FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

MESTRADO EM ENGENHARIA BIOMÉDICA

EM DIRECÇÃO A UMA LARINGE ARTIFICIAL ELECTRÓNICA – FUNDAMENTOS TÉCNICO-CIENTÍFICOS E ENSAIOS PRELIMINARES

Dissertação

Luis Ricardo Monteiro Jacinto

Dezembro 2007

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
MESTRADO EM ENGENHARIA BIOMÉDICA

EM DIRECÇÃO A UMA LARINGE ARTIFICIAL ELECTRÓNICA – FUNDAMENTOS TÉCNICO-CIENTÍFICOS E ENSAIOS PRELIMINARES

Luis Ricardo Monteiro Jacinto
Licenciado em Biologia
pela Faculdade de Ciências da Universidade do Porto

Dissertação submetida para satisfação parcial dos requisitos do grau de Mestre em Engenharia Biomédica (área de especialização de Sinais e Imagens Médicas)

Dissertação realizada sob supervisão de
Prof. Doutor Diamantino Freitas (orientador)
do Departamento de Engenharia Electrónica e de Computadores
da Faculdade de Engenharia da Universidade do Porto

Prof. Doutora Carla Moura (co-orientador)
do Serviço de Otorrinolaringologia
do Hospital de São João, EPE / Faculdade de Medicina da Universidade do Porto

RESUMO

A remoção da laringe por laringectomia total – devido a cancro ou traumatismo – impede a produção de fala nos humanos. A restauração da capacidade de comunicação falada é crítica para a prevenção de consequências psicológicas, sociais e económicas negativas. O estado presente do desenvolvimento das electro-laringes, apesar de permitir um restabelecimento da comunicação com os outros, com mais ou menos dificuldade de cada uma das partes, deixa ainda muito a desejar, nomeadamente quanto à qualidade do som produzido, à funcionalidade para expressar entoação e acentuação, e à sua forma de manuseamento. Perceber as correntes limitações, estudar os avanços no domínio das electro-laringes mas também do processamento da fala e outras áreas complementares relevantes, indicar possíveis caminhos de investigação futuros bem como analisar e demonstrar certas metodologias passíveis de ser aplicadas e implementadas são objectivos obrigatórios numa fase inicial de desenvolvimento de uma electro-laringe que pretenda ultrapassar, pelo menos, alguns dos defeitos actuais. Focou-se, então, esta dissertação na obtenção e análise de um sinal glotal, de qualidade e passível de parametrização, obtido em conjunto com o estudo acústico do tracto vocal. O sinal glotal foi obtido através de um método de estimação conjunta da fonte e tracto vocais, na qual se aplicaram estratégias de optimização com recurso ao filtro de Kalman e a um algoritmo de solidificação simulada, como parte de um modelo auto-regressivo com entrada exógena, ARX, do processo de produção da fala, onde o modelo glotal RK serve como entrada. O tracto vocal foi reconstruído tridimensionalmente a partir de imagens de ressonância magnética, RM, existentes, de forma a permitir a produção de um modelo sólido em silicone, por prototipagem rápida, que se aproxima do tracto vocal real em termos anatómicos e acústicos, passível de ser utilizado em ensaios acústicos de análise de fontes vocais e caracterização de falantes. Mostrou-se que é possível obter boas estimativas da fonte vocal de forma parametrizada com recurso a um modelo glotal e à sua integração em procedimentos de estimação conjunta da fonte e do tracto vocais. Os modelos RK estimados mostraram ser uma boa aproximação da fonte glotal real. A igualização, através da determinação de filtros que simulam o sistema inverso, de um altifalante comum, com uma resposta em frequência adequada, permitiu o acoplamento do mesmo à entrada do tracto vocal sólido e consequente realização dos ensaios de demonstração acústicos pela aplicação do sinal RK estimado e pré-filtrado, ao altifalante. O tracto vocal sólido mostrou ser uma boa aproximação do tracto vocal real e permitiu a recuperação dos formantes existentes na fala real do sujeito que lhe deu origem, a menos de um ligeiro desvio no valor do quarto formante. As razões de tal erro não são ainda claras, mas deverão estar relacionada com erros introduzidos durante o processo de construção do modelo 3D. Os testes perceptuais subjectivos mostraram um sinal, produzido pela estimativa da fonte e sua aplicação ao modelo sólido do tracto vocal, razoavelmente próximo de um sinal de fala real e uma identificação perceptual inequívoca da vogal do português europeu que se pretendia sintetizar.

ABSTRACT

The removal of the voice organ, the larynx, due to cancer or trauma, eliminates the possibility of speech production in humans. The rapid restoration of speech is critical to prevent further negative psychological, social and economic consequences. The present state of electronic artificial larynxes, though permitting the re-establishment of communication with others, with more or less difficulty on each part, is still very unsatisfactory, namely the quality of the sound produced, the ability to express intonation and stress, and its handling. Understanding the current limitations, studying the recent advances on the field of electrolarynx design but also speech processing and other complementary relevant areas, identifying possible future research paths, as well as demonstrating methodologies that can be implemented, are all key concerns on a first stage of development of a prototype that can overcome, at least some, of the present flaws. This dissertation has thus focused on obtaining and analysing a glotal signal, of good quality, in a parametric way, and studying in conjunction the acoustics of the vocal tract. The glotal signal was obtained by method of simultaneous estimation of the vocal tract and source, in which optimization strategies where applied including a Kalman filter and a simulated annealing algorithms, as part of an autorregressive with exogenous input, ARX, model of the speech production process, where a RK model of the glotal signal serves as source. The vocal tract was reconstructed in three-dimensions, 3D, from Magnetic Ressonance images, MRI, with the objective of producing a silicone solid model of the vocal tract, by rapid prototyping, that can approximate the real vocal tract anatomically and acoustically, to be used in acoustical experiments of voice source analysis and speaker characterization. It was shown that it is possible to obtain good estimates of the voice source, in a parametric way, using a glotal source model and integrating it in a source and vocal tract simultaneous estimation procedure. The RK model showed to be a good approximation of the real voice source. The equalization, through inverse system filters, of consumer speakers, with the adequate frequency response, allowed their coupling to the vocal tract solid model's input orifice and consequent acoustical experiment. On this, the estimated RK signal was initially pre-filtered and then .fed to the speakers. The solid vocal tract showed to be a good approximation of the real vocal tract, allowing the recovery of the first three formants, despite a small frequency shift, present in the real speech obtained from the speaker that was the model for the silicone vocal tract. The higher frequency formant was not accurately recovered. The causes of such error are not yet clear but should be related to errors introduced in the process of 3D modelling. Subjective perceptual tests have shown that the signal produced by the silicone model of the vocal tract excited by the estimated glotal source was reasonably close to the real speech signal and that there was a positive identification of the European Portuguese vowel meant to be synthesized.

PREFÁCIO

A voz humana é a fundação base da auto-expressão e da comunicação com os outros. A remoção da laringe por laringectomia total – devido a cancro ou traumatismo – afecta profundamente a fisiologia da respiração e da deglutição bem como a produção da fala nos humanos. A perda de produção de som glótico é a principal restrição que os doentes adquirem após a cirurgia. Assim, a restauração efectiva da produção de fala é crítica para a prevenção de consequências negativa de ordem psicológica, social e económicas.

Presentemente, apesar das as electro-laringes permitirem um restabelecimento da comunicação com os outros, com mais ou menos dificuldade de cada uma das partes, apresentam ainda algumas limitações, nomeadamente quanto à qualidade do som produzido, à funcionalidade para expressar entoação e acentuação, e à sua forma de utilização manual.

Pretende-se com a presente dissertação estudar as fundações técnico-científicas para o desenvolvimento de uma electro-laringe que venha a superar as limitações das electro-laringes actuais comercialmente disponíveis, servindo também como estudo preparatório de um projecto que se pretende continuar e aprofundar

Perceber os defeitos correntes, estudar os avanços no domínio das electro-laringes mas também do processamento da fala e outras áreas complementares relevantes, analisar e demonstrar certas metodologias passíveis de ser aplicadas e implementadas, bem como indicar possíveis caminhos de investigação futuros são os objectivos obrigatórios numa fase inicial de desenvolvimento de uma electro-laringe que pretenda ultrapassar algumas das limitações actuais.

Os principais objectivos da dissertação, nesta primeira fase, foram a obtenção e análise do sinal glotal, de qualidade e passível de parametrização, através da implementação de um algoritmo de identificação conjunta da fonte e tracto vocais. Para tal recorreu-se a técnicas avançadas de estimação e optimização como o filtro de Kalman e a solidificação simulada; bem como o estudo acústico do tracto vocal, através

da construção de um modelo tridimensional do mesmo a partir de grupos de imagens existentes, obtidas por ressonância magnética, RM, com o intuito de produzir um modelo sólido, nesta fase construído em silicone, que se aproxime das características acústicas do tracto vocal real.

AGRADECIMENTOS

O autor quer agradecer a todos aqueles que, de uma forma ou outra, tornaram esta dissertação possível, especialmente: Avelino Marques, Carla Pinto Moura, Cláudia Falcão Reis, Diamantino Freitas, Ema Monteiro, João Canossa, José Pedro Fernandes, Luís Jesus, Maria da Silva Ferreira, Pedro Marques, Leandro Ribeiro, Rui César e Sandra Rua.

