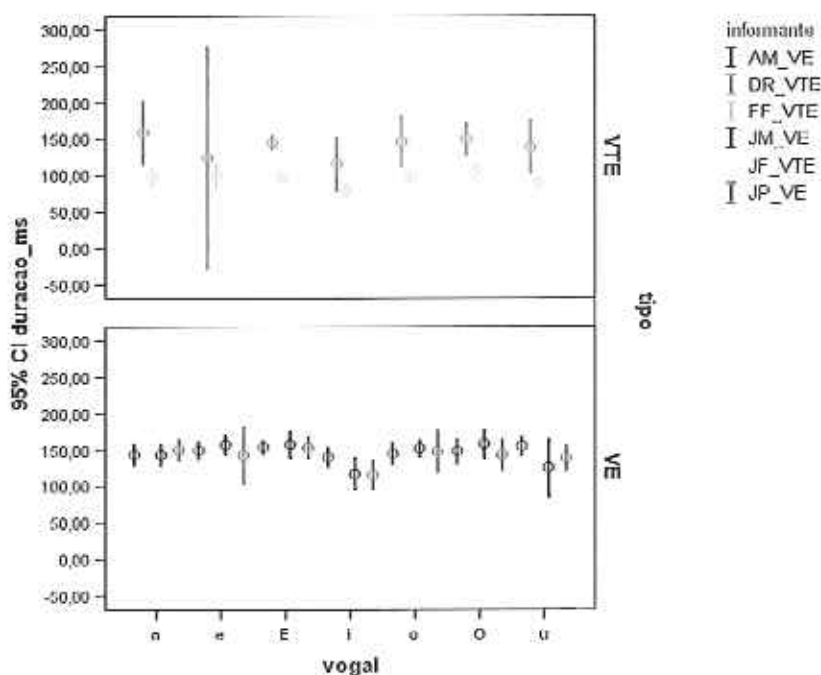


Figura 15 – Comparação das médias de duração das vogais por grupo de utilizador

Analisando os dois gráficos da Figura 15, pode observar-se que a duração média das vogais produzidas pelos utilizadores de VE é menos variável do que a do grupo de VTE, apresentando valores mais homogêneos. No grupo de VTE, o informante J.F. tem valores médios de duração de todas as vogais superiores a qualquer outro elemento dos dois grupos e, o informante F.F., apresenta os valores médios de duração mais baixos (≈ 100 ms). Este valor é o que mais se aproxima dos dados obtidos para a normalidade

no estudo de Teixeira et al (2001). O facto de os valores médios de duração serem mais elevados para o grupo de VE poderá ocorrer, tal como descrito na literatura, pelo facto de haver um fraco controlo da neoglote neste grupo de utilizadores. De referir também que os dois elementos escolhidos (de forma aleatória) para o teste perceptivo, a apresentar no próximo capítulo, J.M. e D.R. apresentam valores médios de duração relativamente semelhantes.

5.1.2 Frequência fundamental

A frequência fundamental (F_0) é uma das medidas acústicas que aparece referenciada na literatura, em diversos estudos, comparando grupos de utilizadores de VE e VTE. Casper & Colton (1998) referem que o valor de F_0 para os utilizadores de VE do sexo masculino é inferior à voz laríngea. Autores como Snidecor & Martin e Behlau, referem, para estes utilizadores, valores na ordem dos 64 Hz e 60 Hz, respectivamente, o que nos indica que é cerca de metade da média encontrada no adulto normofalante que apresenta valores de referência entre 120 e 130 Hz (Casper & Colton, 1998; Jorge et al, 2004; Guimarães, 2007).

Os bons utilizadores de VE tendem a produzir valores mais elevados de F_0 e utilizadores menos capazes produzem valores inferiores. Nestes últimos, parece haver uma dificuldade acrescida em controlar F_0 durante a produção de discurso espontâneo. Consequentemente, os falantes de VE têm uma gama de valores superior aos falantes normais. Alguns autores acreditam que este factor está relacionado com a localização exacta do segmento faringo-esofágico, mas não há grandes evidências que suportem esta teoria (C.J.van As, 2001; Casper & Colton, 1998).

Os utilizadores de VTE tendem a produzir F_0 com valores mais próximos do normal, pelo menos em utilizadores do sexo masculino. A variabilidade de F_0 é menor do que em utilizadores de VE, mas existem algumas diferenças individuais (Casper & Colton, 1998).

Na Figura 16, são apresentados os valores médios e respectivos intervalos de confiança a 95% de F_0 para cada um dos grupos.

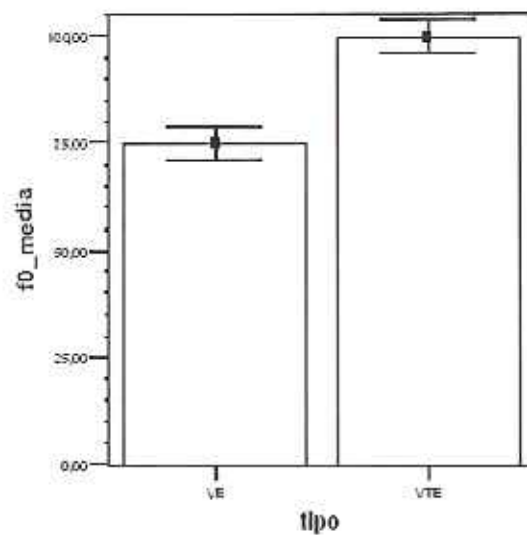


Figura 16 - Valores de F_0 para os dois grupos de Laringectomizados

Analisando a figura anterior, podemos verificar que o valor de F_0 é superior para os indivíduos utilizadores de VTE e que essa diferença é significativa. Assim temos valores médios de F_0 de 100Hz para os utilizadores de VTE e valores de 75Hz para o grupo utilizador de VE. Estes valores estão de acordo com a literatura consultada que referem que, os utentes que utilizam VE têm valores de frequência fundamental inferior ao grupo que utiliza VTE e que estes têm valores mais próximos do valor normal de F_0 . Contudo, o valor de F_0 obtido para a voz esofágica é superior ao obtido nos estudos de Behlau e Snidecor, já referidos anteriormente.

De forma a observar se existe variabilidade dentro dos grupos estudados, apresenta-se, em seguida a análise da Figura 17 onde se comparam os valores médios e respectivos intervalos de confiança a 95% de F_0 , por informante.

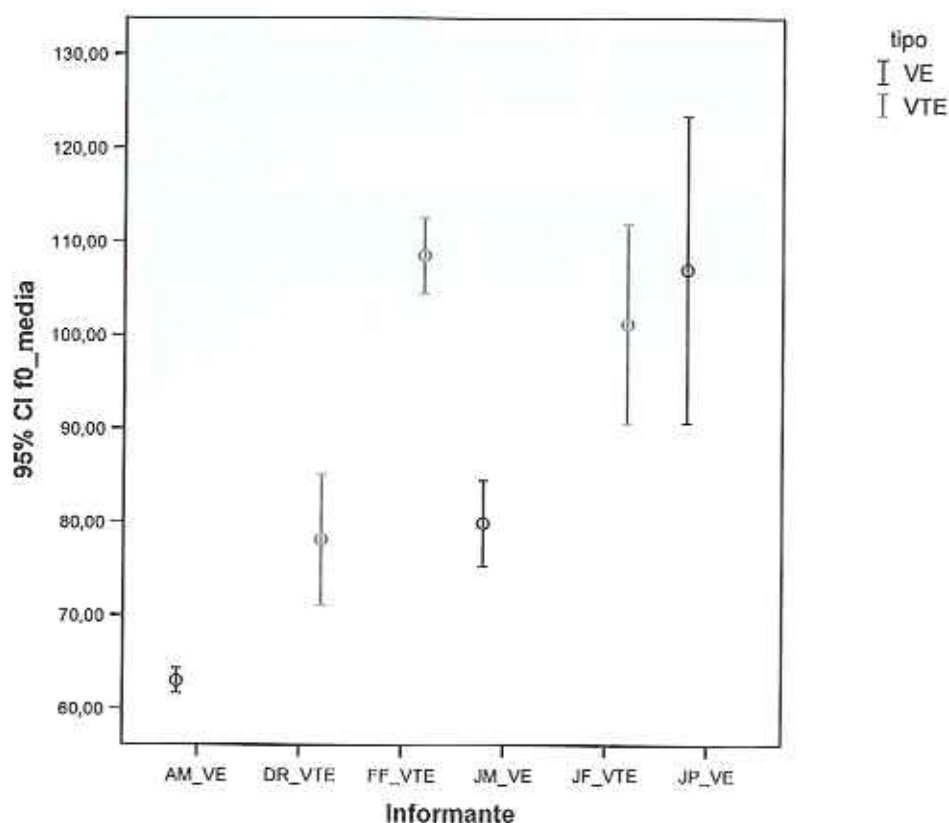


Figura 17 - Valores médios de F_0 por informante

Pode então verificar-se pela observação da figura 17, que em cada um dos grupos há casos especiais. O informante A.M. (do grupo de VE) apresenta um valor aproximado de 60 F_0 Hz que pode ser considerado um valor bastante baixo e supor uma qualidade vocal fraca, podendo este valor influenciar, de forma significativa, os resultados obtidos, dado o reduzido número de informantes estudados. Contudo, dentro do mesmo grupo, o informante J.P. apresenta um valor de F_0 de 100 Hz que é mais próximo dos valores obtidos para o grupo de VTE e para o grupo de utilizadores de voz laríngea. No grupo de VTE, o informante D.R. apresenta um valor de F_0 abaixo dos 80 Hz, que está situado abaixo dos valores encontrados noutros estudos para os grupos de VTE. Aqui podemos também comparar os valores de F_0 dos falantes D.R. e J.M. (seleccionados aleatoriamente para o estudo perceptivo) e verificar que o valor médio obtido neste parâmetro, parece ser bastante próximo nos dois falantes.

Os valores obtidos sugerem que, dentro de cada um dos grupos, há alguma variabilidade que poderá estar relacionada com a capacidade individual de utilização de cada um dos tipos de voz.

Na figura 18, apresentam-se os valores médios e respectivos intervalos de confiança a 95% de F_0 para cada vogal, obtida em cada um dos grupos.

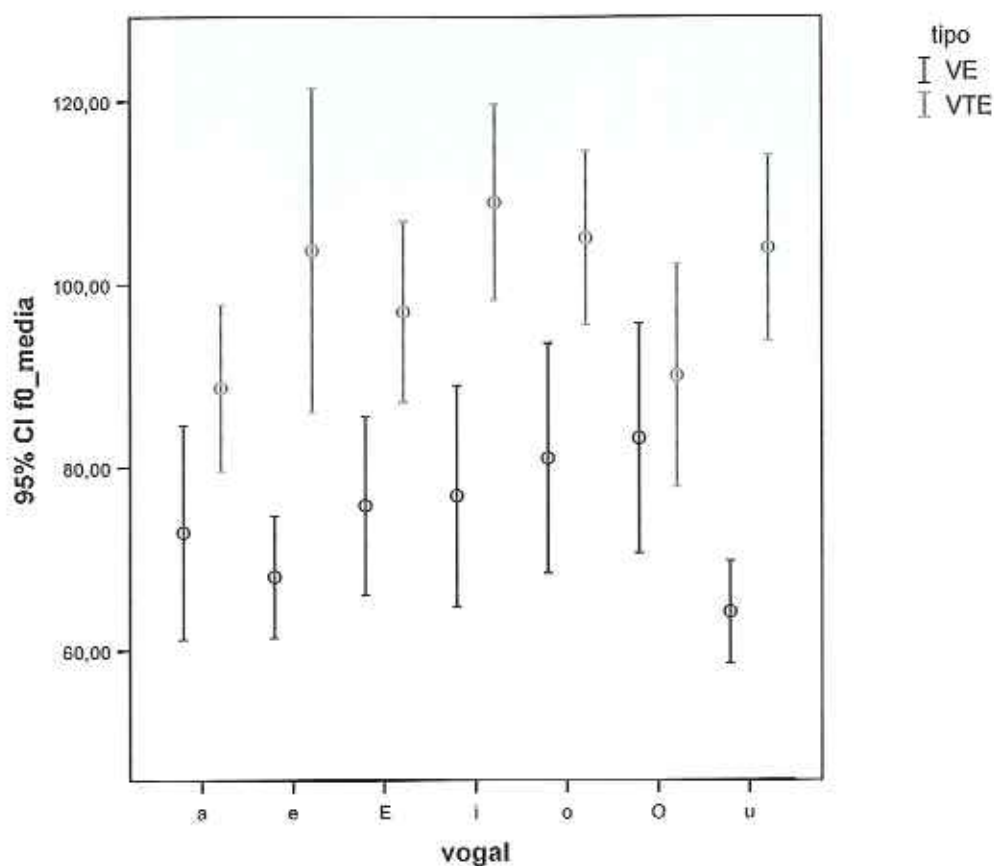


Figura 18 - Valores médios de F_0 para as vogais

Analisando os dados do gráfico anterior, verificamos que os valores de F_0 de todas as vogais são superiores nos utilizadores de VT em relação aos utilizadores de VE.

No gráfico, podemos observar que para as vogais [e], [E], [i], [o] e [u] as diferenças são significativas (os intervalos de confiança não se sobrepõem) e que para as vogais [a] e [O] as diferenças não são significativas, tendo a vogal [O] para os dois grupos um valor de F_0 médio bastante próximo, comparando os grupos de utilizadores de VTE e VE. A vogal com maior valor médio de F_0 para o grupo de VE foi a vogal [O] e a que apresentou menor valor médio de F_0 foi a vogal [u]. No grupo de VTE a vogal

com maior valor médio de F_0 foi a vogal [i] e a de menor valor foi a vogal [a]. Comparando estes valores com os dados normativos disponíveis para a voz laríngea, verificamos que o maior valor médio de F_0 , neste caso, ocorre na vogal [u] e o menor valor médio ocorre na vogal [o].

Verifica-se ainda que o valor médio de F_0 nas vogais analisadas, apresenta menor variação no grupo utilizador de VE do que no grupo utilizador de VTE.