ÍNDICE

ÍNDICE DE FIGURAS

LISTA DE ABREVIATURAS E SÍMBOLOS

LISTA DE ABREVIATURAS

3D	Três-Dimensões
ACS	American Câncer Society
AMT	Limiar de Mascaramento Auditório (*Auditory Masking Threshold*)
APLV	Associção Portuguesa dos Limitados pela Voz
AR	Autorregressivo
ARMA	Média Movemente Autorregressiva
ARMAX	Média Movemente Autorregressiva com Entrada Exógena
ARX	Autorregressivo com Entrada Exógena
ASCII	*American Standard Code for Information Interchange*
CAD	Desenho Assistido por Computador (*Computer Assisted Design*)
EGG	Electroglotografia
EL	Electro-laringe
EMA	Articulografia Electromagnética
EMG	Electromiografia
EMGG	Electroglotografia Magnética
EPG	Electropalatografia
EUA	Estados Unidos da América
DAP	Modelação Discreta só com Pólos (*Discrete All-Pole Modelling*)
DEGG	Derivada do Electroglotograma
FFT	Transformada Rápida de Fourier (*Fast Fourier Transform*)
FIR	Resposta Impulsional Finita (*Finite Impulse Response*)
GAR	*Glotal ARX*
GARMAX	*Glottal ARMAX*
GCI	Instante de Fecho Glotal (*Glottal Closure Instant*)
GOI	Instante de Abertura Glotal (*Glottal Opening Instant*)
IF	Infravermelho
IFFT	Transformada Rápida de Fourier Inversa (*Inverse Fast Fourier Transform*)
IIR	Resposta Impulsional Infinita (*Infinite Impulse Response*)
IPA	Alfabético Fonético Internacional
IPO	Instituto Português de Oncologia
LED	Díodo Emissor de Luz (*Light Emitting Diode*)
LPC	Codificação por Predição Linear (*Linear Prediction Coding*)
MA	Média Movente
MFA	Análise Multi-Frame (*Multi-Frame Analysis*)
MIS	Sistema de Identificação de Modelo (*Model Identification System*)
OPG	Optopalatografia

OQ	Quociente de Abertura (*Open Quotient*)
P0	Período Fundamental
RF	Rádio-Frequência
RM	Ressonância Magnética
ROI	Região de Interesse (*Region of Interest*)
SA	Solidificação Simulada
SAMPA	*Speech Assessment Methods Phonetic Alphabet*
SQ	Quociente de Velocidade (*Speed Quotient*)
SUMT	Minimização Sequencial Sem Restrições
T0	Frequência Fundamental
TAC	Tomografia Axial Computorizada
TE	Traqueoesofágica
TL	Pendente Espectral (*Spectral Tilt*)
UV	Ultra-Violeta

LISTA DE SÍMBOLOS

A	Área
a_k	Coeficientes do modelo AR
b_j	Coeficientes do modelo ARX
$C(n)$	Matriz de Observação
c	Parâmetro de Controlo
d_g	Largura da Corda Vocal
dug^{RK}	Derivada do Impulso Glotal Gerada pelo Modelo RK
$e(n)$	Sequência de Sinal de Erro
F_0	Frequência Fundamental
$K(n)$	Ganho de Kalman
k	Coeficiente de Rigidez
k_b	Constante de Boltzmann
l_g	Comprimento da Corda Vocal
m	Massa
$P(n)$	Matriz de Covariância do Vector de Estado
P_g	Pressão de Ar Glotal
p	Ordem de modelo

$Q(c)$	Constante de Normalização dependente de c
$q(n)$	Ruído do Sistema
$r(n)$	Ruído de Medição
r	Coeficiente de Atenuação
$s(n)$	Sinal de Fala
T	Temperatura
T_0	Período Fundamental
U_g	Velocidade Volumétrica Glotal
x	Deslocamento
$X(n)$	Vector de Estado
$y(n)$	Vector de Observação
$Z(T)$	Factor de Normalização
ρ	Densidade do Ar

1 INTRODUÇÃO

1.1 INTRODUÇÃO À ANATOMIA E FISIOLOGIA GERAL DO SISTEMA DE PRODUÇÃO DE FALA

A onda da fala é uma onda acústica de pressão sonora originada em movimentos voluntários de estruturas anatómicas pertencentes ao sistema de produção de fala humano. Os componentes genéricos deste sistema são os pulmões, a traqueia, a laringe, o órgão por excelência de produção da voz, a cavidade faríngea, a cavidade oral ou bucal e a cavidade nasal nariz. As cavidades faríngea e oral são geralmente agrupadas com a designação de tracto vocal, e a cavidade nasal designada de tracto nasal. Existem ainda outras estruturas que são determinantes para a produção de fala: pregas vocais, palato mole ou véu palatino, língua, dentes e lábios. Estas estruturas anatómicas, que se movem para diferentes posições de modo a produzir os vários sons da fala, são denominados articuladores, já que são responsáveis por movimentos que afectam o tamanho e a forma do tracto vocal bem como a posição dos outros articuladores – ver Figura 1.1. De acordo com este sistema, o tracto vocal inicia-se no limite superior da laringe e termina na entrada dos lábios. O tracto nasal tem início no véu palatino e termina nas narinas.

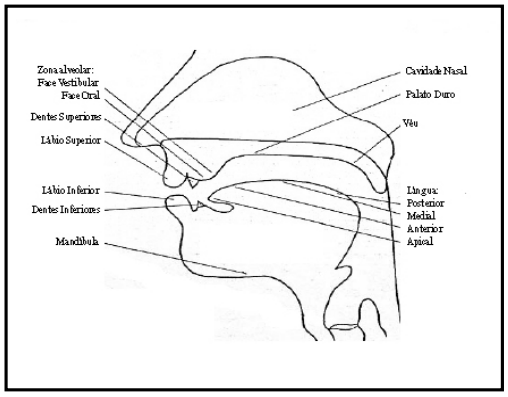

Figura 1.1 - *Diagrama simplificado de alguns articuladores; adaptado de Huang, X. et al [1].*

Considerando o sistema em termos acústicos, o tracto vocal e nasal compreendem o filtro acústico principal. O filtro é excitado pelos órgãos inferiores a este, e está carregado na sua saída principal por uma impedância de radiação devida aos lábios e às

narinas. Os articuladores, a maioria dos quais associados ao próprio filtro, são usados para alterar as propriedades do sistema, as formas de excitação e a carga apresentada à saída, ao longo do tempo.

As características espectrais da forma de onda da fala são variantes no tempo, ou não-estacionárias, já que o sistema físico se altera rapidamente ao longo do tempo.

No homem adulto médio, o comprimento do tracto vocal é de cerca 17 cm, e da mulher adulta média cerca de 14 cm. O comprimento do tracto vocal de uma criança média é de cerca de 10 cm. O posicionamento dos articuladores do tracto vocal leva a variações da área seccional do tracto vocal ao longo do seu comprimento que podem ir desde zero, estrangulamento total, até 20 cm^2. O tracto nasal constitui um caminho auxiliar para a transmissão de som. O comprimento típico do tracto nasal num homem adulto é de cerca de 12 cm. A ligação entre os tractos vocal e nasal é controlada pelo tamanho da abertura do véu palatino. Em geral, a ligação nasal pode influenciar substancialmente as características frequenciais do som radiado pelo falante. Se o palato mole estiver relaxado, permitindo a continuidade entre a faringe e a nasofaringe, o tracto nasal está acusticamente activo de modo a produzir os sons nasais da fala. A abertura velar pode variar de zero a 5 cm^2 para um homem adulto médio. Para a produção de sons não nasais, o véu é elevado em direcção à parte posterior da cavidade faríngea, selando a entrada da cavidade nasal. Deller, J. *et al* [2].

De um modo genérico, a função da laringe consiste em gerar a excitação periódica do sistema para a produção de sons denominados vozeados. Do ponto de vista anatómico e fisiológico a laringe é um órgão complexo. A descrição anatómica e funcional, mais detalhada, será realizada adiante.

Uma das principais características de qualquer som de fala é a sua forma de excitação. Os tipos de excitação podem dividir-se em dois grandes grupos: vozeados e não vozeados. No entanto, para efeitos de modelação e síntese, podem considerar-se quatro tipos de excitação: vozeado, não vozeado, misto e silêncio. Os sons vozeados são produzidos forçando o ar através da glote ou de uma abertura entre as pregas vocais. A tensão das pregas é ajustada de modo a que vibrem de uma forma oscilatória sustentada.

A interrupção periódica do fluxo de ar subglótico[1] resulta em impulsos de ar quase-periódicos que excitam o tracto vocal. O som produzido pela laringe é denominado voz ou fonação. Os sons não vozeados são gerados pela formação de uma constrição em algum ponto ao longo do tracto vocal, forçando o ar através dessa constrição para produzir turbulência. Um som que apresenta simultaneamente aspectos de vozeado e de não-vozeado é denominado som misto.

Ainda é de referir, que alguns sons da fala são compostos por um curto período de silêncio seguido por uma região de som vozeado, não-vozeado ou misto. Estes sons são denominados plosivos e formados através de uma fase de encerramento completo, normalmente próximo do final do tracto vocal, com a subsequente acumulação de pressão de ar a montante do ponto de encerramento, e posterior libertação súbita dessa pressão.

1.2 PRODUÇÃO DA FALA

A fonação é conseguida quando os músculos abdominais forçam a subida do diafragma, empurrando ar para fora dos pulmões em direcção à traqueia e, depois, para a laringe onde o fluxo é periodicamente interrompido pelas pregas vocais. O ciclo de abertura e encerramento repetido da glote ocorre em resposta à pressão subglótica proveniente da traqueia. As forças responsáveis pelo ciclo do impulso glotal afectam a forma de onda glotal e estão consequentemente relacionadas com as características espectrais correspondentes. A teoria mioelástica-aerodinâmica da fonação, Van Den Berg, J. [3], fornece a mais completa visão unificada das forças que intervêm no ciclo vibratório da glote. Esta descreve o início do ciclo pela ocorrência de uma acumulação de pressão de ar subglótica, inferiormente às verdadeiras pregas vocais, forçando a separação das mesmas. As pregas vocais começam a abrir e o ar escapa da traqueia através da glote. A pressão de ar subglótico continua a forçar o afastamento das pregas vocais do qual resulta um aumento de fluxo de ar através da glote. Considerando a noção de conservação de energia, a energia cinética é representada como o quadrado da velocidade do ar, enquanto que a energia potencial é proporcional à pressão do ar. À

[1] O termo *glótico* foi utilizado, nesta dissertação, em referência ao sistema anatómico e aos fluxo e pressão de ar. Quando se designa o sistema acústico e temporal utiliza-se a designação *glotal*.

medida que as pregas vocais se afastam, a velocidade do ar aumenta significativamente através da glote, o que causa uma queda de pressão no ar local. Assim, quando as pregas vocais estão fechadas, a pressão do ar e a energia potencial são elevadas. À medida que a glote abre, a velocidade do ar e energia cinética aumenta, enquanto que a pressão e a energia potencial diminuem. A glote continua a abrir até a tensão elástica natural das pregas vocais igualar a força separadora da pressão aérea. Neste momento, a abertura glótica e a taxa de fluxo de ar atingem o seu máximo. A energia cinética que foi recebida pelas pregas vocais durante a abertura fica armazenada como energia elástica que, por sua vez, leva a que as pregas vocais comecem a fechar. Devido à força de restauro elástico, o movimento de encerramento das pregas vocais adquire quantidade de movimento e ocorre um efeito de sucção causado pela força de Bernoulli quando a glote se tornou suficientemente estreita. Tanto a força de restauro elástico como a força de Bernoulli actuam para fechar as pregas vocais abruptamente. A pressão subglótica e as forças de restauro elástico durante o encerramento da glote levam o ciclo a repetir-se. A Figura 1.2 mostra esquematicamente este ciclo.