5.1.3 Jitter

O *jitter* é uma medida de curto termo (ciclo a ciclo), de variabilidade não voluntária de F_0 que permite determinar o grau de estabilidade do sistema fonatório. (Guimarães, 2007)

Existem diversas possibilidades para analisar o *jitter*. O programa *Praat*, utilizado neste estudo, analisa as medidas acústicas dos dois grupos de laringectomizados utilizando várias medições: *jitter local*, *jitter local absoluto*, *jitter rap*, *jitter ppq5*, que já foram descritas previamente neste estudo (Boersma & Weenink, 1992-2005; Casper & Colton, 1998).

A VE é mais instável do que a voz laríngea, o que reflecte valores de *jitter* superiores (Behlau, 2001; Casper e Colton, 1998; Doyle, 1994).

O estudo de Robbins et al (1984) com 15 utilizadores de VE, 15 utilizadores de VTE e 15 falantes normais revelou que os maiores valores de *jitter* pertenciam ao grupo de VE, os valores intermédios ao grupo de VTE e os valores mais baixos atribuídos aos normafalantes (Casper&Colton, 1998).

Na Figura 19, podemos observar os valores médios e respectivos intervalos de confiança a 95% de todos os parâmetros de *jitter*, analisados pelo programa *Praat*, comparando os dois grupos de falantes

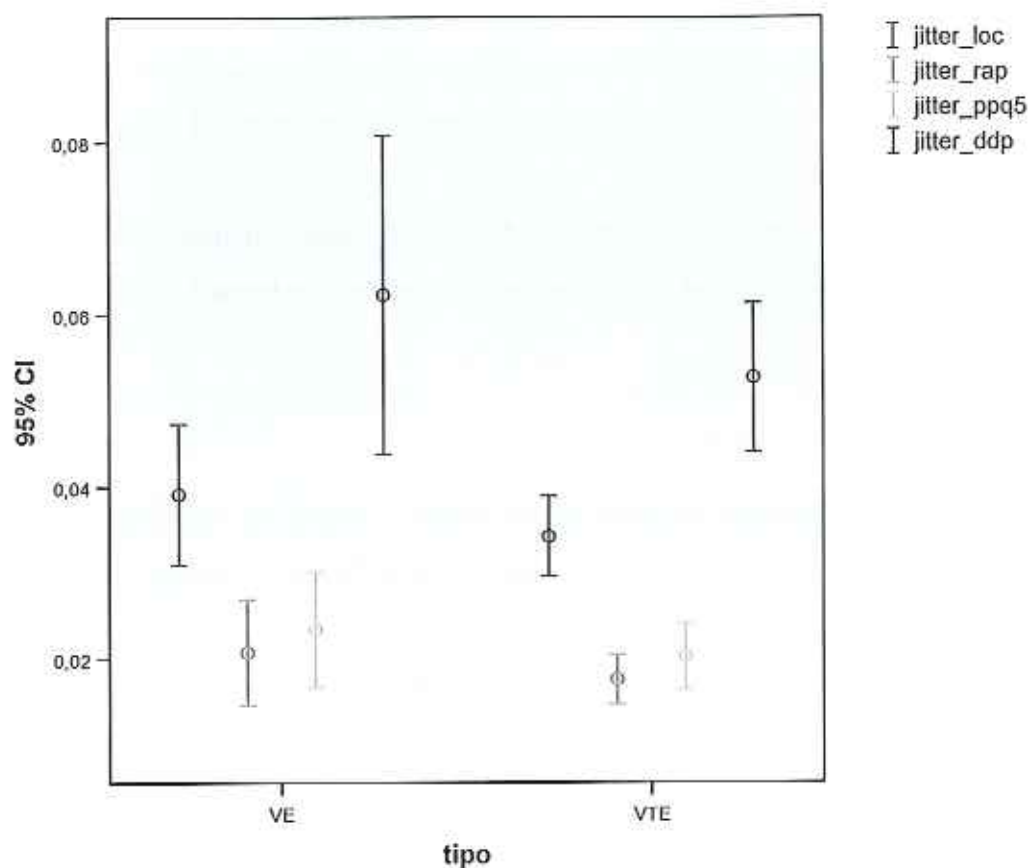


Figura 19 – Imagem representativa dos diversos valores de *jitter* por grupo de informante

Da análise da figura 19, podemos concluir que o valor médio de *jitter* para todas as formas de cálculo contempladas no nosso estudo, é superior nos indivíduos utilizadores de VE, se olharmos apenas para a média. Neste gráfico, podemos verificar que, apesar da sobreposição dos intervalos de confiança da figura, testes estatísticos revelaram como significativa a diferença para *jitter local* [$t(227) = -2.827, p = 0,005$]² e não significativa para *jitter - RAP* [$t(116,127) = -1.511, p = 0.172$] e para *jitter ppq5* [$t(79.419) = -0.812, p = 0.419$]³. Tal como é referido na literatura, os valores de *jitter* de VE e VTE são muito semelhantes e são superiores aos valores obtidos na voz laríngea. Esta situação poderá ocorrer, tal como já foi referido, pelo facto de os dois grupos de laringectomizados utilizarem o mesmo ponto de vibração, o segmento faringo-esofágico.

² Não assumindo igualdade de variância

³ o valor de ddp não é referido uma vez que está relacionado com outro dos parâmetros sendo igual a três vezes RAP

De modo a poder verificar se há diferenças significativas entre os informantes de cada um dos grupos, analisemos a Figura 20 onde se mostram as diferentes medidas de *jitter* por informante, num intervalo de confiança de 95%

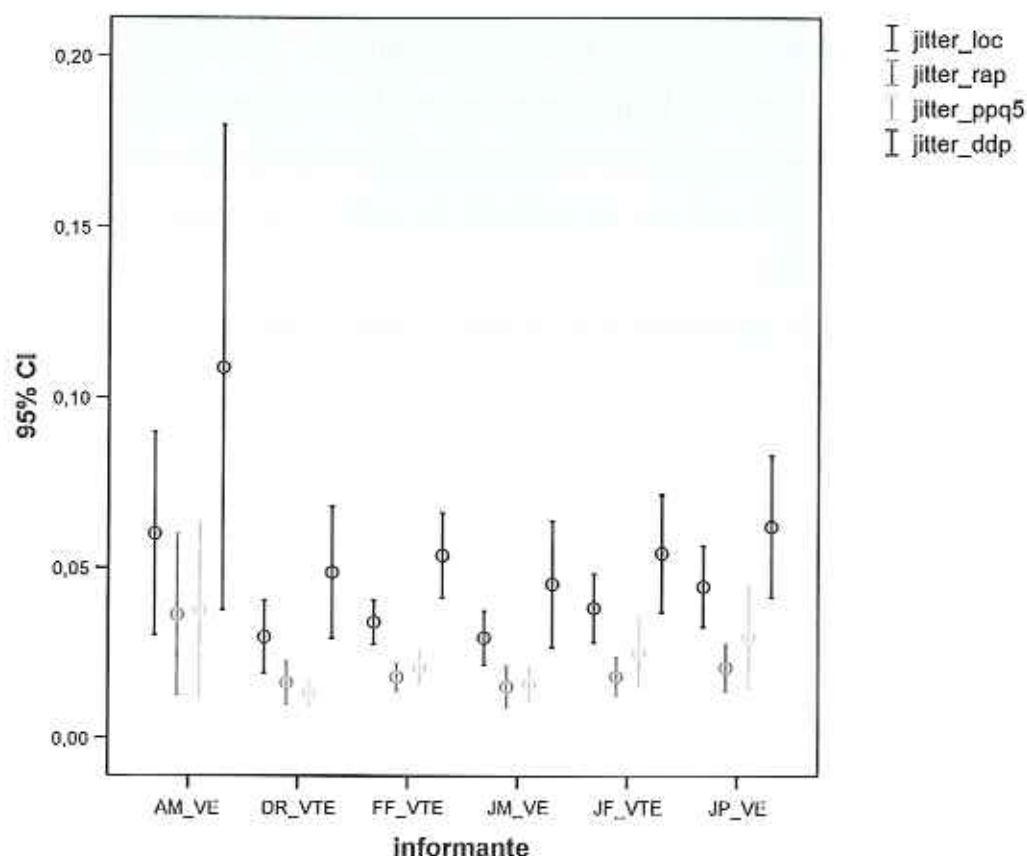


Figura 20 – Imagem representativa dos valores de *jitter* obtidos, por Informante

Nesta figura, pode observar-se que os valores de *jitter* mais elevados pertencem ao informante A. M. que apresentava no parâmetro F_0 o valor mais baixo. Este resultado está de acordo com a literatura que nos diz que os valores de *jitter* mais elevados ocorrem nos falantes com voz mais grave e com menor grau de intensidade (Behlau, 2001). De notar, mais uma vez, a semelhança de valores médios dos informantes D. R. e J. M., que foram utilizados no teste de identificação e avaliação da qualidade vocal.

Nos restantes informantes, apresentados neste estudo, não parece haver diferenças significativas entre os valores de *jitter* apresentados, tendo os resultados sido confirmados por testes estatísticos t- student realizados separadamente para cada um dos parâmetros.

5.1.4 Shimmer

Tal como o *jitter*, o *shimmer* tende a ser ligeiramente elevado nas perturbações da voz, apresentando valores mais altos nas frequências graves e sons de baixa intensidade (Behlau, 2001; Guimarães, 2007)

Segundo a literatura, as medidas de *shimmer* são superiores nos falantes de VE e VTE relativamente às medidas obtidas nos falantes normais. Esta conclusão parece óbvia, dado que, quer a VE quer VTE apresentam valores de F_0 e de intensidade baixos.

Para o *shimmer*, o programa *Praat* também apresenta várias medidas já referidas anteriormente neste trabalho, pelo que se analisará, de seguida, os valores médios de *shimmer* obtidos neste estudo.

Na Figura 21, estão representados os valores médios e respectivos intervalos de confiança a 95% dos diferentes valores de *shimmer*, obtidos no estudo.

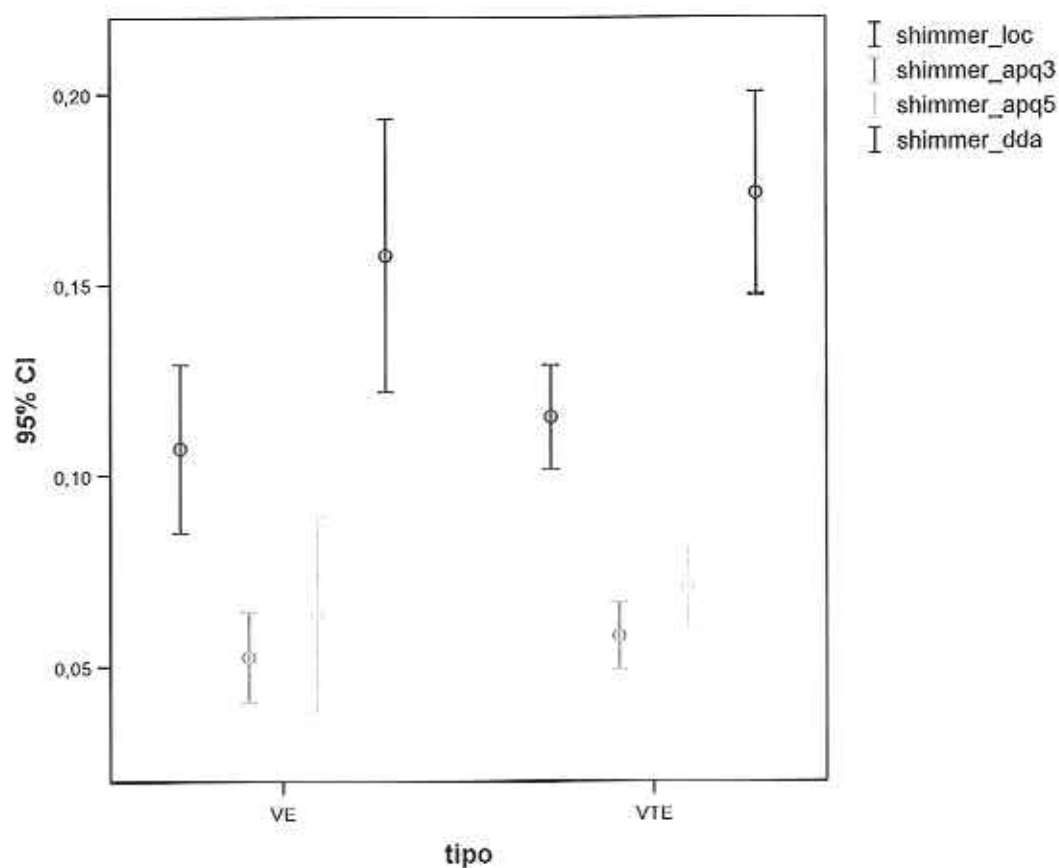


Figura 21 - Imagem representativa dos diversos valores de *shimmer* por grupo de informantes

Na figura 21, podemos observar que, apesar de os valores de *shimmer* para a VTE serem ligeiramente superiores, não há diferenças significativas em nenhuma das medidas apresentadas, tendo os resultados sido confirmados por testes estatísticos t-student, realizados separadamente para cada um dos parâmetros.

5.1.5 Medidas de ruído (Proporção Harmónico-Ruído)

A proporção harmónico-ruído, conhecida em inglês como *Harmonic-to-noise-ratio* (HNR), é uma medida que relaciona a componente harmónica versus a componente de ruído da onda acústica. Isto é, se o ruído espectral (inter-harmónico) aumenta, parece lógico que a energia da frequência dos harmónicos baixe. (Guimarães, 2007). Por esta razão, criou-se o cálculo matemático designado por índice sinal-ruído que é a divisão entre a média da amplitude do sinal e a medida da amplitude do componente de ruído da onda. A unidade usada é o dB (Guimarães, 2007)

Yumoto et al. (citado por Baken & Orlikoff, 2000) consideram que, para a voz normal, a média de HNR normal é de 11,9 dB. Behlau (1997) refere que os valores médios de normalidade para os homens são de 11,8 dB (no registo modal), considerando patológicos os valores inferiores a 7 dB.

Na Figura 22, analisam-se apenas os valores médios e respectivos intervalos de confiança a 95% de HNR obtidos neste estudo.