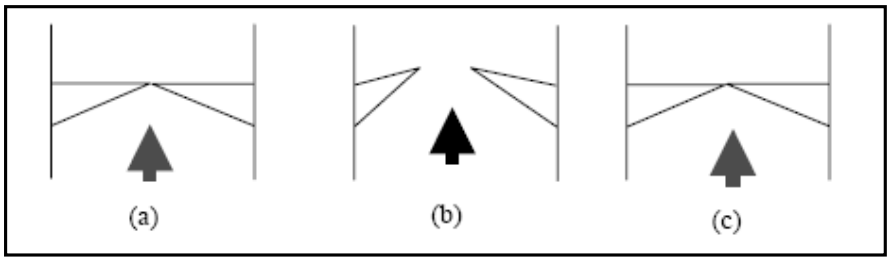

Figura 1.2 - *Resumo esquemático do ciclo vibratório da laringe. (a) Laringe fechada e acumulação de pressão sub-glótica; (b) afastamento das pregas vocais e de fluxo de ar através da glote; (c) encerramento temporário da laringe por igualização da pressão e força de restauro elástica. De Huang, X. et al [1].*

O tempo decorrido entre aberturas sucessivas das pregas vocais é designado por período fundamental, T_0, enquanto que a respectiva taxa de vibração é denominada frequência fundamental, $F_0 (= 1/T_0)$. Por vezes, o termo tom é utilizado como sendo equivalente à frequência fundamental. Mas este é, na verdade, definido como a percepção psicoacústica da frequência fundamental de um som, quer esta esteja presente ou não na forma de onda.

Algumas das características temporais do ciclo glotal são descritas com detalhe mais adiante.

1.3 PERDA DA FUNÇÃO DE FALA / REABILITAÇÃO DA FALA

1.3.1 LARINGECTOMIA E CANCRO DA LARINGE

Entende-se por laringectomia total a remoção cirúrgica da laringe, incluindo os dois primeiros anéis da traqueia, inferiormente, e do osso hióide, superiormente. Consequentemente, estabelece-se uma separação entre o tracto respiratório e o tracto vocal. Tal como mostra a Figura 1.3, o laringectomizado respira por intermédio de um traqueostoma – um orifício criado pela ligação da traqueia com à pele da região anterior do pescoço.

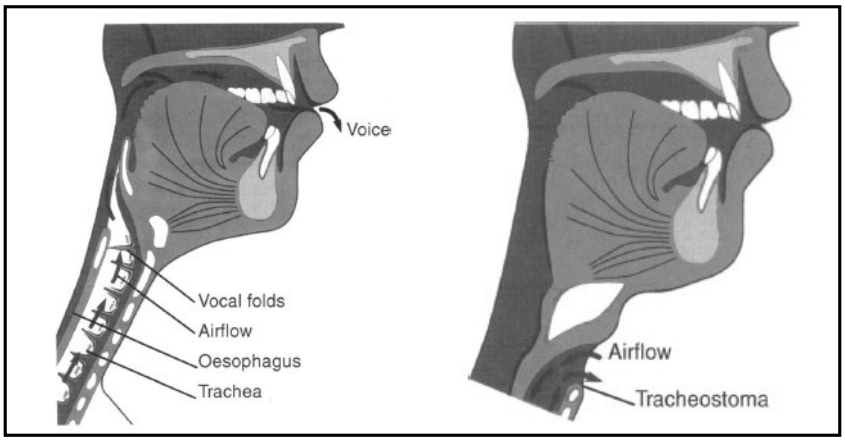

Figura 1.3 - *Desenho esquemático da anatomia antes da laringectomia total – esquerda – e após – direita. De Jassar, P. et al [4].*

Apesar da laringectomia poder ter indicação após um traumatismo, cervical, a esmagadora maioria das indicações para esta cirurgia são as vítimas de cancro da laringe. Segundo a American Cancer Society, ACS, [5], a maioria dos cancros da laringe desenvolve-se a partir de células escamosas estratificadas que formam o epitélio da mucosa. A maior parte dos cancros da laringe tem início em condições pré-cancerígenas denominadas displasias. Note-se, contudo, que nem todas as displasias evoluem para carcinoma. Mais raramente, o cancro da laringe pode ter origem em glândulas adenocarcinomas, distinguindo-se dos carcinomas de células escamosas pela constituição e organização celular. Ainda mais raros, são os cancros da laringe que têm origem nos tecidos conjuntivos ou cartilagens, originando os condrossomas e sarcomas sinoviais. A Figura 1.4 mostra uma laringe normal e outra afectada por cancro.

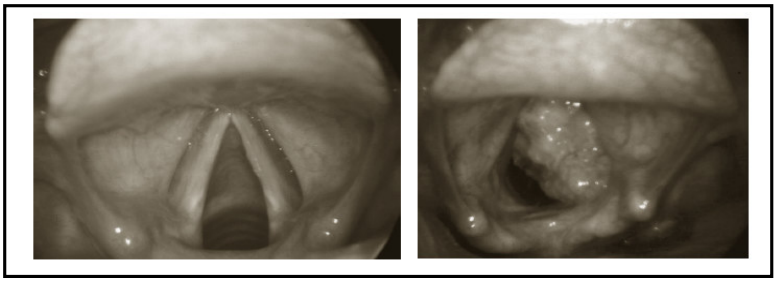

Figura 1.4 - *Esquerda: laringe normal; Direita: laringe cancerígena. De Geertsema, A. [6].*

A mesma fonte aponta como principais factores de risco para a ocorrência do cancro da laringe: o consumo de tabaco; o consumo abusivo de álcool; um estado nutricional deficiente; o Papiloma Vírus Humano; a existência do sistema imunitário enfraquecido; riscos ocupacionais como serrilha, gases de tintas, amianto e químicos das indústrias metalúrgica, petroquímica, têxtil e dos plásticos; o sexo do indivíduo, sendo 4 a 5 vezes mais frequente nos homens do que nas mulheres; a faixa etária, em que mais de 50% de novos diagnósticos ocorrem em indivíduos com idade superior a 65 anos; a raça, sendo, por exemplo, 50% mais comum em Afro-Americanos do que em Brancos Caucasianos; e o refluxo gastroesofágico.

A ACS [7] estima que, durante o ano de 2007, nos EUA, 11300 novas pessoas terão sido diagnosticadas como tendo cancro da laringe e que 3740 terão falecido devido a essa mesma condição. Apesar do número total de novos casos estimados ter aumentado em relação ao ano de 2006, [8], prevê-se também que menos doentes serão sujeitos a remoção total da laringe devido ao aumento da detecção precoce, bem como à melhoria dos métodos de tratamento não-cirúrgicos [9].

Em Portugal, segundo a Associação Portuguesa dos Limitados da Voz, APLV, estima-se existem cerca de 25 mil laringectomizados. Este número deve ser aceite como uma estimativa dado que não existem quaisquer estudos oficiais que o confirmem.

1.3.2 REABILITAÇÃO DA FALA

As principais formas de voz não-laríngea[2] são a voz esofágica, traqueoesofágica e a voz artificial / electromecânica / aeromecânica. Podem, contudo, considerar-se duas grandes categorias deste tipo de fala: a intrínseca e a extrínseca, Weinberg, B. [10]. As diferenças essenciais relacionam-se com a forma como o paciente gera a fonte não-laríngea: através de um mecanismo vozeador vicariante, intrínseca, ou através de uma fonte vibratória artificial não sistémica, extrínseca. Assim, as vozes esofágica e traqueoesofágica são métodos intrínsecos já que dependem do uso de estruturas anatómicas residuais como fontes de produção de falam; em ambas essa estrutura é o segmento faringoesofágico, também designado de pseudoglote. Por outro lado, o uso de uma laringe artificial é um método extrínseco.

Apesar de existirem mais duas formas de voz não-laríngea, as formas bucal e faringeal, estas não são muito desejáveis, Weinberg, B. e Westerhouse, J. [11, 12] e, apesar de nem todos os dados existentes serem conclusivos, Khaila, H. *et al* [13], costumam ser desencorajados. A voz bucal é o resultado da compressão de ar entre as bochechas e os dentes, ou entre a língua e a fossa incisiva dos maxilares, sendo portanto produzida na cavidade oral. A voz faríngeal é o resultado da compressão da língua contra as estruturas orais e / ou faríngeas; tal envolve o contacto da língua com a parede faríngea posterior, ou o palato duro ou o palato mole.

1.3.2.1 VOZ ESOFÁGICA

A voz esofágica é conseguida pela prévia injecção de ar da cavidade oral para o esófago, insuflando-o na sua região superior. Esta área é formada pelas mesmas estruturas do esfíncter do esófago superior, sendo composta por membrana mucosa e musculatura, nomeadamente o músculo crico-faríngeo e os músculos constritores da faringe. O ar previamente insuflado é depois libertado, causando a vibração do segmento faringoesofágico. A técnica de injecção de ar é descrita como um movimento ininterrupto, ocorrendo com o fonema inicial da palavra a produzir.

[2] [2] Ou “alaríngea” ou “não-vozeada”, ou seja, que não implicam a utilização da laringe para a produção de voz

A fala produzida pela voz esofágica, devido à fonte vibratória, tem características acústicas muito diferentes das de uma fala normal, Snidecor, J. [14]. Tais diferenças incluem aspectos como: a frequência fundamental, a intensidade e o débito vocal. A frequência fundamental da voz esofágica está demonstrada como apresentando cerca de metade do valor da voz de um adulto masculino normal, Snidecor, J. e Curry, E. [15]. A intensidade da voz também é reduzida em comparação à de falantes normais, Robbins, J. *et al* [16]. Devido à necessidade frequente de insuflar o reservatório esofágico, que apresenta uma pequena capacidade, Diedrich, W. [17], o débito vocal é reduzido consideravelmente, com efeitos negativos na prosódia conseguida.

A principal vantagem da voz esofágica é a de não requerer à utilização de qualquer aparelho externo para a produção de fala. Devido a tal, as mãos do falante estão livres para outros propósitos. Falantes proficientes com voz esofágica exibem a capacidade para codificar aspectos linguísticos tais como a entoação e a acentuação. Gandour, J. e Weinberg, B. [18, 19] e Gandour, J. *et al* [20].