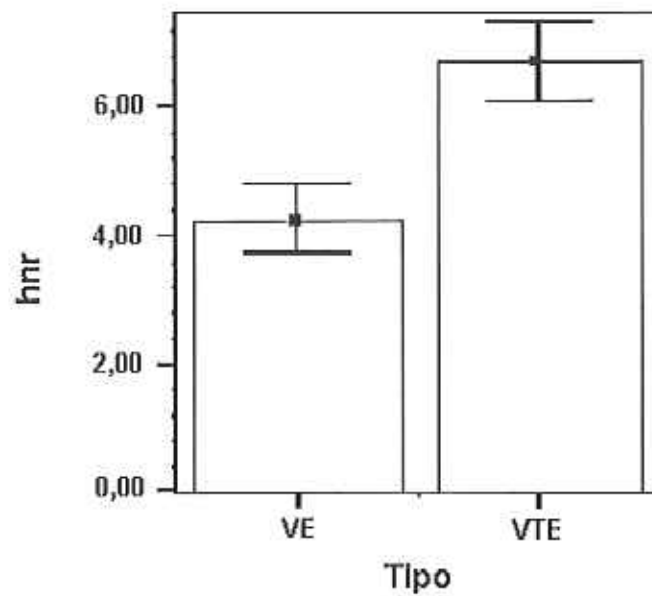


Figura 22 - Valores médios de HNR por tipo de voz

Na figura 22, podemos observar que o HNR é inferior nos utilizadores de VE em relação aos utilizadores de VTE, mas que a média de ambos os grupos é inferior a 7dB, concluindo que as vozes dos dois grupos são, obviamente, patológicas. O facto de a VE ter uma média inferior a VTE sugere que esta voz apresenta bastante ruído na sua emissão, com um valor elevado da amplitude do componente de ruído da onda.

Será interessante observar a Figura 23, onde é apresentado o valor médio e respectivo intervalo de confiança a 95% de IINR, por vogal, em cada um dos grupos de informantes.

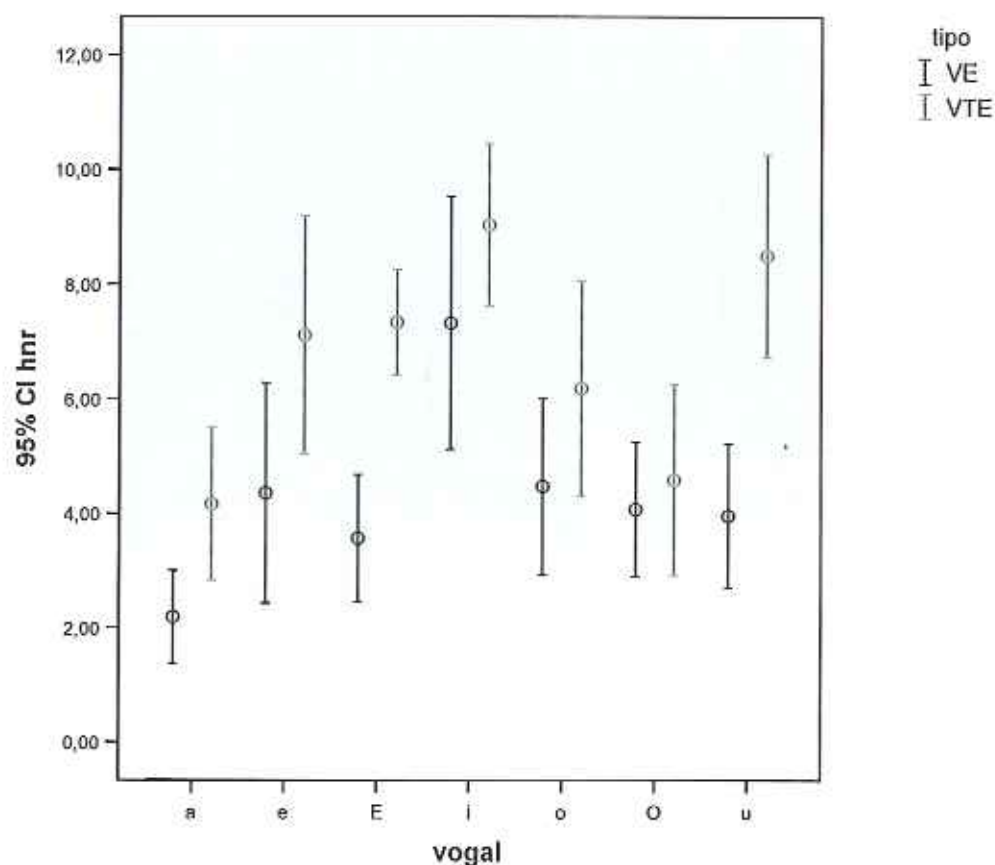


Figura 23 – Valores médios de HNR por vogal em cada um dos tipos de voz

Na figura 23, pode observar-se que os valores médios de HNR para VE são mais baixos em todas as vogais, o que sugere uma produção destas com maior índice de ruído do que o do grupo de VTE. Nas vogais [E] e [u] as diferenças são significativas e também o são para a vogal [a] ($p = 0,011$). Para as restantes vogais ($p > 0,05$) as diferenças não são significativas, comprovado pela realização de testes t-student realizados separadamente para cada um dos parâmetros. Podemos observar ainda que os valores mais baixos de HNR pertencem às vogais [a] e [E] para o grupo de VE e às vogais [a] e [O] para o grupo de VTE. O valor mais elevado de HNR pertence à vogal [i] para os dois grupos.

Vejamos de seguida, na Figura 24, o valor médio e respectivo intervalo de confiança a 95% de HNR, por informante.

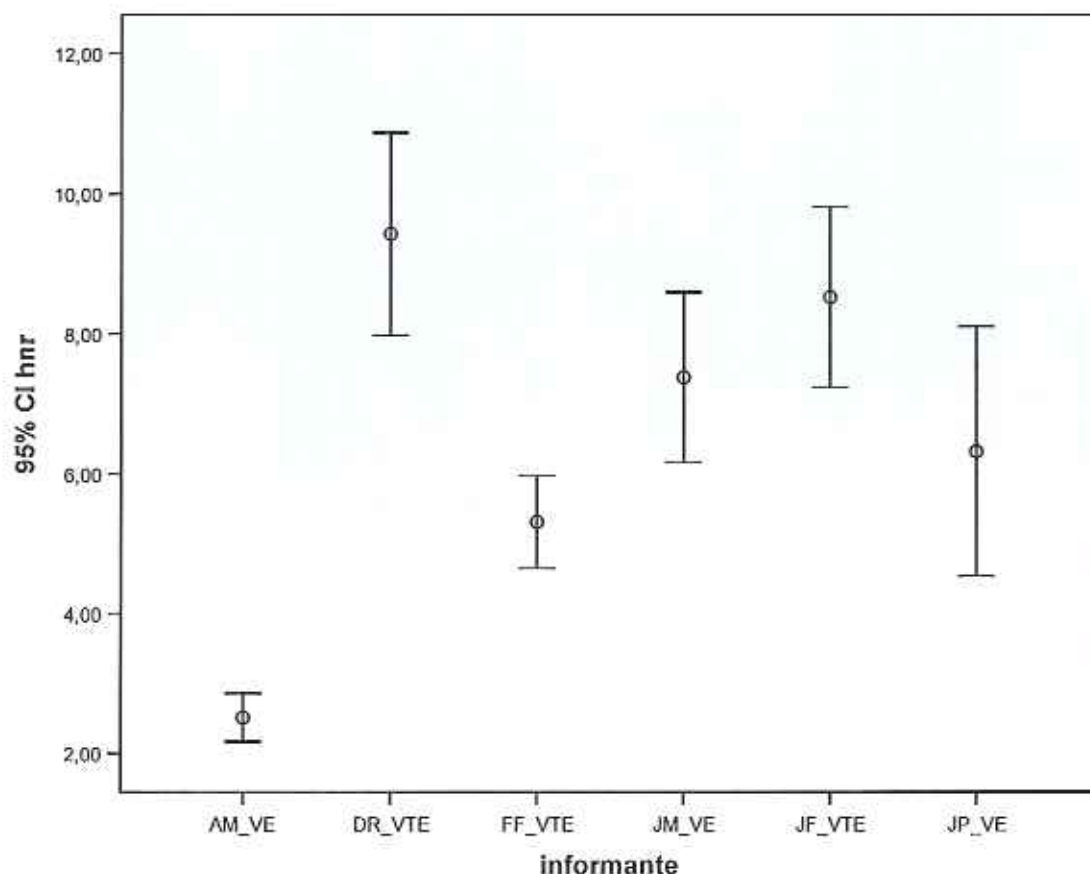


Figura 24 – Valores médios de HNR por Informante

Nesta figura, podemos observar que o informante que tem o valor médio de HNR mais baixo é A. M., o que sugere uma voz com poucos harmónicos e muito ruído. O valor médio mais elevado de HNR pertence ao informante D. R., cujo valor é o que mais se aproxima da normalidade (11,8, segundo Behlau, 2001). Os informantes D.R., J.F. e J.M. apresentam vozes com valores superiores a 7 dB e os restantes apresentam valores médios inferiores ao valor citado. Podemos também salientar que há dois indivíduos utilizadores de VTE com valores superiores a 7 dB. De qualquer forma, os utilizadores de VE e VTE apresentam valores médios de HNR bastante baixos, o que indica que, de uma forma geral, apresentam uma voz pobre em harmónicos e com bastante ruído.

5.1.6 Formantes

Na aprendizagem da voz esofágica podem ser utilizados diferentes métodos de introdução de ar no esófago para que o segmento faringo-esofágico vibre e se produza a voz. O método mais utilizado e o mais eficaz, para a grande maioria dos utilizadores de VE, é o método de injeção de ar.

De forma a injectar ar no esófago para utilização de VE, é necessário um movimento considerável da língua, o mesmo acontecendo para a produção individual de sons. A necessidade que os utilizadores de VE têm de colocar a língua em determinada posição de forma a conseguir a injeção de ar para a fala, afecta a posição da língua durante a produção dos sons. (Hoops & Noll, 1969 citados por Casper & Colton, 1998).

A frequência das formantes (F1 e F2) é maior nos utilizadores de VL e VLE do que nos utilizadores de VL, o que sugere a presença de um tracto vocal mais curto, resultante do procedimento cirúrgico a que foram submetidos. Os falantes poderão compensar este facto fazendo protrusão dos lábios ou alterando a posição da língua durante o discurso. (Casper & Colton, 1998; Jorge M. e al, 2004)

5.1.6.1 F1

Tal como já foi referido no capítulo de definição do *corpus*, para a análise acústica das vogais, consideram-se mais importantes as frequências das duas primeiras formantes (F1 e F2). O valor da frequência da primeira formante (F1) está relacionada com a altura da língua, pelo que vogais com posição alta da língua apresentam F1 reduzido e vogais com posição baixa da língua apresentam valores de F1 elevado.

Verifiquemos os resultados obtidos no nosso estudo, analisando a Figura 25, onde se encontram os valores médios e respectivo intervalo de confiança a 95% para a média de F1, nas vogais estudadas.

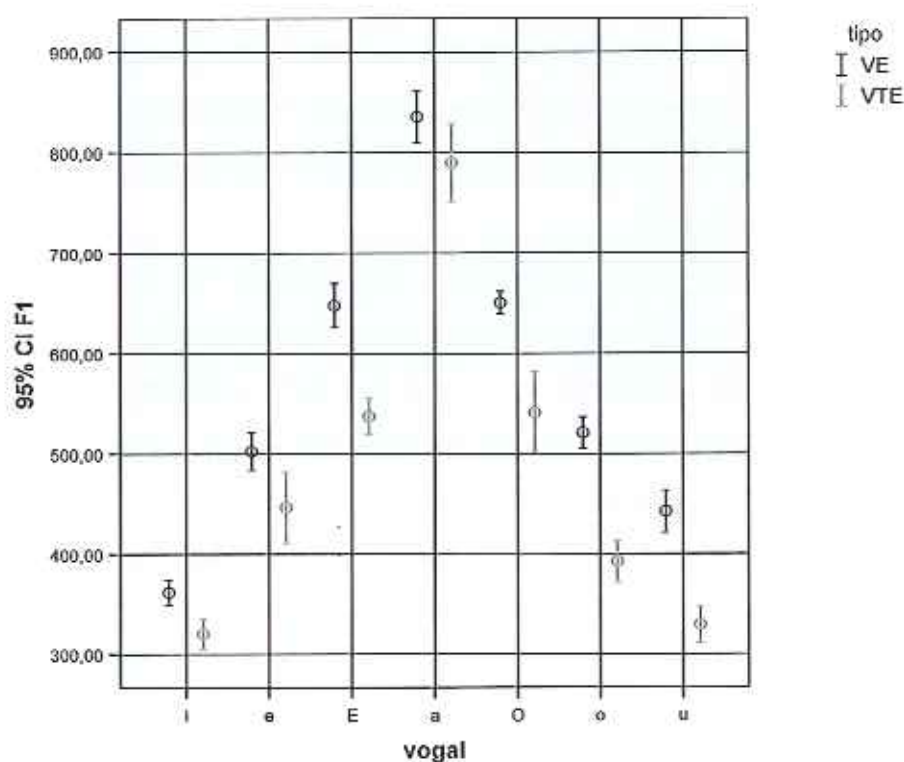


Figura 25 – Comparação dos valores médios F1 para os falantes de VE e VTE

Numa primeira análise, verificamos que na realidade, F1 é superior para todas as vogais nos falantes com VE relativamente aos utilizadores de VTE, o que não está de acordo com a literatura consultada.

Os valores médios de F1 são elevados para todas as vogais, tendo sido mantida, entre todas as classes, a diferença de valor médio (relacionado com a altura da língua). Esta situação contribui para que elas sejam reconhecidas como vogais diferentes e pode observar-se, no gráfico, um esboço do triângulo das vogais.

Verificamos também que as diferenças mais marcadas se verificam para as vogais [E], [O], [o], [u] e [i]. Estas diferenças são significativas, uma vez que os intervalos de confiança não se sobrepõem.

Podemos também concluir que, para o Português Europeu, as vogais que apresentam F1 mais baixo são [i] e [u], que são aquelas onde se utiliza uma posição mais elevada da língua e que a vogal com valor de F1 mais elevado é a vogal [a] que apresenta posição mais baixa da língua.

Parece ser unânime, na literatura, o facto de F1 na VTE ser superior à VE e à voz laríngea e, tal como referem Christensen & Weinberg (1976) e Cervera et al (2001), o que contraria os dados obtidos no nosso estudo.

Comparando com o estudo Espanhol que serviu de base ao nosso trabalho, verifica-se que o comportamento relativamente a este parâmetro é diferente nas duas línguas. Enquanto no estudo para a língua espanhola os valores de F1 na VTE são superiores à VE, no estudo para o PE os valores de VE são superiores aos obtidos para a VTE. Comparando os valores de VE para as duas línguas, verifica-se uma semelhança de valores, excepto para a vogal [u]. No caso da VTE, os valores obtidos no estudo Espanhol são superiores aos obtidos para o PE, excepto para a vogal [a], tal como se verifica na tabela seguinte.

VOGAIS	[i]	[e]	[ɛ]	[a]	[O]	[o]	[u]
F1 VE (PE)	350 Hz	500 Hz	650 Hz	840 Hz	650 Hz	520 Hz	450 Hz
F1 VE (E)	348 Hz	532 Hz	--	866 Hz	--	540 Hz	326 Hz
F1 VTE (PE)	320 Hz	450 Hz	540 Hz	790 Hz	540 Hz	400 Hz	320 Hz
F1 VTE (E)	355 Hz	546 Hz	--	781 Hz	--	603 Hz	459 Hz
F1 VL (PE)	305 Hz	431 Hz	566 Hz	694 Hz	564 Hz	438 Hz	308 Hz

Tabela 4 - Valores de F1 obtidos nos estudos para o PE e para a língua Espanhola

Optou-se ainda pela inclusão dos valores obtidos para a VL num estudo realizado por Teixeira et al (2002) para o PE, de forma a comparar a média de F1 obtida para cada um dos tipos de voz (VE, VTE e VL). Verificou-se, de uma forma geral, que os valores obtidos para a VTE e a VE são superiores aos obtidos para a VL. Esta situação está de acordo com os resultados obtidos no estudo de Cervera et al, que serviu de base ao nosso estudo.

5.1.6.2 F2

O valor da segunda formante (F2) está relacionado com a posição antero-posterior da língua e com o grau de arredondamento dos lábios e, por isso, vogais anteriores, produzidas sem arredondamento dos lábios, apresentam valores de F2

elevados enquanto vogais posteriores e produzidas com arredondamento dos lábios apresentam valores de F2 baixos.

Na figura 26, pode observar-se o que acontece no estudo para o Português Europeu relativamente aos valores médios e respectivo intervalo de confiança a 95% de F2, por vogal, em cada um dos grupos de informantes.

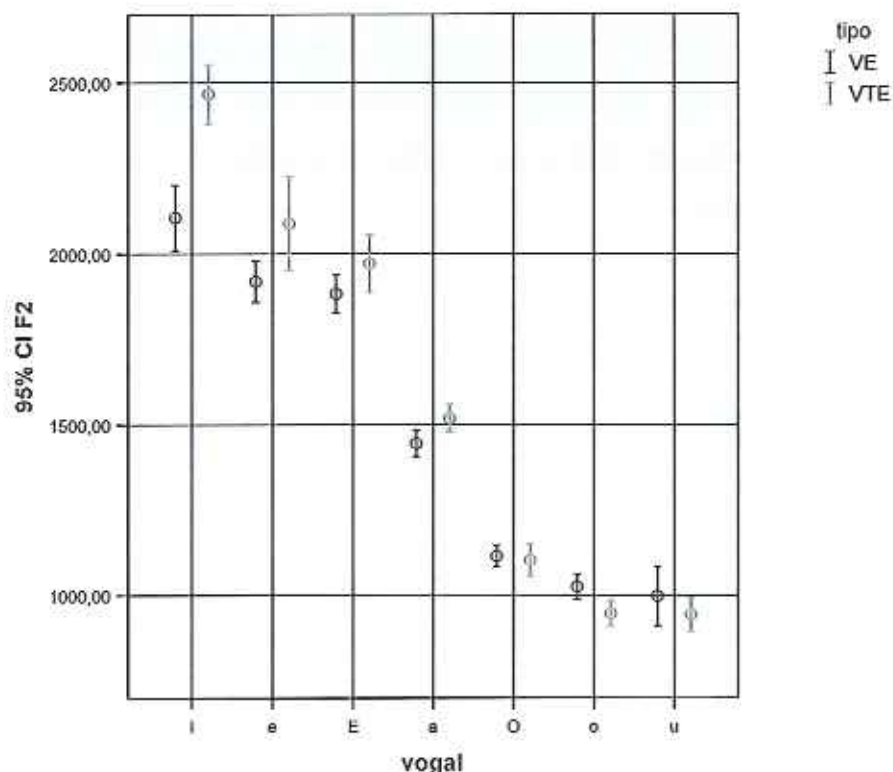


Figura 26 - Comparação dos valores médios de F2 obtidos para o grupo de VE e VTE

Na figura 26, podemos observar que os valores de F2 seguem a descrição encontrada na literatura, ou seja, as vogais mais anteriores e com menor grau de arredondamento dos lábios como [i, E, e] e [a] têm valores de F2 superiores aos das vogais mais posteriores e com maior grau de arredondamento dos lábios como [o, O, u].

É possível também observar que os valores de F2 são superiores nos utilizadores VTE do que nos utilizadores de VE, excepto para a vogal [O], onde a média de F2 é semelhante e nas vogais [o] e [u], onde o valor é ligeiramente inferior, embora não significativa, comprovado pela realização do teste t-student para cada uma das variáveis. Ou seja, as vogais anteriores produzidas pelos utilizadores de VTE têm valores de F2 superiores aos utilizadores de VE, mas o mesmo parece não se verificar

para as vogais mais posteriores, uma vez que as diferenças não são significativas. Ao compararmos estes dados com os resultados obtidos para a Língua Espanhola observamos que, se olharmos apenas para a média, no Português Europeu, se verificam dois comportamentos. O valor de F2 aumenta de VE para VTE para as vogais [i,e,E e a] – as vogais anteriores e central- e diminui para [O,o e u] - as posteriores. Na Língua Espanhola, em geral, o valor de F2 aumenta de VE para VTE, excepto para a vogal [e].

VOGAIS	[i]	[e]	[E]	[a]	[O]	[o]	[u]
F2 VE (PE)	2100 Hz	1920 Hz	1900 Hz	1450 Hz	1110 Hz	1020 Hz	1000 Hz
F2 VE (E)	2518 Hz	2180 Hz	--	1498 Hz	--	970 Hz	832 Hz
F2 VTE (PE)	2470 Hz	2100 Hz	1980 Hz	1500 Hz	1100 Hz	980 Hz	980 Hz
F2 VTE (E)	2636 Hz	2087 Hz	--	1576 Hz	--	1079 Hz	883 Hz
F2 VL (PE)	2134Hz	1865Hz	1838Hz	1341 Hz	1045Hz	1100Hz	890 Hz

Tabela 5 - Valores de F2 obtidos nos estudos para o PE e para a língua Espanhola

5.2 Discussão

Partindo do exposto, procederemos à discussão dos resultados obtidos, tentando a comparação, em todos os parâmetros possíveis, com o estudo de Cervera et al (2001), que serviu de base ao nosso trabalho.

O primeiro parâmetro a ser analisado relaciona-se com a duração das vogais e no nosso estudo, podemos concluir que, em termos médios, o valor da duração das vogais para o Português Europeu é superior nos utilizadores de VE e VTE, quando comparado com utilizadores de VL. Este resultado é concordante com os estudos consultados. Na realidade, verifica-se uma duração superior das vogais para os utilizadores de voz não laríngea (Cervera et al, 2001). No estudo para o PE, verifica-se que a duração média de todas as vogais é superior na VE, quando comparada com VTE, o que nos levaria a considerar o facto de haver dificuldades de controlo voluntário do segmento faringo-esofágico e, daí, ser difícil controlar o tempo de produção de uma determinada vogal. O facto da duração da VTE ser mais próxima à da VL poderá estar relacionado com o facto de, no contexto de palavra, a duração da vogal ser normalmente mais curta devido à velocidade articulatória. Segundo a literatura (Corinna, 2001; Cervera et al, 2001), a VTE parece ser a que mais se aproxima em termos de velocidade de discurso e comprimento médio de frases da VL. Assim é de supor que estes valores estejam mais próximos do que os valores obtidos para a VL.

Os resultados obtidos para as vogais do PE contrariam os resultados do estudo de Cervera et al (2001), que refere que os valores médios de duração das vogais são superiores em indivíduos utilizadores de VTE, em relação aos utilizadores de VE. Esta diversidade de resultados pode estar relacionada com as diferenças linguísticas (número de vogais) encontradas para o PE e o Espanhol, e o facto de, os contextos CVCV utilizados para os dois estudos, terem algumas diferenças.

Outro dos parâmetros analisado neste estudo foi o valor de F_0 , que se verificou estar em concordância com a literatura consultada. Isto significa que, a VE e VTE têm valores médios mais baixos que VL e, que, ao fazermos a comparação entre as duas primeiras, verificamos que os valores de VE são inferiores à VTE. Os resultados médios obtidos são concordantes com os obtidos na literatura que referem valores de F_0 para VE entre 77 e 103 Hz (Corina, 1998), com alguns estudos a referir valores de 60 e 64 Hz. No estudo aqui apresentado, este valor situa-se pelos 75 Hz o que vai de encontro aos intervalos já referidos. Relativamente à VTE, que neste estudo apresentou valores

de cerca de 100 Hz, o resultado também vai de encontro à literatura consultada que revela valores de F_0 entre os 86 e os 112 Hz. Os valores médios de F_0 para homens adultos dentro desta faixa etária, tal como já foi referido anteriormente, encontra-se entre os valores médios de 120 a 130 Hz. O estudo espanhol de Cervera et al não contemplou este parâmetro pelo que não podemos aqui realizar qualquer tipo de comparação.

No que se refere aos parâmetros de perturbação *jitter* e *shimmer*, verificamos que, de uma forma geral, os dados obtidos neste estudo são concordantes com a literatura consultada. No caso dos valores de *jitter*, observou-se que os valores mais elevados pertencem à VE. Como este tipo de voz é mais grave do que VTE, são aceitáveis as diferenças encontradas. Ainda de referir que estes valores são superiores aos encontrados para a VL, o que será também fácil de compreender uma vez que estes falantes (de VI) apresentam uma voz menos grave e com intensidade superior relativamente às duas vozes apresentadas neste estudo. Quanto aos valores de *shimmer*, não foram encontradas diferenças entre os dois grupos estudados. Contudo, foi possível verificar que estes valores são superiores aos obtidos para a VL. A obtenção de valores semelhantes para os dois tipos de voz analisada neste estudo, pode estar relacionada com a utilização do segmento faringo-esofágico como fonte de vibração sonora em ambos os casos. O estudo dos valores de *jitter* e *shimmer* para a população de utentes laringectomizados não é abordado com grande regularidade na literatura, pelo que se torna complicada a retirada de conclusões mais aprofundadas sobre os valores obtidos neste estudo.

Analisando agora os valores de HNR obtidos, verificou-se que, de uma forma geral, a VE apresenta valores inferiores relativamente à VTE, o que sugere que a VE é produzida com maior índice de ruído e fraca em harmónicos. Este dado poderá estar relacionado com diversos aspectos subjacentes à utilização do segmento faringo-esofágico, e também à forma como os utilizadores de VE utilizam o ar para a produção de voz que é diferente em cada um dos casos. Assim, o ar, quando passa pelo referido segmento, nos casos de VE, tem uma pressão diferente e é libertado com maior tensão do que nos outros casos (VTE), onde o ar apenas é desviado para a prótese fonatória, permitindo uma vibração mais homogénea do segmento. Esta situação é perceptível quando se compara auditivamente os dois tipos de voz. Contudo, dado que não foram encontradas referências relativas a este parâmetro na literatura, esta análise requer um estudo mais aprofundado que será sugerido posteriormente.

Os valores de F1 e F2 foram outros parâmetros analisados neste estudo e que merecem uma discussão mais alargada. Verificou-se, pois, que os valores de F1 para a VE são mais elevados do que os valores obtidos para a VTE. Esta situação poderá estar relacionada com o facto de o método de injeção de ar, utilizado preferencialmente pelos utilizadores de VE, que participaram neste estudo, obrigar a um abaixamento da altura da língua no momento de produção das vogais o que origina uma subida nos valores de F1. Os resultados obtidos neste estudo contrariam os valores apresentados no estudo espanhol, que revelam F1 na VE com valores inferiores aos da VTE. A diferença obtida nos dois estudos poderá estar relacionada com vários factores: (a) o método de introdução de ar utilizado pelos falantes da língua espanhola poder não ser o mesmo do que é utilizado pelos falantes de PE (não está referido no artigo publicado); (b) o contexto em que foram produzidas as vogais, sendo que no PE se escolheu a combinação CVCV (palavras iniciadas com consoante oclusiva surda) e no estudo espanhol ter sido utilizada a mesma combinação, mas para todas as consoantes da língua em referência. Apesar de não ser um dos objectivos iniciais deste estudo, optou-se pela comparação dos valores obtidos para a VE e VTE com dados obtidos para a VL, com base num estudo realizado por Teixeira et al (2002) e verificou-se que F1 é superior na VE e VTE quando comparado com a VL. Estes dados são concordantes com os resultados encontrados na literatura consultada.