As desvantagens da voz esofágica estão primariamente relacionadas com as baixas taxas de aquisição, Gates, G. *et al* [21]. A maioria das estimativas sugere que menos de metade dos pacientes que tentam este método conseguem de facto adquiri-lo. Outra desvantagem prende-se com o facto do processo de aquisição exigir um comprometimento de tempo substancial por parte do paciente. Contudo, as desvantagens mais significativas, como já anteriormente descrito, estão associadas às alterações das características acústicas e temporais. A grande maioria dos falantes com voz esofágica também não possui proficiência para efectuar e controlar alterações na fonte vocal, Angermeier, C. e Weinberg, B. *et al* [22]. Finalmente, a voz esofágica pode ter como consequência comportamentos secundários, como esgares faciais indesejados, que podem interferir negativamente em interacções comunicativas.

1.3.2.2 VOZ TRAQUEOESOFÁGICA

A voz traqueoesofágica, TE, é conseguida através de um fístula, criada cirurgicamente, entre a traqueia e o esófago, constituindo uma fonte vocal vicariante. Nesta fístula é inserida uma prótese vocal, denominada prótese traqueoesofágica. Esta

prótese actua como uma válvula de sentido único, capaz de desviar o ar pulmonar da traqueia para o esófago, quando selada pelo dedo do falante – tal como se vê na Figura 1.5. Note-se que o ar pulmonar é a fonte directa deste tipo de voz – a insuflação esofágica é obtida quase instantaneamente com a exalação, Singer, M. e Blom, E. [23], e a quantidade de ar insuflado é muito maior, Van den Berg, J. *et al* [24] – sendo que grande parte das diferenças acústicas entre a voz esofágica e a traqueoesofágica se devem a este facto. Existem várias próteses comerciais disponíveis. São geralmente fabricadas em silicone e em dimensões variáveis de modo a permitirem a adaptação às diferentes profundidades das punções traqueoesofágicas.

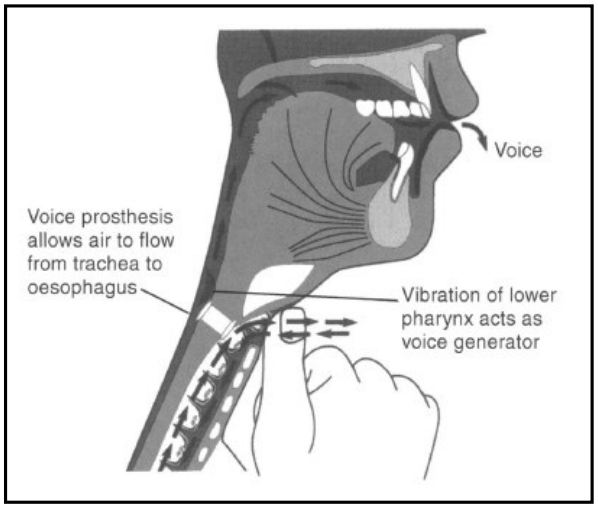

Figura 1.5 - *Desenho esquemático pós-laringectomia com fístula TE. De Jassar, P. et al [4].*

As principais vantagens da voz TE estão directamente relacionadas com o acesso ao ar pulmonar por parte do falante. O ar pulmonar permite um aumento substancial da capacidade de excitação do segmento faringoesofágico. Assim, a duração das frases e, consequentemente, o débito vocal, são melhorados e mais próximos dos de um falante normal. Tal permite também que o falante exiba uma prosódia mais natural e a capacidade de realizar, sistematicamente, alterações linguísticas associadas, Gandour, J. e Weinberg, B. [18, 19] e Gandour, J. *et al* [20]. O acesso ao ar pulmonar, o maior controlo e o fluxo de ar acrescido, levam também a um aumento da frequência e da intensidade da voz, Robbins, J. *et al* [25], o que pode ser vantajoso em ambientes ruidosos. A inteligibilidade da fala parece comparar-se favoravelmente com outros métodos de voz não-laríngea, Blom, E. *et al* [26]. Existem ainda outras vantagens,

associadas ao facto de este tipo de voz poder ser adquirida de uma forma relativamente rápida, após a cirurgia, quando comparada com a fala esofágica.

As principais desvantagens da voz traqueoesofágica estão relacionadas com o uso e a manutenção da prótese. Também é necessário aprender várias técnicas e desenvolver aptidões de modo a maximizar a fala, requerendo ainda a utilização de uma mão durante a fonação. Nem todos os indivíduos são candidatos para a implementação da prótese, e mesmo alguns que de facto o são, não têm capacidade para apreender este método. Quando a punção TE é feita secundariamente, isto é, realizada num período posterior à laringectomia, mas num período posterior, acarretando todos os riscos que um processo cirúrgico envolve.

1.3.2.3 LARINGE ARTIFICIAL ELECTRÓNICA

A laringe artificial, ou electrolaringe, EL, proporciona uma fonte vocal externa ao falante. Existem dois tipos de aparelhos comercialmente disponíveis e clinicamente prescritos, os aparelhos intraorais e os transcervicais. Ambos assentam no princípio da introdução de uma vibração electromecânica no tracto vocal que pode ser ouvida como um tom. A Figura 1.6 apresenta um desenho esquemático de utilização de laringe electrónica transcervical. Os aparelhos intraorais permitem a introdução da fonte vocal electrónica directamente na cavidade oral, onde a fala é articulada, por intermédio de um tubo. Os aparelhos transcervicais são os mais comuns e requerem o posicionamento em tecidos do pescoço ou face, sendo o sinal transferido para o tracto vocal / cavidade oral. A maioria das electro-laringes transcervicais, comercialmente disponíveis, oferece a possibilidade de adaptação à utilização intraoral por intermédio de um tubo ajustável.

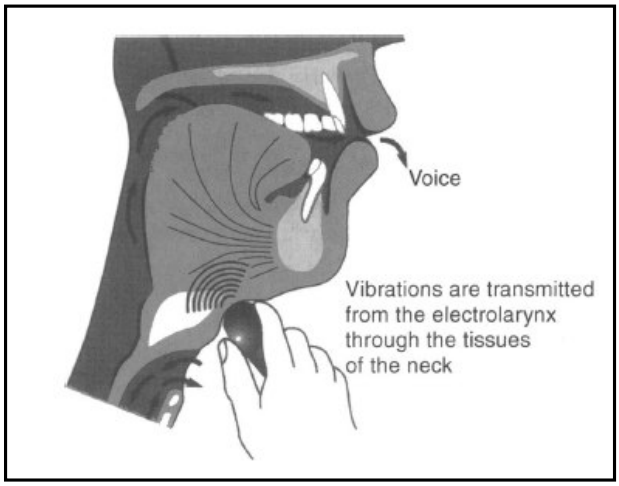

Figura 1.6 - *Desenho esquemático pós-laringectomia com utilização de electrolaringe transcervical. De Jassar, P. et al [4].*

A voz laríngea artificial, em relação aos restantes métodos intrínsecos, tem a vantagem distinta de poder ser adquirida rápida e facilmente pela maioria dos pacientes, tornando-se uma opção extremamente viável no período pós-operatório inicial. Contudo, alguns factores físicos podem impedir o uso de alguns dos aparelhos existentes, como por exemplo, no caso de pacientes com fibrose induzida pela radioterapia, em que os aparelhos transcervicais podem não ser apropriados. A inteligibilidade geral da fala pode ser bastante boa se os indivíduos forem efectivamente treinados, podendo, contudo, variar conforme o aparelho, Stalker, J. *et al* [27]. Como com outros métodos não-laríngeos, foram já relatadas dificuldades associadas a um contraste de voz eficaz, Weiss, M. e Basili, A. [28], mas a intensidade conseguida parece ser melhor do que na voz esofágica, podendo oferecer ao paciente uma vantagem em ambiente ruidoso. A grande maioria das electro-laringes comercialmente disponíveis também permite ao falante a possibilidade de ajustar a intensidade de forma variável.

As principais desvantagens da fala quando produzida por intermédio de uma laringe artificial foram atribuídas à qualidade mecânica da voz produzida. Apesar de, desde a sua implementação, terem sido introduzidas algumas modificações, em electro-laringes comerciais, e terem efectuados testes em electro-laringes experimentais, os problemas associados à qualidade mecânica e "artificial" da voz ainda persistem. O falante pode também ter dificuldade em sinalizar componentes linguísticos do código falado, Gandour, J. e Weinberg, B. [29]. O facto da electro-laringe ser um aparelho electrónico, e necessitar de baterias para o seu funcionamento, pode impedir ou debilitar

seriamente a comunicação no caso de avaria, de defeito ou de falha na alimentação. Independentemente da capacidade do falante para comunicar, efectivamente, com uma laringe artificial, o nível de aceitação, quer pelo falante quer pelo ouvinte, parece ser reduzido em relação a outros métodos, Bennet, S. e Weinberg, B. [30]. A laringe artificial também requer o uso de uma das mãos, sendo claramente visível para outros, o que pode criar uma preocupação no falante. A desvantagem final prende-se com o custo elevado associado à compra e manutenção do material

1.4 CARACTERIZAÇÃO DE LARINGES ARTIFICIAIS ELECTRÓNICAS ACTUAIS

De acordo com Hillman, R. [31] as electro-laringes presentemente disponíveis produzem uma fala pouca natural – mecânica, robótica, monótona – com reduzida inteligibilidade e intensidade e que chama alguma atenção não desejada para o falante. A fraca qualidade da fala produzida por intermédio de uma electro-laringe pode ser atribuída às limitações dos transdutores actuais e à perda de controlo fino do tom, amplitude, inicialização e terminação da fala. A perda do controlo fino causa défice no vozeamento segmental, distinção vozeado / não vozeado para consoantes, e suprasegmental, entoação e stress silábico. A utilização de uma mão para o controlo da EL também é fisicamente limitadora impedindo a função bi-manual normal.

Já Barney, H. [32], em 1958, aponta as principais características que uma EL deveria ter, incluindo: qualidade da fala produzida e inflexão do tom próximas da fala normal; intensidade da fala igual à intensidade da fala normal; pequena e não obstrutiva, sem fios, tubos ou outros acessórios visíveis; capacidade de operação durante longos períodos de tempo; higienicamente aceitável para o utilizador; e reduzido custo de compra e manutenção.

1.4.1 NOTAS HISTÓRICAS

Shute, B. [9] descreve sucintamente a história das laringes artificiais. A primeira electro-laringe é atribuída a Thermistocles Gluck em 1909 – o mesmo foi também responsável pela revolução nos procedimentos da laringectomia no final do século XIX, introduzindo a base da técnica que é hoje praticada. Gluck utilizou um aparelho semelhante ao fonógrafo de Edison onde a gravação de voz de um cantor era utilizada para criar um sinal enviado para uma pequena coluna, que podia ser colocada no nariz ou no palato duro do paciente, transmitindo-o para o tracto vocal.