Os valores de F2, apesar de seguirem a descrição da literatura que se refere a valores superiores para vogais mais anteriores e com menor grau de arredondamento dos lábios como [i, E, e] e [a] do que os das vogais mais posteriores e com maior grau de arredondamento dos lábios como [o, O, u], não foi conclusivo para os dois grupos de laringectomizados estudados. Esta situação deve-se ao facto de terem sido encontrados valores superiores para a VTE vogais mais anteriores e, nas posteriores, essa situação não se verificar. Apesar de múltiplas tentativas para encontrar uma resposta que apoiasse estes resultados, não foi possível chegar a nenhuma conclusão. Contudo, ao compararmos os dados obtidos na VTE e VE com os valores da VL, verificamos que, tal como vem descrito na literatura, os valores obtidos para os dois grupos de laringectomizados são superiores aos obtidos para o grupo de normofalantes. As diferenças encontradas para os valores de F1 e F2 para o PE poderão estar relacionadas com um diferente posicionamento da língua durante a utilização de dois métodos de aquisição de voz que são substancialmente diferentes. Enquanto na VE o método de injeção implica um maior movimento da língua para a produção das vogais, na VTE,

sendo a fonte de ar pulmonar e não bucal, a língua apresenta um movimento mais semelhante ao realizado pelos utilizadores de VL.

Contudo, a diferença de valores de F1 e F2 obtida para a VTE e VL, relativamente à VL, poderá ser justificada pelo facto de os falantes laringectomizados apresentarem, devido ao procedimento cirúrgico a que são submetidos, um tracto vocal mais curto que os utilizadores de VL. Estes resultados estão de acordo com estudos realizados para a voz laríngea onde, ao serem comparados os valores de F1 e F2 para grupos de Homens, Mulheres e Crianças se verificou que, quando mais curto o tracto vocal, maiores os valores de F1 e F2.

Estas conclusões, relativas ao posicionamento da língua na produção de palavras ou grupos de palavras e ao tamanho do tracto vocal só poderão ser verificadas através de estudos realizados para utentes laringectomizados totais, durante a produção de palavras isoladas, utilizando Ressonância Magnética e comparando com estudos já realizados para o PE. Desta forma, será possível observar qual o posicionamento da língua durante a produção de palavras, tentando perceber as diferenças entre os utilizadores de VL, VE e VTE.

Capítulo 6 - Avaliação Perceptiva

6.1 Introdução

A fala humana é essencialmente utilizada para comunicação. Isto implica que o julgamento da voz, sendo esta uma propriedade da fala, é de carácter subjectivo (Hammarberg & Gauffin, 1995 citados por Corinna, 2001). Na rotina clínica dos Terapeutas da Fala, a avaliação perceptiva é a avaliação clássica da qualidade vocal e, embora muitas críticas sejam feitas à subjectividade e à terminologia envolvida neste procedimento, nem mesmo a análise acústica ameaça a sua soberania (Behlau, 2001). Assim, Hirano (1989) define-a referindo que se baseia na análise subjectiva do examinador de acordo com protocolos internacionais padronizados, de forma a permitir a comparação de resultados em tempos diferentes ou efectuados por diferentes técnicos (Pinto Moura, 2006).

Nas últimas décadas, esta forma de avaliação tem vindo a ser desenvolvida, com a criação de diversas escalas ou índices, de forma a padronizar e obter um carácter menos subjectivo deste tipo de avaliação. Foram, então, criadas escalas de base fonética como o 'Vocal Profile Analysis Protocol', desenvolvida por Laver (1980), a escala GRBAS (Grade, Roughness, Breathiness, Asthenic, Strained) proposta pela Sociedade Japonesa de Logopedia e Foniatria (Hirano, 1980), traduzida e validada para português por Pinho e Pontes (2002) com o nome RASAT (Rouquidão, Aspreza, Soproidade, Astenia e Tensão) e escalas de diferenciação semântica como a criada por Osgood et al (1957) e descritas por outros autores que a aplicaram e estudaram para o tipo de vozes em análise neste trabalho.

Note-se que a avaliação da qualidade vocal dos indivíduos utilizadores de VE e VTE também é baseada em aspectos perceptivos e a literatura refere a realização de alguns trabalhos (Nieboer, 1998; Cain, 1988 citados por Corinna, 2001) onde as escalas de diferenciação semântica foram utilizadas para diferenciar aspectos relacionados com a qualidade vocal, em indivíduos utilizadores de VE e VTE.

Corinna (2001) descreve alguns estudos realizados por diferentes autores, que tendo recorrido à utilização de escalas para análise e comparação de determinados itens, avaliam a qualidade vocal e nos quais parece haver uma tendência generalizada para

que a VTE tenha uma avaliação positiva relativamente à VE. O trabalho desta autora refere-se apenas ao estudo da qualidade vocal em utilizadores de VTE, uma vez que, no centro onde desenvolve as suas funções, é esta a única opção de comunicação existente. O estudo consistia na audição de um texto em Holandês com 151 palavras que era lido pelos utilizadores de VTE e posteriormente apresentada a sua gravação a dois grupos de avaliadores: um dos grupos constituído por ouvintes não experientes e, um outro, grupo composto por 4 avaliadores experientes. Para este estudo foi utilizada uma escala semântica de 7-pontos, baseada na escala de Fagel et al (1983) para avaliação de vozes normais. A escala é apresentada na Tabela 6:

<i>Dutch term</i>	<i>English term</i>
goed-matig-slecht ¹	good-reasonable-poor ¹
afwijkend-normaal	deviant-normal
onaangenaam-aangenaam	unpleasant-pleasant
lelijk-mooi	ugly-beautiful
veel bijgeluid-geen bijgeluid	noise-no noise
eentonig-melodius	monotonous-melodious
uitdrukkingsloos-expressief	expressionless-expressive
zwak-krachtig	weak-powerful
onvast-vast	unsteady-steady
hortend-vloeiend	jerking-fluent
langzaam-snel	slow-quick
laag-hoog	low-high
diep-schel	deep-shrill
borrelig-niet borrelig	bubbly-not bubbly
hees-niet hees (wilde lucht)	breathy-not breathy
schor-niet schor	rough-not rough
krakerig-niet krakerig	creaky-not creaky
gespannen-ontspannen ²	tense-relaxed ²
dof-helder	dull-clear
hypertoon-niet hypertoon ²	hypertonic-not hypertonic ²
hypotoon-niet hypotoon ²	hypotonic-not hypotonic ²
onverstaanbaar-verstaanbaar	unintelligible-intelligible

¹Only for trained raters
²Only for naive raters

Tabela 6 - Escala Semântica utilizada por Corina (2001) para avaliação da qualidade vocal em indivíduos com VTE. Fonte: Tracheoesophageal Speech – A multidimensional Assessment of Voice Quality.

As conclusões do trabalho referido revelaram que, apesar de a VTE apresentar bons índices de qualidade, ainda se mostra como desviante relativamente à VL veio também demonstrar que a utilização de um grupo de avaliadores experientes é recomendada visto que apresentam um conhecimento bastante profundo da anatomia e morfologia da neoglote e têm uma “audição” mais analítica deste tipo de voz (Corinna, 2001).

6.2 Método

Na realização do estudo para o PE, apesar da excelência do trabalho atrás citado, por questões de ordem metodológica e de tempo, optou-se pela realização de um teste com facilidade de aplicação e rápido de analisar, servindo este para estudar a inteligibilidade de cada uma das vozes e avaliar de forma sumária a sua qualidade.

Para se realizar a identificação e avaliação perceptiva da qualidade vocal das vozes dos indivíduos utilizadores de VE e VTE, foi realizado um teste de identificação com respostas abertas (sem lista pré-estabelecida) onde os ouvintes tiveram a possibilidade de escrever livremente o estímulo ouvido. Cada estímulo (palavra) era apresentado duas vezes em momentos diferentes para que fosse analisada a consistência da resposta. Simultaneamente, foi realizado um teste de avaliação da qualidade de voz, segundo uma escala de Likert com 5 categorias de resposta - Muito Mau, Mau, Razoável, Bom e Muito Bom. O teste acima referido foi criado pelo orientador deste trabalho e administrado em programa Tcl/Tk, sendo baseado num teste já existente e utilizado em trabalhos de investigação no âmbito deste Mestrado. Foram utilizadas 14 palavras do *corpus* constituinte deste trabalho, duas para cada um dos dois informantes seleccionados aleatoriamente, um utilizador de VE e outro utilizador de VTE. Procurou-se que fossem utilizadas todas as vogais em contexto Oclusiva-Vogal e Fricativa-Vogal, mas, por dificuldades de anotação, duas palavras não foram incluídas neste teste (uma para cada um dos informantes), que, de alguma forma, poderá alterar os resultados. O teste de identificação e avaliação da qualidade vocal apresentava-se aos examinadores com a configuração electrónica que de seguida se apresenta.

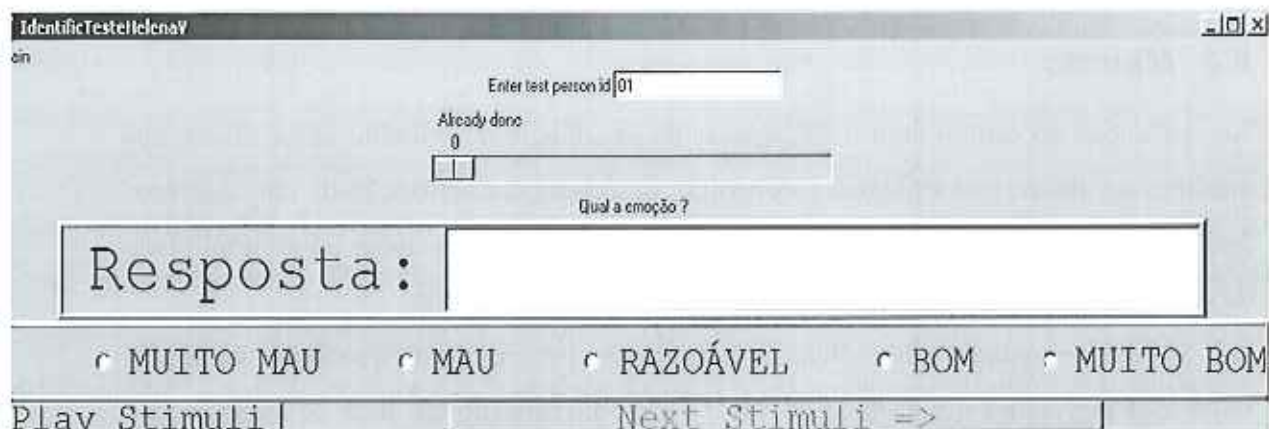


Figura 27 - Teste de Identificação e Avaliação da Qualidade Vocal

Foram seleccionados dois grupos de ouvintes com 4 elementos cada, um grupo composto por ouvintes não familiarizados com este tipo de voz (não experientes) e por um grupo composto por elementos experientes na comunicação com estes pacientes (médico ORL, Terapeuta da Fala, Estagiária de Terapia da Fala e Familiar). O grupo de ouvintes não familiarizadas era composto por 2 elementos do sexo feminino e 2 do sexo masculino com uma média de idades de 42 anos e o grupo de elementos experientes na comunicação com estes indivíduos era constituídos por 2 elementos de cada um dos sexos com uma média de idades de 30,75 anos. Os ouvintes foram informados dos objectivos do estudo e accitaram participar na realização deste teste.

6.3 Resultados

Após a realização do teste de identificação e avaliação já referido, tentaremos neste capítulo analisarmos, neste capítulo, os dados obtidos e responderemos a algumas questões previamente levantadas: que tipo de voz (VE ou VTE) apresenta melhores índices de identificação dos estímulos? Há diferenças de *performance* entre os dois grupos de avaliadores? Essas diferenças mantêm-se alterando os contextos linguísticos?

6.3.1 Resultados do Teste de Identificação

Nesta secção serão apresentados os resultados relativos ao Teste de Identificação de palavras apresentado aos dois grupos de avaliadores.

Na figura 28, optou-se por incluir todos os ouvintes sem se realizar qualquer tipo de referência ao grupo a que pertenciam, com o intuito de verificar qual o tipo de voz que obteve maior taxa de acerto para os dois grupos. Assim, obteve-se um gráfico com os resultados obtidos para os dois tipos de voz.

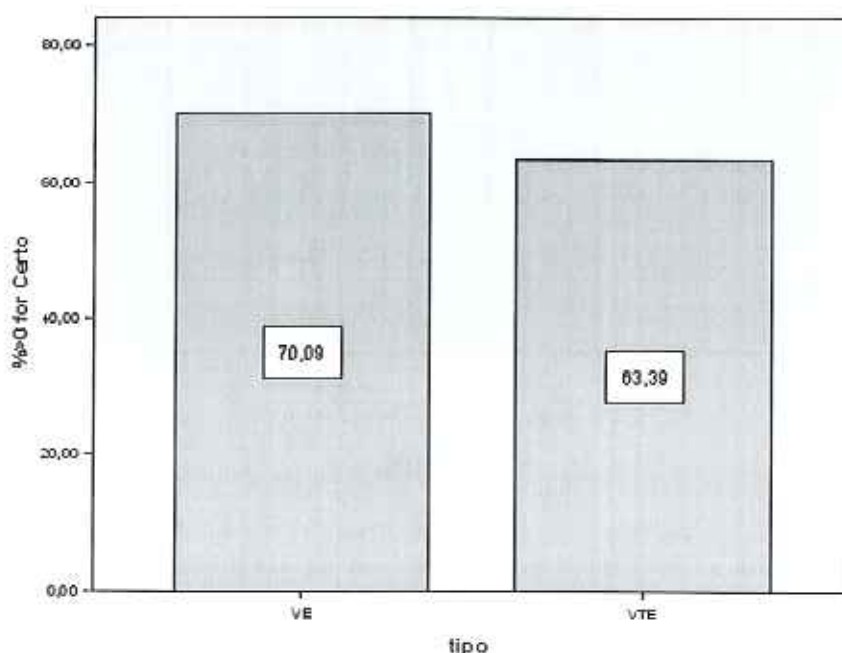


Figura 28 – Percentagem de acerto obtida para os dois tipos de voz, utilizando os dois grupos de ouvintes.

Pela análise da Figura 28, verificou-se que a VE obteve uma taxa de acerto de estímulos superior à VTE. A VE obteve uma taxa de acerto 70,09% e a VTE uma taxa de 63,39%.

Na figura 29, são apresentados os resultados obtidos relativos à taxa de acerto dos estímulos ouvidos, para cada um dos ouvintes.

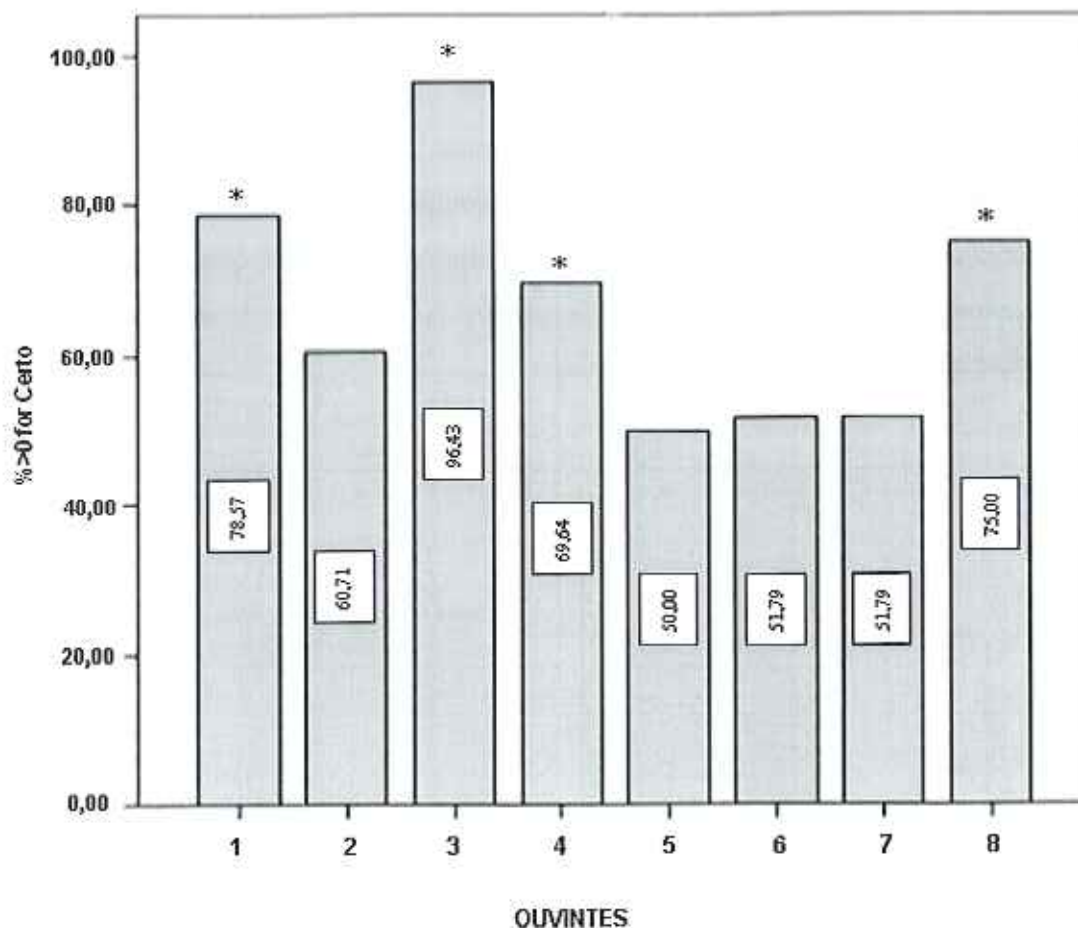


Figura 29 - Taxa de acerto/identificação para os ouvintes dos dois grupos. Assinalado com * o grupo de avaliadores experientes.

Na figura 29, verifica-se que os elementos representados com os números 1,3,4 e 8 e que integram o grupo de avaliadores experientes, tiveram uma taxa de acerto superior ao grupo de avaliadores não experientes, com valores superiores a 65%. Esta situação era previsível e confirma os dados publicados na literatura consultada.

Na figura 30, inclui-se um gráfico com a percentagem de acerto dos estímulos auditivos para cada um dos ouvintes (representados por números) e a sua preferência relativamente a cada um dos tipos de voz.

Para um melhor entendimento do gráfico, acrescenta-se que, os ouvintes representados com os números 1,3,4 e 8 representam o grupo de ouvintes experientes e os elementos representados pelos números 2,5,6 e 7 pertencem ao grupo de ouvintes não experientes.

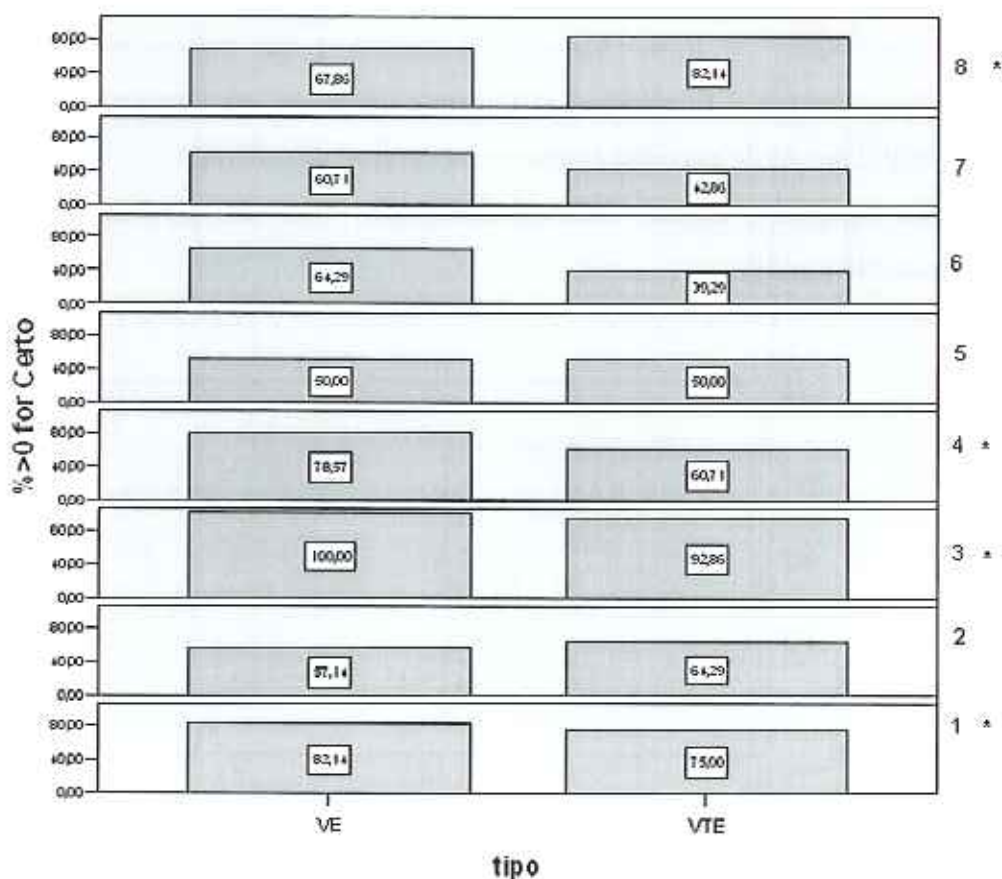


Figura 30 – Percentagens de acertos por ouvinte relativamente a cada um dos tipos de voz. Os ouvintes assinalados com * representam o grupo de avaliadores experientes.

Pela análise da Figura 30, pode verificar-se que, no grupo de avaliadores, apenas 2 elementos apresentaram uma maior percentagem de acertos para a VTE; todos os outros elementos obtiveram uma taxa de acerto superior para a VE e um dos elementos (número 4, do grupo não experiente) obteve uma taxa de acerto igual para ambos os grupos. Os elementos que tiveram uma melhor taxa de identificação para a VTE foram o elemento representado com o número 1 (elemento do grupo experiente) e o elemento 7 (do grupo experiente). Neste gráfico, também se pode verificar que existem diferenças de prestação entre os dois grupos. Neste momento, podemos dizer que para estes dois grupos de ouvintes, há uma taxa de acerto superior para o grupo mais experiente e que a voz com maior perceptibilidade é a VE.

Os estímulos apresentados no Teste foram, como já foi referido, retirados do corpus produzido pelos utilizadores de VE e VTE e apesar de seguirem todos a

estrutura silábica CVCV, foram gravadas palavras iniciadas por consoantes oclusivas surdas e outras iniciadas por consoantes fricativas surdas. Na VE e VTE, estas palavras são percebidas de forma diferente e mesmo a sua aquisição por parte dos utilizadores, também é diferente. Verifiquemos então se, ao alterarmos a consoante inicial, há diferenças de percepção para os dois grupos de ouvintes.

Na Figura 31, iremos observar um quadro onde são apresentados os dados relativos a este aspecto.

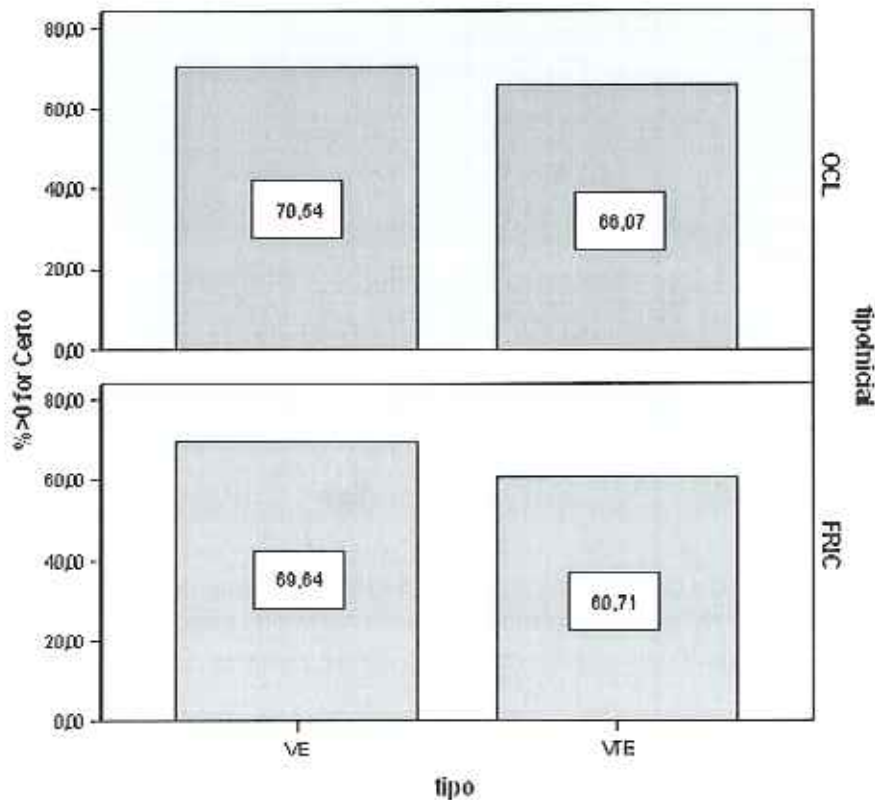


Figura 31 – Taxas de acerto para a VE e VTE em palavras com diferentes contextos iniciais

Tal como se verifica na Figura 31, a alteração do contexto inicial da palavra não parece alterar a taxa de acerto, ou seja, os ouvintes continuam a ter maior taxa de acerto para VE apesar do contexto inicial ser diferente. Este dado parece estar de acordo com o que foi referido anteriormente relativamente à percepção de palavras iniciadas por oclusivas e por fricativas.

Até este momento, parece manter-se a tendência para a VE ser a voz mais perceptível para o grupo de ouvintes seleccionados, mesmo que seja alterado o contexto inicial do estímulo produzido. Verifiquemos se esta situação se mantém com as

diferentes vogais analisadas. Na figura 32, observam-se as taxas de acerto para cada um dos grupos de falantes, por vogal com contextos consonânticos iniciais diferentes.

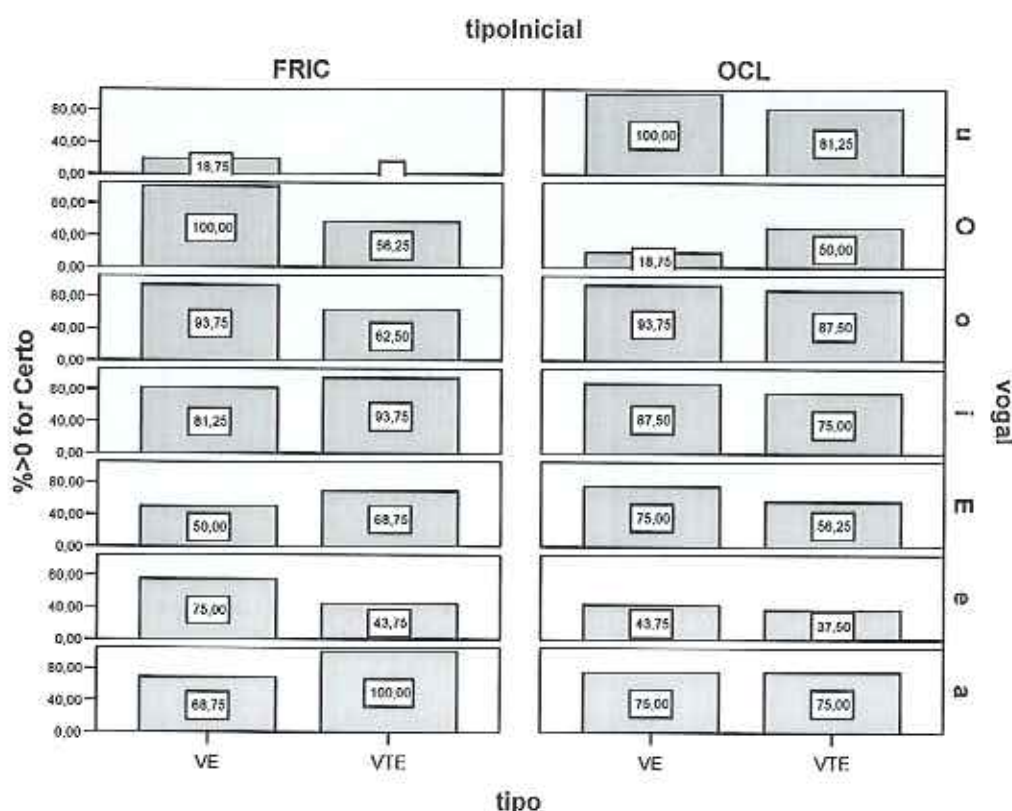


Figura 32 – Taxas de acerto em cada um dos grupos de laringectomizados por vogal em diferentes contextos iniciais

Na Figura 32, verifica-se que a percentagem de acerto dos ouvintes varia conforme a vogal apresentada. Assim, se dividirmos por contexto consonântico, podemos inferir que, em contexto inicial com fricativa, a maior taxa de acertos ocorre na VE para as vogais [O], [o], e [c]. As vogais [i], [E] e [a] tiveram uma taxa de acerto superior na VTE, sendo que a vogal [a] teve uma taxa de acerto de 100%. A vogal [u] teve uma taxa de acerto nula para a VTE, mas obteve taxas de acerto muito elevadas quando se inicia o contexto por oclusiva, para os dois grupos. Se o contexto consonântico for iniciado por oclusiva verificamos que, a maior taxa de acertos se verifica para VE em todas as vogais, excepto para a vogal [O] e para a vogal [a] cuja taxa de acertos foi igual para os dois grupos.