Gilbert M. Wright é creditado pela invenção da primeira electro-laringe transcervical em 1942, de seu nome Sonovox. Inicialmente desenhada para a produção de efeitos sonoros antropomórficos em filmes – foi, por exemplo, utilizada no filme de animação da Disney *Dumbo* – uma outra versão especial foi produzida com o intuito de ser utilizada por laringectomizados. A companhia americana Aurex Corporation foi a responsável pela produção da primeira electro-laringe comercial, baseada na Sonovox, e denominada Aurex Neovox M-520T, que estabeleceu as fundações da electro-laringe moderna e de todos os modelos a serem produzidos posteriormente.

Em 1959, Barney, H *et al.* [33], dos Laboratórios Bell, apresentaram o primeiro modelo de uma laringe electrónica com base em transístores. O aparelho tinha três transístores, dois utilizados para criar um pulso oscilatório e outro para amplificar o sinal, e era operado através de um potenciómetro que permitia ajustar o tom durante a conversação. O sinal era transferido para o pescoço através de um receptor telefónico. Mais tarde, denominada de Western Electric e comercializada pela AT&T, tornou-se bastante ubíqua entre utilizadores e investigações científicas. Deixou de ser produzida em meados dos anos 90, tendo outras marcas e aparelhos, que entretanto emergiram, dominado o mercado.

1.4.2 ESTADO DA ARTE INDUSTRIAL

Actualmente existem no mercado várias empresas que fabricam e comercializam electro-laringes, das quais se destacam: Servox, na Alemanha, e Griffin Laboratories, Mount Precision Manufactoring, Luminaud, Romet, OptiVox, UltraVoice, todas nos Estados Unidos da América, EUA. Verifica-se claramente que o mercado é dominado pelos EUA em termos de número de fabricantes. Segue-se agora uma breve descrição dos modelos disponibilizados por cada empresa bem como das suas características principais.

Os modelos disponibilizados pela Servox são:

A Servox Digital é uma electro-laringe transcervical, com controlo digital de volume e tom programáveis, por meio de dois botões de pressão e um botão de tambor. Vem acompanhado de um *software* para PC específico e um cabo de ligação que permite ajustes mais finos de programação da "melodia" do tom gerado. Apresenta também um invólucro de plástico revestido a titânio para maior protecção e durabilidade. Utiliza pilhas recarregáveis NiMH e vem acompanhado de um carregador. Oferece a possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

Figura 1.7 - *Servox Digital. De [34]*

A Servox Inton é uma electro-laringe transcervical com controlo do volume e tom, permitindo variações básicas da voz, pela utilização de um tom de base e de acentuação. Permite ainda programação de uma redução automática do tom durante cada fonação. Utiliza pilhas recarregáveis e vem acompanhado de um carregador. Oferece a

possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

Figura 1.8 - *Servox Inton. De [34]*

A Servox Eco é uma electro-laringe transcervical, com ajuste de volume e tom, num botão multifunções único. Utiliza pilhas alcalinas convencionais. O invólucro é de plástico.

Os modelos disponibilizados pela Griffin Laboratories são:

A TruTone ElectroLarynx é uma electro-laringe transcervical com capacidade de controlo do tom, em tempo real, através de um botão, permitindo entoação durante a conversação. Utiliza pilhas alcalinas de 9V convencionais. Oferece a possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

A SolaTone Artificial Larynx é uma electro-laringe transcervical de um só tom, operada por um único botão. Utiliza pilhas alcalinas de 9V convencionais. Oferece a possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

Figura 1.9 – *SolaTone. De [35]*

Para ambos os modelos é possível adquirir um suporte de mãos-livres que se coloca na zona do peito, fazendo-se a activação do tom utilizando o queixo.

Os modelos disponibilizados pela Mount Precision Manufactoring são:

A NuVois III Digital é uma electro-laringe transcervical, que apresenta controlo digital dual e independente de volume e tom. Permite a utilização de baterias recarregáveis ou baterias alcalinas convencionais. Oferece a possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

Figura 1.10 - *Nu Vois III digital, esquerda; Nu Vois I, direita; Nu Vois II, centro. De [36]*

A NuVois II Digital é uma electro-laringe transcervical, com controlo digital de volume e de tom. Utiliza baterias NiMH recarregáveis e vem acompanhada de um carregador. Oferece a possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

A NuVois I trata-se de uma electro-laringe transcervical, com controlo manual de volume e tom. Utiliza baterias NiMH recarregáveis e vem acompanhada de um carregador. Oferece a possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

XtraVois é uma electro-laringe transcervical básica de tom único, e é operada por meio de um único botão. Utiliza baterias NiMH recarregáveis e vem acompanhada de um carregador.

O único modelo disponibilizado pela Luminaud é:

A Cooper-Rand é a única electro-laringe intra-oral adaptada à utilização mãos-livres ou com pouca utilização das mesmas. Atinge esse objectivo pois não introduz vibração contra o tecido do pescoço, mas sim através de um tubo de plástico posicionado na boca e direccionado para o palato. Como o tom é introduzido directamente na boca, não interfere com o processo de recuperação pós-cirurgica, podendo ser utilizado imediatamente após a cirurgia, e permite a utilização simultânea com um ventilador. Tem um desenho em duas peças: um gerador de tom, transportado na mão, no bolso ou à cintura, e um emissor de tom, colocado na base do tubo que se insere na boca. Possui dois controlos com botão de tambor para o volume e o tom, no módulo gerador de tom. Utiliza duas pilhas alcalinas de 9V convencionais.

Figura 1.11 - *Cooper-Rand. De [37]*

O único modelo disponibilizado pela Romet é:

A Romet Electronic Larynx é uma electro-laringe transcervical de dimensões reduzidas. Está equipada com controlo de volume e de tom, sendo operada apenas por um botão. Utiliza baterias NiMH recarregáveis, e vem acompanhada por um carregador. Oferece a possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

O único modelo disponibilizado pela OptiVox é:

A OptiVox Plus é uma electro-laringe transcervical com controlo por intermédio de um botão de tambor para volume e tom, sendo operada apenas por um botão. Utiliza baterias NiMH recarregáveis ou baterias alcalinas convencionais. Oferece a

possibilidade de utilização de um tubo oral no caso de impossibilidade de utilização transcervical.

O único modelo disponibilizado pela UltraVoice é:

A UltraVoice Plus, uma electro-laringe intra-oral é constituída por duas unidades: a intraoral e a de controlo. A unidade oral é especialmente montada na dentadura superior, através de um fixador ortodontico, e contém um transdutor de sinal que emite o tom, e um circuito RF que recebe o sinal proveniente da unidade de controlo, convertendo-o num sinal eléctrico que activa o transdutor. A unidade oral é alimentada por uma bateria de pequenas dimensões, recarregável. A unidade de controlo possui um microprocessador, de alta velocidade, que amostra a fala durante a fonação, de modo a controlar a prosódia em tempo real. A electro-laringe possui também um adaptador mãos-livres que é constituído por um sensor de pressão revestido a velcro, que se liga directamente à unidade de controlo, através de um cabo *jack* vulgar, e permite a activação da electro-laringe através da criação de pressão no sensor pelo movimento, por exemplo de um braço, ou perna, podendo ser montado em qualquer lugar, desde que o utilizador lhe tenha acesso para proceder à activação.

1.4.3 ESTADO DA ARTE CIENTÍFICO-TECNOLÓGICA

Os desenvolvimentos que ocorreram no estado da arte cientifício-tecnológica das electro-laringes focaram-se tipicamente em dois grandes campos: a melhoria da qualidade do som produzido / redução de ruído e a implementação de funções de controlo de activação e / ou tom. A melhoria do som produzido recorre tipicamente a sintetizadores de fala enquanto a redução de ruído é conseguida por intermédio de várias técnicas de processamento de sinal, incluindo a filtragem adaptativa e a subtracção espectral.

O controlo de activação depende na maioria das electro-laringes do controlo manual por intermédio de diversos botões. Trabalhos recentes apontam para outras possibilidades de controlo, nomeadamente o controlo electromiográfico. O controlo de tom também tem sido alvo de grande atenção, já que a qualidade monótona do som

produzido pelas electro-laringes se deve, em parte, à incapacidade destas produzirem acentuação e entoação durante a fonação. Em relação ao controlo de tom, podem encontrar-se os seguintes tipos: o controlo manual; o controlo pela pressão da expiração; e o controlo neuronal / electromiográfico.

Pretende também destacar-se um trabalho que apresentou um protótipo de uma electro-laringe implantável, e a consequente recolha de dados dos resultados obtidos após a implantação. Apesar de parecer ser uma solução viável e promissora, este campo não tem atraído a atenção de investigadores no domínio das electro-laringes.

1.4.3.1 PROTÓTIPOS DE LARINGES ARTIFICAIS

Duas electro-laringes comercialmente disponíveis já permitem o controlo manual do tom durante a conversação, a TrueTone e a UltraVoice Plus, ambas descritas na secção anterior, através de sensores de pressão. A questão do controlo manual através de sensores de pressão foi também abordada por Choi, H. *et al* [38] que desenvolveram um protótipo para uma nova electro-laringe transcervical, denominada EVADA, com controlo simultâneo da frequência e da intensidade, durante a conversação, por intermédio de dois botões com sensores de pressão – a Figura 1.12 mostra em detalhe o sensor de pressão desta electro-laringe. Através de testes perceptuais, os autores afirmam que a capacidade desta electro-laringe expressar entoação e acentuação é maior que a das outras electro-laringes actualmente disponíveis no mercado.

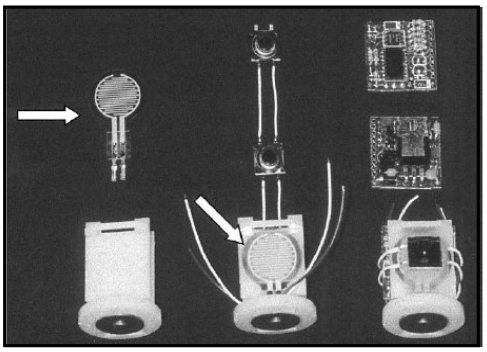

Figura 1.12 - *Detalhe do sensor de pressão da electolaringe EVADA. De Choi, H. et al [38].*

Mais recentemente, Takahashi, H. *et al* [39] desenvolveram um protótipo de uma laringe intra-oral com controlo manual sem fios das funções de activação e do tom. Os interruptores manuais enviam informação binária via luz infravermelha, IV, para um controlador – este por sua vez já se encontra ligado ao vibrador electromecânico intra-oral por meios de fios – que implementa a geração do tom no vibrador – tal como se vê na Figura 1.13. Desta forma, permite o controlo da variação do tom durante a conversação, recorrendo a uma modelo de implementação de controlo do tom baseado no modelo de Fujisaki – ver Fujisaki, H. [40]. A inteligibilidade conseguida por esta electro-laringe é próxima daquela conseguida pelas electro-laringes comerciais, tendo a vantagem de libertar, em parte, as mãos do utilizador. No entanto, a parte do aparelho que ainda recorre a fios, a ligação entre o controlador e o vibrador, parece perturbar os utilizadores.

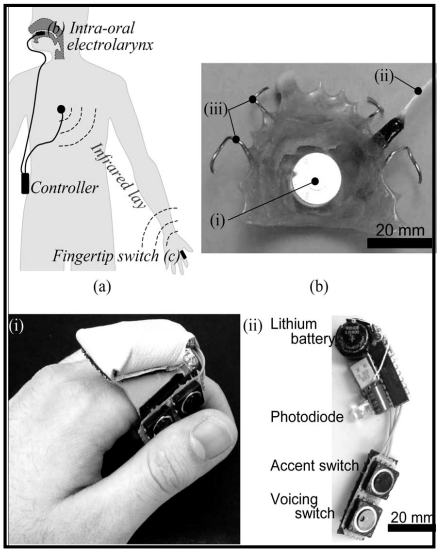

Figura 1.13- *Electro-laringe controlada via IV; a) diagrama de funcionamento; b) vibrador intraoral fixado em prótese palatina; i) detalhe dos interruptores manuais para controlo de activação e acentuação; ii) detalhe dos interruptores manuais e célula de alimentação. De Takahashi, H. et al [39]*

Note-se que o vibrador intra-oral deste último protótipo se encontra fixado numa prótese palatina semelhante à estrutura e material das próteses dentárias. O primeiro sistema deste tipo foi descrito por Stern, K. [41] que detém mesmo a patente americana do sistema.

Ainda no domínio do controlo manual do tom, Kibuchi, Y. e Kasuya, H. [42] propuseram uma electro-laringe cuja activação e controlo da frequência fundamental é conseguida pelo movimento lateral – activação – e vertical – frequência fundamental – do dedo polegar. Os movimentos lateral e vertical são detectados e medidos através de um sistema de detecção da luz emitida por um LED.

Matsushima, J. *et al* [43] e Ifukube, T. [44] descreveram o único protótipo de uma electro-laringe que permite o controlo da entoação através da pressão de expiração produzida no estoma. A electro-laringe é constituída por três partes: um sensor de pressão que detecta a pressão da expiração; uma unidade de controlo que converte a pressão da expiração em frequência fundamental de voz; e um vibrador electromecânico transcervical que gera o tom. A Figura 1.14 mostra esta electro-laringe com destaque para o sensor de pressão. Após um período de treino, o utilizador testado era capaz de produzir uma entoação aproximada da de um falante normal, sendo a fala produzida mais próxima da natural quando comparada com uma electro-laringe normal. No entanto, o processo de treino parece ser o lado mais complexo desta abordagem, sendo mais difícil para os utilizadores de longa data de electro-laringes convencionais, que tipicamente sustêm a respiração durante a fonação. A melhoria da qualidade da voz produzida também é limitada, devido às variações de tom que é possível detectar, sendo a determinação da função óptima de transformação de pressão de expiração em frequência uma tarefa bastante difícil.

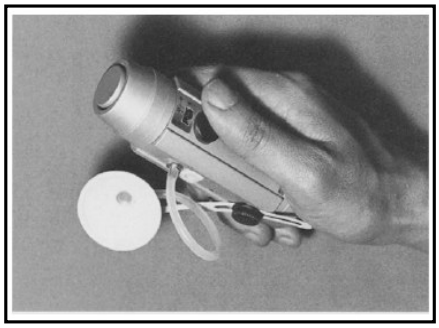

Figura 1.14 - *Electro-laringe com controlo de entoação através da pressão de expiração. De Ifukube, T. [44].*

A possibilidade de controlo de uma electro-laringe a partir de sinais electromiográficos data já do final dos anos 70. A primeira tentativa deveu-se a Sugie, N. e Tsunoda, K. [45], que, baseados nos resultados de Cole, K. *et al* [46], desenvolveram uma prótese de fala, com recurso a um sintetizador para voz para a produção de vogais, cuja activação e discriminação das vogais a serem sintetizadas era conseguida pela recolha de informação electromiográfica de três músculos da boca e / ou face. Os movimentos da boca são detectados através de três canais de EMG, com recurso a eléctrodos de superfície, colocados nos músculos digástrico, zigomático maior e orbicular da boca. Para obter a discriminação de vogais, o sistema foi treinado de modo a determinar que músculos são activados para a produção de cada vogal. Através de uma operação de limiarização – quando um sinal ultrapassa o limiar determinado é produzido um impulso de amplitude constante – foi possível obter uma discriminação muito satisfatória da vogal que se pretende sintetizar. Apesar de encorajadores, os resultados apresentam ainda limitações em relação à robustez da discriminação, bem como no momento da síntese, talvez relacionados com assincronia entre a actividade dos músculos e a respiração associada à produção de fala, sendo também requerido ao utilizador um período de treino. A discriminação de consoantes, apesar de não fazer parte dos objectivos iniciais deste estudo, também se torna extremamente difícil utilizando o método descrito. Contudo, os autores sugerem a possibilidade de utilização de sinais posicionais da língua, através de palatometria dinâmica e sensorização óptica, de modo a atingir esse objecitos.

Houston, K. *et al* [47] introduziram a necessidade de um mecanismo de geração da forma de onda, próxima da onda glotal naturalmente produzida, apontando para as possibilidades da filtragem inversa, apropriadamente compensada para as distorções introduzidas pelo processo de transdução, centrando-se este trabalho no desenvolvimento de um melhor componente transdutor. O sistema proposto introduz também um módulo de controlo da fonte sonora baseado em sinas neuronais – não tendo sido contudo implementado, simplesmente conceptualizado, durante este estudo já que o objectivo principal era o melhoramento da fonte e do som produzido. À data do artigo, tinha sido realizado o primeiro transplante de nervo laríngeo humano para um músculo do pescoço, próximo da superfície, e esperava-se que eléctrodos EMG de superfície fossem capazes de detectar a actividade do nervo laríngeo transplantado,

gerando sinais de controlo que poderiam ser utilizados da mesma forma que eram utilizados os sinais de controlo para membros prostésicos.

No ano seguinte, Heaton, J. *et al* [48], demostraram, em cobaias, existir uma correlação entre os sinais do músculo esterno-hióide, para onde tinham transplantado o nervo laríngeo recorrente, e aqueles dos músculos laríngeos. Observações histoquímicas também indicaram uma transformação unilateral da maioria das fibras do músculo esterno-hióide, ipsilaterais à transposição, para um fenótipo mais típico de músculo laríngeo. Estas descobertas davam força à ideia de que o nervo laríngeo recorrente, no momento da laringectomia, deveria ser conservado por transposição, para um outro músculo, mantendo-se os sinais de controlo laríngeo intactos.

Heaton, J. *et al* [49], apresentaram os resultados do estudo, ao longo de um ano, da actividade electromiográfica de superfície, durante tarefas fonatórias e não-fonatórias, em músculos do pescoço para os quais tinha sido transplantado o nervo laríngeo recorrente, no momento da laringectomia. Em todos os oito indivíduos foram obtidos sinais que se correlacionavam com a produção de fala.

Goldstein, E *et al* [50], implementaram um sistema que utiliza sinais electromiográficos de um músculo do pescoço – omo-hióide esquerdo – para o qual tinha sido transplantado o nervo laríngeo recorrente no momento da laringectomia, para controlo de uma electro-laringe transcervical mãos-livres. Os sinais eléctricos do músculo são detectados por eléctrodos de superfície bipolares e processados para produzir um controlo de activação na electro-laringe. Como a amplitude do sinal recolhido por EMG é de poucas dezenas de microvolts, e concentra a sua energia numa banda dos 10 aos 500 Hz, o sistema também utiliza um amplificador que amplifica e filtra o sinal melhorando a razão sinal-ruído. A Figura 1.15 mostra o sistema destacando os eléctrodos de superfície, a unidade de controlo e o vibrador electromecânico. Apesar de este estudo ter comprovado o conceito de utilização de sinais de EMG para controlo sem mãos de uma electro-laringe, o protótipo é ainda pouco ergonómico e funcional, para além de não possuir uma bateria, não podendo portanto ser utilizado no dia a dia.

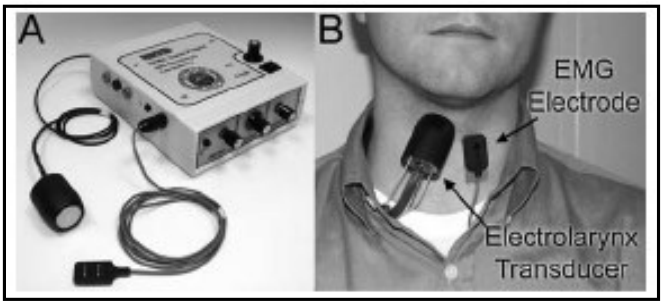

Figura 1.15 - *Electro-laringe com controlo de activação por intermédio de sinais electromiográficos; a) sistema EMG – Electro-laringe; b) disposição em sujeito. De Goldstein, E. et al [50].*

Estudos futuros deverão concentrar-se no desenvolvimento de um protótipo compacto e funcional que poderá ser implantado intra-oralmente. O processamento digital do sinal electromiográfico poderá ser, também, um caminho para melhorar a resposta e a resolução do sistema. A incorporação de uma bateria com uma autonomia adequada é indispensável. Segundo o artigo, está também em desenvolvimento um controlo da frequência fundamental a partir dos mesmos sinais EMG.