Observemos então, pela análise da figura 33, os estímulos que obtiveram melhores taxas de acerto para os dois grupos de laringectomizados.

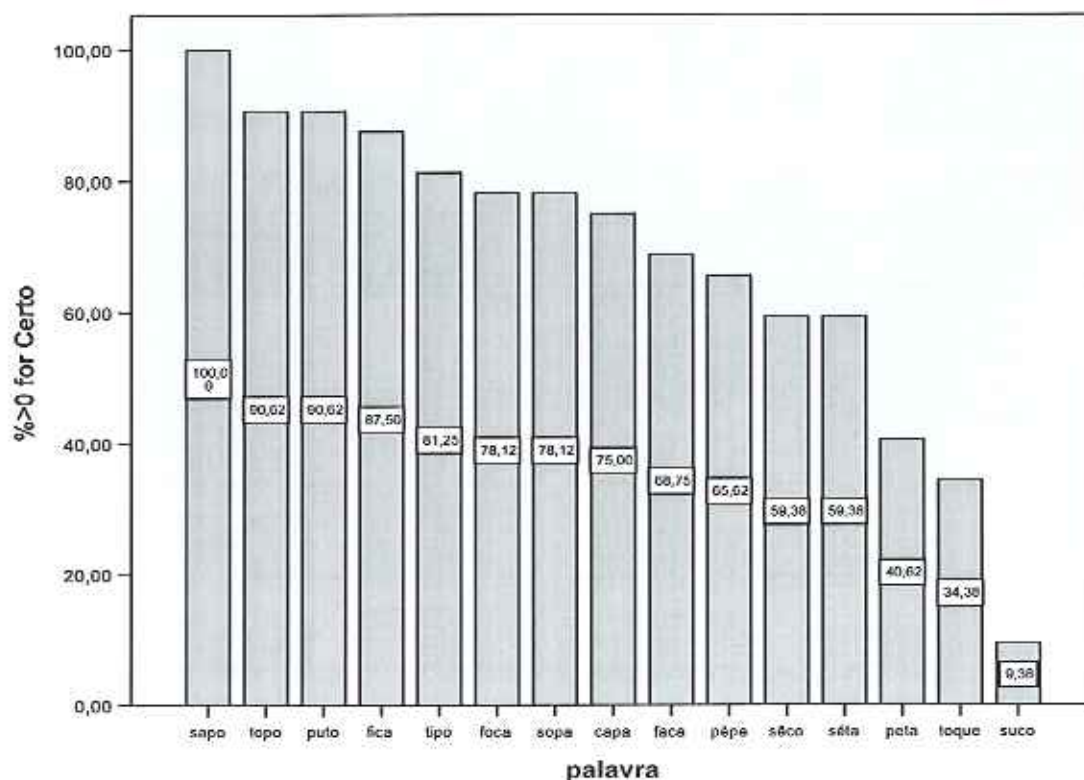


Figura 33 – Percentagem de acertos por palavra nos dois grupos de Laringectomizados

Nesta figura, verificamos que o único estímulo com uma taxa total de acertos foi a palavra [sapo] e que a grande maioria das palavras teve uma percentagem superior a 50%. Podemos também verificar que as palavras [peta, toque e suco] foram as que obtiveram valores percentuais de acerto mais baixos. Se compararmos estes dados com os da figura 32, podemos verificar que a vogal [e] presente na palavra [peta], teve taxas de acerto inferiores a 50%, excepto para VTE com contexto inicial com fricativa. A vogal [u] tem uma taxa de acerto de 9,38% para a palavra [suco] e uma taxa de 90,62% na palavra [puto]. Perante esta situação podemos considerar que a palavra [suco] por apresentar duas sílabas com a vogal final [u] se torna menos perceptível para os ouvintes. Esta palavra deverá ser retirada do *corpus* em trabalhos a ser desenvolvidos no futuro.

6.3.2 Resultados do Teste de Avaliação da Qualidade Vocal

O Teste de Avaliação da Qualidade Vocal, como já foi referido anteriormente, avalia de forma subjectiva a qualidade vocal dos indivíduos utilizadores de VE e VTE, através da apresentação de 5 categorias de resposta. Com este teste pretendia-se verificar qual a voz que apresentava melhor qualidade e se os avaliadores experientes seriam mais exigentes na avaliação deste tipo de voz.

Comecemos por verificar, na Figura 34, qual das vozes obteve melhores índices de qualidade vocal, para os dois grupos de avaliadores.

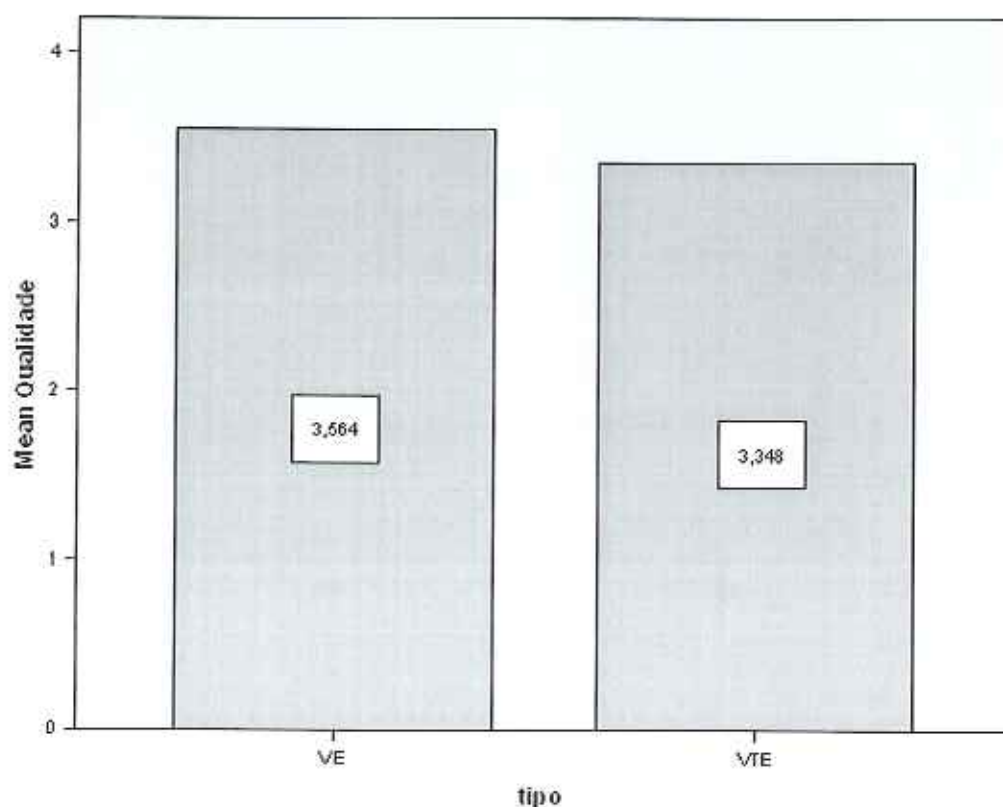


Figura 34 – Valores médios de qualidade vocal para os dois grupos de indivíduos

Na Figura 34, podemos observar que o valor médio de qualidade obtido no teste avaliação da qualidade vocal é superior na VE, sendo este resultado concordante com o teste de identificação dos estímulos que apresenta taxas de acerto superiores para a VE. Contudo, como o teste apresenta valores bastante próximos, verificamos se a diferença é estatisticamente significativa, aplicando, neste caso, um teste-t para a igualdade das médias. Verificou-se que, na realidade, a diferença é estatisticamente significativa uma

vez que $[t(446) = 3.025, p = 0,003]$, o que significa que os ouvintes consideram a qualidade da VE superior à VTE.

De forma a respondermos à segunda questão levantada com a utilização deste teste, relativa ao grau de exigência dos avaliadores, analisemos a Figura 35.

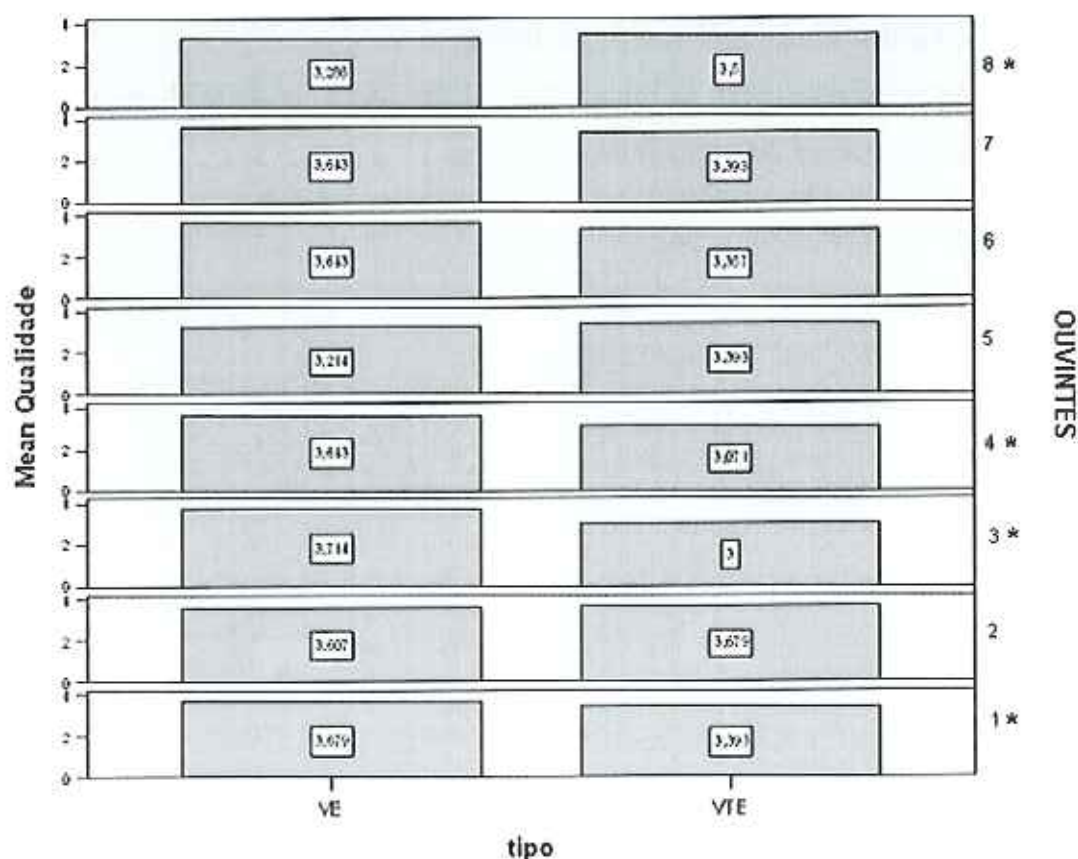


Figura 35 – Grau de exigência dos avaliadores relativo aos estímulos ouvidos. Os avaliadores experientes estão assinalados com o símbolo *.

Pela análise da figura 35, verificamos que os resultados obtidos neste teste variam entre os níveis 3 e 4, ou seja, as duas vozes têm uma qualidade que varia entre o razoável e o bom. Também podemos concluir que dos elementos pertencentes ao grupo experiente (1,3,4,8) apenas 1 considera a VTE com qualidade superior. No grupo não experiente, verificamos que dois elementos preferem a VE e outros dois a VTE. Este resultado aponta para uma indecisão neste grupo, pelas dificuldades encontradas na avaliação da qualidade vocal destes elementos. Podemos também verificar que, de uma forma geral, os elementos pertencentes ao grupo mais experiente obtêm valores médios de avaliação da qualidade mais baixos do que o grupo não experiente. Esta situação ocorre pelo facto

deste grupo de avaliadores ser mais exigente na sua avaliação subjectiva, que decorre da sua experiência em comunicar de forma diária com este tipo de utentes.

De forma sumária, podemos inferir que a VE, neste estudo, apresenta melhor qualidade vocal e que o grupo de avaliadores experientes é mais exigente no seu julgamento relativamente ao grupo não experiente. Se esta última conclusão parece estar de acordo com a literatura consultada, já o parâmetro relativo à voz com melhor qualidade difere do que foi consultado. Será necessária, sem dúvida, uma análise mais aprofundada, aumentando o número de falantes utilizadores dos dois tipos de voz.

Capítulo 7 - Conclusões e Trabalho Futuro

Iniciaremos este capítulo com um resumo do trabalho realizado, em seguida abordaremos os principais resultados e conclusões e, finalmente, abordaremos as limitações relativas encontradas e referiremos os possíveis desenvolvimentos posteriores deste trabalho.

7.1 Resumo do trabalho realizado

A voz é um instrumento precioso de comunicação no mundo que nos rodeia. Ela transmite os nossos sentimentos e emoções e tem um papel preponderante ao longo da nossa existência.

Existem diversas patologias que podem causar alterações da voz, mas nenhuma é tão devastadora como o carcinoma laríngeo, que pode levar à realização de uma Laringectomia Total. Este procedimento causa a perda imediata da voz e traz imensas alterações para os indivíduos que a realizam. É um processo lento, complicado e com muitas etapas difíceis a superar, desde o conhecimento do problema em si até à reabilitação.

A forma corajosa como estes doentes conseguem ultrapassar as fases do processo e a força interior que revelam, levou à realização deste trabalho.

Conhecendo-se os diferentes métodos de reabilitação terapêutica destes doentes e tomando como base alguns artigos e livros publicados sobre o tema, chegou-se à conclusão que, em Portugal, não havia nenhum trabalho publicado sobre a análise vocal destes doentes e que a maior parte das análises feitas aos métodos eram puramente subjectivas.

O presente trabalho iniciou-se com a escolha dos métodos de intervenção a comparar e optou-se pelo estudo da voz esofágica e traqueo-esofágica, que seriam aquelas que maior número de utilizadores reunia. Procedeu-se então, à selecção do *corpus*, com a escolha de palavras dissilábicas em contexto CVCV, por ser o tipo de construção mais frequente da língua portuguesa. Propôs-se a análise da vogal que se encontra na primeira sílaba, replicando desta forma o estudo de Cervera et al (2001). As vogais são os sons que maior número de informações acústicas transmitem e a escolha da primeira vogal para análise relaciona-se, sobretudo, com o facto de ser a mais marcada na utilização deste tipo de voz.

Encontrado o *corpus*, procedeu-se à gravação dos informantes (9 no total), todos sócios da API.V, que aceitaram participar, voluntariamente, neste estudo. Após as gravações, fez-se a anotação das palavras e fones no programa *SFS* de onde se extraíram, usando *Praat*, os parâmetros a analisar. Posteriormente, procedeu-se à análise estatística dos dados.

Após a análise dos resultados, percebeu-se que os parâmetros analisados não eram muito diferentes para os dois grupos. Optou-se então pela realização um teste de avaliação perceptiva e da qualidade vocal, seleccionando-se 1 dos elementos de cada um dos grupos de utilizadores de voz não-laríngea. Reuniu-se um grupo de ouvintes treinados com estes dois tipos de voz e um grupo *naive* (sem contacto com este tipo de voz) e propôs-se que identificassem cada uma das palavras da lista produzida pelos falantes e, de seguida, classificassem a produção de voz quanto à sua qualidade.

7.2 Principais resultados e conclusões

O principal resultado deste trabalho foi o seu carácter pioneiro em estudos de VE e VTE efectuados para o PE. Existe, neste momento, um estudo acústico de voz para uma população que até agora só era referida e avaliada de forma subjectiva, ou seja, existem dados (que poderão obviamente, ser melhorados) quantitativos e objectivos que poderão permitirão a todos os que lidam com este tipo de utentes, reflectir melhor nos seus processos de decisão terapêutica.

Há, portanto, com este trabalho, também uma avaliação perceptiva e da qualidade vocal que, devendo ser alargada, nos coloca perante dados que exigem uma reflexão madura.

Uma das principais conclusões deste trabalho, está relacionada com o facto de não nos ter sido possível decidir qual o método de reabilitação mais eficaz, concluindo-se que, cada caso é um caso e que em Portugal ainda estamos a dar os primeiros passos até podermos decidir (juntamente com o utente) qual o método de intervenção que devemos aplicar. Os métodos não servem de igual modo para os doentes e há situações que devem ser muito bem ponderadas antes de se optar por um ou por outro. Contudo, em trabalhos como o de Corinna (2001) e Cervera et al (2001), não parece haver grandes dúvidas quanto ao facto de o método mais eficaz a utilizar ser, a VTL.

7.3 Limitações do estudo

A maior limitação deste estudo está relacionada com o número reduzido de informantes analisados, o que dificultou a obtenção de conclusões mais efectivas.

Outra limitação prende-se com o facto de não havendo estudos fisiológicos prévios relativos ao tamanho do tracto vocal dos utentes laringectomizados, à forma como posicionam a língua durante a produção de palavras e à localização exacta do segmento faringo-esofágico, não ser possível interpretar determinados valores que daqui resultaram.

O Teste de Percepção realizado deveria ter maior número de informantes e de ouvintes, sendo que a amostra destes últimos deveria ser mais homogénea. Por último, a dificuldade em encontrar estudos nacionais ou internacionais que se propusessem a analisar os parâmetros deste trabalho, tornou algumas das conclusões meras suposições que necessitam de confirmação urgente.

7.4 Desenvolvimentos Futuros

Sabemos que este estudo foi pioneiro em Portugal e que servirá como ponto de partida para um estudo mais alargado, com maior número de informantes utilizadores de VTE e VE e com um *corpus* mais abrangente que permita verificar o comportamento das vogais noutros contextos linguísticos. O trabalho mais importante a desenvolver consistirá na captação de imagem, através de Ressonância Magnética ou Videofluoroscopia, para o estudo das medidas do tracto vocal dos utentes laringectomizados, do posicionamento da língua durante a produção de sons isolados e palavras e de todo o comportamento fisiológico da região da cabeça e pescoço durante o momento da fonação. Depois de recolhidos estes dados e comparados com os normofalantes, já será possível a retirada de conclusões mais exactas.

Bibliografia

AKISKAL, H. S.; [et al.] Manual Merck: Saúde para a família [em linha]. Brasil – Merck Sharp & Dohme (revisto em 2006) [citado em de 2006 -]. Disponível em URL: http://www.msd-brazil.com/msd43/m_manual/nota_ao_leitor.htm

BAKEN & ORLIKOFF. Clinical Measurement of Speech and Voice. 2nd Edition. Singular Publishing Group, Inc, 2000.

BEHLAU M, PONTES P. Avaliação e Tratamento das Disfonias. São Paulo: Lovise, 1995.

BEHLAU, Mara. Voz: O livro do Especialista – Volume I. Rio de Janeiro: Livraria e Editora Revinter Ltda, 2001.

BEHLAU, Mara. Voz: O livro do Especialista – Volume II. Rio de Janeiro: Livraria e Editora Revinter Ltda, 2005.

CASPER, J.K and COLTON R.II. Clinical Manual for Laryngectomy and Head/neck Cancer Rehabilitation. 2^a Ed. London: Singular Publishing Group, Inc, 1998.

CASPER, Janina K.; COLTON, Raymond H. Compreendendo os Problemas da Voz: Uma Perspectiva Fisiológica ao Diagnóstico e ao Tratamento. Porto Alegre: Editora Artes Médicas.

CLOSE, L.G.; LARSON, D.I., and SHAI, J. P. Essentials of Head and Neck Oncology. New York: Thieme Medical Publishers, Inc., 1998.

CLOSE, Lanny Grath; LARSON David L.; SHAH Jatin P.. Essentials of Head and Neck Oncology. Nova Iorque: Thieme Medical Publishers, Inc., 1998.

CORINNA J. Van As. Trachoesophageal Speech: A Multidimensional Assesment of Voice Quality. Institute of Phonctic Sciences of the University of Amsterdam, 2001.

COSTA, Claudiney C.; [et al.] Revista Brasileira de Otorrinolaringologia [em linha]. São Paulo – Revista Brasileira de Otorrinolaringologia São Paulo – Brasil, Setembro 2001 (revisto em 2006) [citado em de 2006 -]. Disponível em URL: <http://www.rborl.org.br/conteudo/accrvo/accrvo.asp?id=2526>

COSTA, Claudiney C.; [et al.] Revista Brasileira de Otorrinolaringologia [em linha]. São Paulo – Revista Brasileira de Otorrinolaringologia São Paulo – Brasil, Setembro 2001 (revisto em 2006) [citado em Abril de 2006]. Disponível em URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-72992001000500017

COUTINHO, Miguel. Diário de Notícias: Sociedade [em linha]. Lisboa: Diário de Notícias, 17 de Abril de 2005. (sem data de revisão) [citado em -- de 2006]. Disponível em URL: http://dn.sapo.pt/2005/04/17/sociedade/400_vitimas_cancro_laringe_ano.html

DENES, Peter B., PINSON, Elliot N.. The Speech Chain: The Physics and Biology of Spoken Language, 2ª Edição. Nova Iorque: W. H. Freeman Company, 1993.

DIKEMAN, Karen J.; KAZANDJIAN, Marta S.. Communication and Swallowing Management of Tracheostomized and Ventilator-Dependent Adults. 2ª Edição. Nova Iorque: Delmar Learning, 2003.

GUIMARÃES, I. A Ciência e a Arte da Voz Humana. Escola Superior de Saúde do Alcoitão, Santa Casa da Misericórdia de Lisboa, 2007.

HARRINGTON, J., CASSIDY, S. Techniques in Speech Acoustics. Kluwer Academy Publishers, 1999

HIRANO, M.; [et al]. Acoustic analysis of pathological voice. Acta Otolaryngologica, 105 (1988), pp 432-438.

HOSPITAL UNIVERSITARIO VIRGEN DE LA ARRIZACA. El portal sanitario de la Región de Murcia: Guía del Laringectomizado (em linha). Murcia Salud: Hospital

Universitario Virgen de la Arrixaca [citado em de 2006]. Disponível em URL:
<http://www.murciasalud.es/pagina.php?id=8232&idsec=45>

INCALUQUE SUCA, Jesus Circilo. Incidencia do câncer da laringe, en el servicio de hospitalización de cirugía de cabeza, cuello y maxilofacial del hospital Guillermo Almenara Irigoyen, EsSalud. Lima-Perú; Junio de 1999 – Mayo de 2002 [em linha]. Lima – Perú: Universidad Nacional Mayor de San Marcos, 2003 [citado de 2006 -]. Disponível em URL:
http://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Salud/incaluque_sj/incaluque_sj.pdf

JORGE, Mariana Silva; GREGIO, Fabiana Nogueira; CAMARGO, Zuleica - Qualidade Vocal de Indivíduos Submetidos a Laringectomia Total – Aspectos Acústicos de Curto e de Longo Termo em Modalidades de Fonação Esofágica e Traqueoesofágica, “Revista CEFAC”. São Paulo: 6:3 (2004), pp. 319 – 328.

KENT, Ray D., READ, Charles. Acoustic Analysis of Speech. 2ª Edição. Canada: Singular (2001).

L. P. FERREIRA et al, Tratado de Fonoaudiologia, Roca Ltda, São Paulo, 2004.

LADEFOGED, P. Vowels and Consonants: an Introduction to the Sounds of Languages. Oxford: Blackwell, 2001.

M. BEHLAU, Voz – O Livro do Especialista, Volume II, Revinter; São Paulo, 2004.

MARTINS, Paula. Vogais do Português Europeu, Aveiro, 2007

MORENTE, Juan Carlos Casado; TORRES, José Antonio Adrián. Manual del laringectomizado. Málaga: Ediciones Aljibe, S.L., 2004.

Mateus M., Freitas I. e Falcó M. Fonética e Fonologia do Português, Universidade Aberta, 2005

MOUTINHO, L.C. Uma Introdução à Fonética e Fonologia do Português. Plátano Edições Técnicas, Lda., 2000.

NETTER, - Larynx, Clinical Symposia: Ciba Pharmaceutical Company, 1964 16(3): Plate VI.

OLIVEIRA, Iára Bittante de; [et al.] Comunicação oral de laringectomizados com prótese traqueoesofágica: análise comparativa pré e pós-treino [em linha]. Brasil – Pró-Fono Revista de Atualização Científica, 2005 [citado em de 2006]. Disponível em URL: <http://www.scielo.br/pdf/pfono/v17n2/v17n2a04.pdf>

PETRA, Jongmans ; [et al.]. An Introduction to the Assessment of Intelligibility of Tracheoesophageal Speech [em linha]. Amsterdão: Institute of Phonetic Sciences, University of Amsterdam, 2003 [citado em de 2006 -]. Disponível em URL: <http://darc.uva.nl/document/7161>

PINHO, S. M. Fundamentos em Fonoaudiologia: Tratando os distúrbios da Voz, Guanabara Koogan S.A., Rio de Janeiro, 2003.

PINTO, José António; [et al.]. Revista Brasileira de Otorrinolaringologia [em linha]. São Paulo – Revista Brasileira de Otorrinolaringologia São Paulo –Brasil, Junho 2005 [citado em de 2006 -]. Disponível em URL: http://www.rborl.org.br/suplementos/detalhes_debates.asp?id=112

PINTO MOURA, C. Alterações de Voz em Crianças Portadoras de Trissomia 21. Porto, 2006.

ROSIQUE, M.; [et al.]. Análisis discriminante aplicado a los formantes de las vocales castellanas en la fonación con prótesis y erigmofonía tras laringectomía total «Acta Otorrinolaringol Esp». Murcia 54 (2003) pp. 361-366

STAJNER-KATUSIC, Smiljka; [et al.] – Voice and speech after laryngectomy «Clinical Linguistics & Phonetics». Taylor & Francis ISSN 0269-9206. (2006) pp. 195-203

T. CERVERA, J. L. Mirales and J. Gonzales-Alvarez, "Acoustical Analysis of Spanish Vowels Produced by Laryngectomized Subjectd", *Journal of Speech, Language and Hearing Research*, Vol. 44 (2001) pp.988-996.

TEIXEIRA, António; [et al.]. *Fonética Aplicada ao Processamento de Fala: As Nasais do Português [em linha]*. Aveiro: Universidade de Aveiro, 2003 [citado em 2006]. Disponível em URI.: http://www.ii.ua.pt/cidlc/gcl/files/publicacoes/FAPLI_2003.doc

VAN AS, Corina J.; [et al.] - Perceptual Evaluation of Tracheoesophageal Speech by Naive and Experienced Judges Through the Use of Semantic Differential Scales «*Journal of Speech, Language, and Hearing Research*» 46 (2003) pp. 947-959

VAN AS, Corina; [et al.] -- Acoustical Analysis and Perceptual Evaluation of Tracheoesophageal Prosthetic Voice «*Journal of Voice*». Singular Publishing Group, Inc. 12:2 (1998) pp. 239-248.

VAN DEN HOOGEN, F. J. A. - A prospective study of speech and voice rehabilitation after total laryngectomy with the low-resistance Groningen, Nijdam and Provox voice prostheses. «*Clinical Otolaryngology*». Blackwell Science Ltd. 23 (1998) pp. 425-431.

VÁZQUEZ DE LA IGLESIA F ; [et al]. *Voz Esofágica*. Revista Medica Universidade de Navarra. Vol 50, Nº 3 (2006) pp 56-64.

ZEMLIN, W. *Princípios de Anatomia e Fisiologia em Fonoaudiologia*. 4ª ed, Porto Alegre, 2000.