No seguimento do trabalho de Griffiths, M. *et al* [51], Painter, C *et al* [52] desenvolveram uma laringe artificial, electromagnética, implantável, constituída por um transdutor implantável – no espaço pré-vertebral oposto a C2-C3 – e uma unidade de controlo que implementa o sinal e permite o ajuste do volume, do tom e do contorno de entoação. No modelo descrito mais recentemente, o transdutor tem acoplado um receptor RF, também implantado subcutaneamente, que recebe sinais de um emissor incluído na unidade de controlo. Uma electro-laringe implantável apresenta as seguintes vantagens: permitir a introdução do sinal mais próximo do local natural, ou seja, na localização real da laringe, optimizando, da melhor forma, as características acústicas do tracto vocal, podendo, desta forma, melhorar a razão-sinal ruído; eliminar a utilização das mãos; e apresentar a capacidade de tornar o mecanismo não visível aos interlocutores do utilizador. No entanto, o modelo proposto pelos autores apresenta várias limitações, nomeadamente: a qualidade do som – pobre desempenho nas baixas-frequências; existência de uma compensação fraca da distorção introduzida pela camada de pele que vai cobrir o transdutor; a impossibilidade de comunicação durante algum tempo após cirurgia; ainda não apresenta soluções para o carregamento das baterias do transdutor e do receptor RF; os custos elevados da cirurgia; e, ainda, os problemas de

biocompatibilidade. A existência de avanços recentes nas áreas da micro e da nanotecnologia, dos biomateriais, bem como na cirurgia, deverão proporcionar o conhecimento necessário ao desenvolvimento de novos protótipos.

1.4.3.2 MELHORIA DO SINAL DE FALA DE ELECTRO-LARINGES

As duas principais abordagens à melhoria do sinal de fala produzido por electro-laringe focam a melhoria da qualidade fonética, através de sistemas com reconhecedores e sintetizadores de fala integrados, e a remoção do ruído aditivo.

1.4.3.2.1 MELHORIA DA QUALIDADE FONÉTICA

No que diz respeito à melhoria da qualidade fonética destacam-se os trabalhos de Shoureshi, R. *et al.* [53] que, tendo como objectivo final incorporar a electro-laringe na prótese dentária, focou este trabalho, inicialmente, no desenvolvimento de uma extensão da electro-laringe, posicionada na boca. Para a criar um som de fala mais natural, dentro da gama de frequências desejada, os sinais originados na boca pela laringe artificial externa são processados e filtrados por um sintetizador de fala. Os sinais do sintetizador são depois enviados para actuadores de pequena escala que, juntamente com os movimentos da boca, vão gerar o sinal final de fala. Dados os efeitos de não linearidade envolvidos na criação da fala através dos movimentos da boca, bem como a existência de um número de fontes de incerteza durante esse mesmo processo, foi desenvolvido um sintetizador neuro-difuso, combinado, não-linear, com uma arquitectura única.

Aguilar, G. *et al.* [54, 55] desenvolveram um sistema de melhoria do sinal de fala produzido por uma electro-laringe, onde as partes vozeadas são determinadas através de um método de reconhecimento de padrões e substituídas, de acordo um livro de código, por sinais de fala vozeados normais, equivalentes, gerados por um sintetizador. No sistema, o sinal de fala distorcido, proveniente da electro-laringe, é primeiramente pré-processado, de modo a reduzir os efeitos de ruído, provocados pela própria electro-laringe nas estimações procedentes. A avaliação do sistema demonstrou uma precisão

próxima de 90% na substituição do segmento de fala electrolaríngea pelo segmento de fala normal. A avaliação subjectiva concluiu que o sistema introduz uma melhoria qualitativa significativa na voz artificial de utilizadores de electro-laringe

1.4.3.2.2 REDUÇÃO DO RUÍDO ADITIVO

Durante a fonação, alguns dos sons produzidos pela electro-laringe transcervical irradiam directamente do aparelho, produzindo um canal de interferência directa, o que afecta a inteligibilidade do discurso e aumenta a percepção da fala como mecânica, Holly, S. *et al* [56]. Vários factores podem contribuir para a radiação de ruído directo, incluindo a forma como a electro-laringe é colocada no pescoço, o padrão vibratório e as características de radiação do tecido cervical.

As primeiras tentativas para reduzir o efeito do ruído de radiação envolveram o isolamento acústico físico, através de variados materiais, como por exemplo em Norton, R. e Bernstein, R. [57] que, para além de entrarem em consideração com a função de transferência do tecido cervical humano, para produzirem um sinal de fonte melhorado, isolaram a electro-laringe através de uma camada de espuma sólida. Apesar de alguma redução do ruído e de uma melhoria na inteligibilidade, este método não era totalmente eficaz e tornava a electro-laringe maior em tamanho.

Devido à limitação das técnicas de isolamento acústico, as técnicas de processamento de sinal parecem as mais indicadas para a remoção de ruído. Até agora, as principais técnicas propostas basearam-se em duas abordagens diferentes: a filtragem adaptativa e a subtracção espectral. Os métodos de filtragem adaptativa concentram-se em estimativas do ruído – com recurso a um sinal base de referência que se admite correlacionado com o ruído mas não correlacionado com o sinal desejado – removendo as componentes de ruído a partir dessas estimativas. Os métodos subtractivos incluem a subtracção da energia espectral, a subtracção de magnitude espectral e a subtracção espectral com base em estimativas do ruído. Estes métodos baseiam-se nos pressupostos que a fala e o ruído aditivo são sinais não-correlacionados e que o sinal de fala é um processo com uma variação lenta, o que é verdade no caso da grande maioria das electro-laringes comerciais, onde o processo pode ser mesmo invariante no tempo,

durante períodos relativamente longos. Os algoritmos subtractivos estimam o ruído aditivo, subtraindo ao sinal ruidoso o valor estimado no domínio das frequências.

No domínio da filtragem adaptativa, destaca-se o trabalho de Espy-Wilson, C. *et al* [58] onde o sinal gravado directamente da electro-laringe é considerado o sinal de referência e o sinal de fala é o sinal ruidoso. O filtro adaptativo, que tenta aproximar o sinal ruidoso, opera sobre o sinal de referência para produzir a saída. Neste caso, a saída reproduz as componentes do sinal ruidoso que se correlacionam com o sinal de referência. O sinal de erro resultante da diferença entre os dois sinais, o de referência filtrado e o sinal de fala ruidoso, é a saída final do sistema e deverá ser o sinal de fala ruidoso, desprovido da componente de ruído aditivo. Este sinal de erro também é utilizado para controlar e modificar os coeficientes do filtro, através de um algoritmo de mínimos quadrados, re-estimados a cada amostra. A natureza adaptativa do filtro permite-lhe reagir a quaisquer variações do ruído causadas por qualquer uma das possibilidades acima descritas. A análise espectral do sinal filtrado apresentou uma redução significativa do ruído de fundo, enquanto a análise perceptual revelou preferência dos intervenientes pelo sinal filtrado, e a análise de inteligibilidade não mostrou melhoria ou degradação significativa da mesma.

Niu, H. *et al* [59] introduziram uma melhoria no método anterior substituindo o algoritmo de mínimos quadrados pela análise de componentes principais. A análise acústica mostra um bom desempenho na redução do ruído. Os resultados perceptuais demonstram uma melhoria na aceitação do sinal filtrado tanto em ambiente silencioso como em ambiente ruidoso. A análise da inteligibilidade não evidencia melhoria ou degradação significativa da mesma.

No domínio dos métodos subtractivos pode destacar-se o trabalho de Cole, D. *et al* [60], que desenvolveram um método híbrido de redução de ruído, que utiliza a subtracção espectral e a subtracção de raízes do cepstro. Os métodos híbridos utilizam o passo da subtracção de raízes espectrais para conseguir redução do ruído residual, após o método de subtracção espectral. O factor crítico passa pela precisão da estimativa do ruído. Este passo é mais simples no caso das electro-laringes, dada a natureza previsível do sinal de excitação que não varia significativamente em amplitude e tom e o facto da electro-laringe ser operada durante toda uma elocução, mesmo em momentos onde

existiria silêncio durante a fala normal. Tipicamente seria utilizada uma forma de descriminação entre vozeamento e silêncio, e o sinal contido no silêncio seria considerado ruído, depois utilizado para ajustar as estimativas. O primeiro factor facilita a detecção do sinal de voz, e o segundo permite obter uma boa estimativa do som do canal de interferência directa. A abordagem deste trabalho baseia-se num método de autocorrelação desenvolvido por Krubsack, D. e Niederjohn, R. [61]. Neste, a autocorrelação normalizada ao longo da gama frequencial, e a energia da janela, são utilizadas para determinar a medida de confiança de vozeamento. Cole, D. *et al* [60], determinam o máximo valor de autocorrelação ao longo de uma gama frequencial muito curta. O limite inferior permite determinar a presença de excitação electrolaríngea, enquanto o limite superior permite determinar a presença de fala. Quaisquer janelas onde a medida de autocorrelação se situe no meio dos limites são considerados ruído, e a estimativa do ruído é ajustada em concordância. A redução de ruído através deste método foi avaliada em três utilizadores de electro-laringe masculinos. Subjectivamente, o ruído aditivo directo é reduzido significativamente, com aparente não redução da inteligibilidade.

Pandey, P. *et al* [62] apresentaram um método de subtracção espectral modificado, onde a magnitude do espectro médio do ruído é obtido com os lábios fechados, através dos quadrados das magnitudes da Transformada Rápida de Fourier, FFT, de janelas adjacentes, e depois subtraído à magnitude do espectro do sinal de fala ruidoso. O estudo mostrou que a cancelamento do ruído aditivo é efectivo quando se utiliza uma janela de análise que inclua dois períodos fundamentais. Contudo, os valores óptimos dos parâmetros de processamento têm de ser obtidos empiricamente.

Pratapwar, S. *et al* [63] propuseram ainda um outro método onde a actualização do espectro do ruído é levada a cabo por intermédio de uma estimação baseada em quantil – no método anterior a estimativa do espectro do ruído era tomada como constante ao longo de toda a produção de fala – não requerendo detecção de segmentos vozeados e não-vozeados.

Mais recentemente, Liu, H. *et al* [64, 65] sugeriram dois métodos de subtracção espectral que tomam em consideração as propriedades de mascaramento de frequências do sistema auditivo humano – experiências de mascaramento de frequências mostram

que, o ruído próximo dos picos das formantes é inaudível para o ouvido humano, ou seja, ruído próximo de pontos onde o sinal de fala tem elevada energia. Num dos métodos. é criado um limiar de mascaramento auditivo, AMT, utilizado na adaptação paramétrica do processo subtractivo; e noutro, é desenhado um algoritmo de limiar de mascaramento suplementar, SAMT, que aplica uma subtracção espectral, baseada na correlação cruzada. Os testes perceptuais destes dois métodos demostram um melhor desempenho que o método de subtracção espectral de potência.

Strothjohann, M. e Buzug, T. [66, 67] também desenvolveram um algoritmo que considera o processo auditivo, desenhando um modelo auto-regressivo do mesmo. O filtro proposto baseia-se num método de predição linear para a audição, separando o sinal percepcionado em componente dinâmica e componente estática. Através de um processo de filtragem inversa, é obtida somente a componente dinâmica, reduzindo então o ruído aditivo.

1.5 A ELECTRO-LARINGE NA PERSPECTIVA DO UTILIZADOR

O único laringectomizado registado como utilizador de electro-laringe no centro da APLV, no Instituto Português de Oncologia, IPO, do Porto, apontou como principais dificuldades da utilização diária da sua electro-laringe: a obrigatoriedade de utilização das mãos para proceder à fonação, referindo, por exemplo, a incapacidade de falar enquanto conduz; a impossibilidade de regular o tom e o volume durante a fonação; complicações acrescidas durante conversação em ambientes ruidosos, onde diminui a capacidade de ser ouvido. Quando questionado sobre qual o principal elemento que, na sua opinião, deveria ser melhorado, referiu, em primeiro lugar a qualidade do som produzido, bem como a capacidade de introduzir entoação, e, em segundo lugar, a possibilidade de utilização de um aparelho do tipo mãos-livres.

2 CARACTERIZAÇÃO DA FONTE GLOTAL, SEUS COMPONENTES E SINAIS

2.1 ANATOMIA DA LARINGE

A anatomia da laringe encontra-se descrita de forma extensiva em vários livros de referência de Anatomia Médica, como por exemplo em Williams, P. [68].

2.1.1 CARTILAGENS E INTERIOR DA LARINGE

A laringe situa-se na zona anterior do pescoço, ao nível dos corpos vertebrais C3 até C6. O esqueleto da laringe consiste em nove cartilagens unidas por ligamentos e membranas. Três cartilagens são ímpares – tiróide, cricóide e epiglótica – e três são pares – aritenóide, corniculada e cuneiforme.

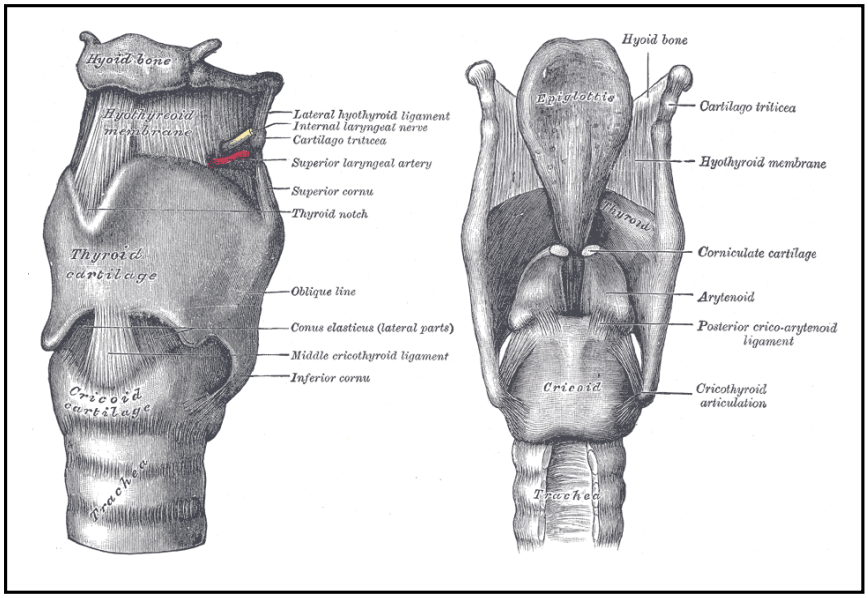

Figura 2.1 - *Ligamentos da laringe: vista antero-lateral – esquerda – e posterior – direita. De Williams P. et al [68].*

A cartilagem tiróide é a maior das cartilagens. Os dois terços inferiores das suas duas lâminas, semelhantes a placas, fundem-se anteriormente no plano mediano para formar a proeminência laríngea, vulgarmente designada por "maçã de Adão". Acima desta proeminência, as lâminas separam-se para formar uma incisura tiróide superior em forma de V. A margem posterior de cada lâmina projecta-se superiormente como corno superior e, inferiormente, como corno inferior. A margem superior e os cornos superiores ligam-se ao osso hóide por meio da membrana tiro-hióideia. A parte mediana, mais espessa, desta membrana é o ligamento tiro-hióideu mediano, e as suas partes laterais designam-se ligamentos tiro-hióideus laterais. Estes ligamentos e o espessamento na sua margem livre – os ligamentos vocais – fazem parte do cone elástico que fecham o ádito da laringe com excepção da fenda glótica, a região de abertura entre os ligamentos vocais. Os cornos inferiores da cartilagem tiróide articulam-se com as faces laterais da cartilagem cricóide nas articulações cricotiróideias. Os principais movimentos nestas articulações são a rotação e o deslizamento da cartilagem tiróide sobre a cartilagem cricóide, o que resulta em mudanças no comprimento e de tensão das pregas vocais.

A cartilagem cricóide tem o formato de um anel de sinete. A parte posterior, sinete, da cartilagem é a lâmina, e parte anterior, aro, é o arco. Embora muito menor que a cartilagem tiróide, a cartilagem cricóide é mais espessa e mais resistente, e é o único anel completo de cartilagem a envolver a via aérea.

As cartilagens aritenóides são um par de estruturas em forma de pirâmide, com três faces, dois processos, uma base e um ápice, que se articulam com a superfície superior do arco da cartilagem cricóide.

O ligamento vocal, elástico, estende-se desde a comissura anterior da cartilagem tiróide, anteriormente, até ao processo vocal da cartilagem aritenóide, posteriormente. O ligamento vocal é formado pela margem superior, espessada e livre, do ligamento cricotiróideu, que faz parte do cone elástico. Este ligamento mistura-se anteriormente com o ligamento cricotiróideu mediano, também ele parte do cone elástico.

A cartilagem epiglótica tem uma constituição fibrocartilagínea, e é uma cartilagem cordiforme, recoberta por mucosa. Situada posteriormente à base da língua e ao osso hióide, e anteriormente ao ádito da laringe, a cartilagem epiglótica forma a região superior da parede anterior da laringe e a margem superior do ádito da laringe. A grande extremidade superior da cartilagem é livre e a sua extremidade inferior, afilada, o pecíolo, liga-se ao ângulo formado pelas lâminas tiróideias pelo ligamento tiroepiglótico.

As cartilagens corniculada e cuneiforme são pequenos nódulos situados na parte posterior das pregas ariepiglóticas. As cartilagens corniculadas situam-se nos ápices das cartilagens aritenóides; as cuneiformes não se ligam directamente a outras cartilagens.

A cavidade da laringe estende-se do ádito da laringe, através do qual se comunica com a faringe, até ao nível da margem inferior da cartilagem cricóide. Aqui, a laringe é contínua com a traqueia. A laringe divide-se em três partes: a supraglote que se localiza entre o ádito da laringe e as pregas vocais; nesta região inclui-se o vestíbulo da laringe – localizado superiormente às pregas vestibulares e o ventrículo da laringe, que está situado entre as pregas vestibulares e as pregas vocais; a glote, a região mais estreita da laringe, delimitada entre as pregas vocais; e a subglote, que compreende a região entre as pregas vocais e a à margem inferior da cartilagem cricóide, onde é contínua com o lúmen da traqueia.

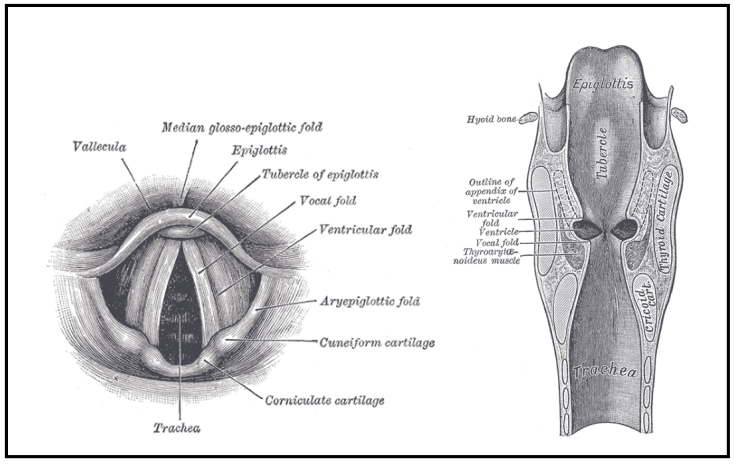

Figura 2.2 - *Interior da laringe – esquerda. Secção coronal da laringe – direita. De Williams P. et al [68].*

As pregas vocais controlam a produção do som. Cada prega vocal é constituída por: um ligamento vocal que consiste em tecido elástico espessado que é a margem livre medial do ligamento cricotiróideu lateral ou cone elástico; um músculo vocal cujas fibras, excepcionalmente finas, formam a maior parte da porção medial do músculo tiroaritenóideu; e pela mucosa que recobre a região mais medial.

A glote, região situada entre as pregas vocais, é o aparelho vocal da laringe. O formato da fenda glótica varia de acordo com a posição das pregas vocais. Durante a respiração, a glote é estreita e cuneiforme, enquanto que durante a respiração forçada é larga e em forma de pipa. Esta região é semelhante a uma fenda quando as pregas vocais estão intimamente próximas durante a fonação.

As pregas vestibulares, falsas pregas vocais, estendem-se entre as cartilagens tiróide e aritenóides, e tomam pouca ou nenhuma parte na produção da voz normal – apresentando maioritariamente uma função protectora. Consistem em duas grossas pregas de mucosa que recobrem os ligamentos vestibulares. O espaço entres estes ligamentos é a rima do vestíbulo. As indentações laterais entre as pregas vocais e as pregas vestibulares denominam-se ventrículos da laringe.

2.1.2 LARINGE MUSCULAR

Os músculos da laringe são divididos em dois grupos: extrínseco e intrínseco. Os músculos extrínsecos movem a laringe como um todo. Os músculos infra-hióideus movem o osso hióide e a laringe para baixo, enquanto que os músculos supra-hióideus e estilofaringeos são elevadores. Os músculos intrínsecos da laringe movem os diferentes componentes da laringe fazendo alterações no comprimento e tensão das pregas vocais e no tamanho e formato da glote. Todos os músculos intrínsecos da laringe, com excepção do músculo crico-tiroideu, são enervados pelo nervo laríngeo recorrente, um ramo do X nervo craniano, o vago; o músculo cricotiróideu é enervado pelo nervo laríngeo externo, um dos ramos terminais do nervo laríngeo superior.

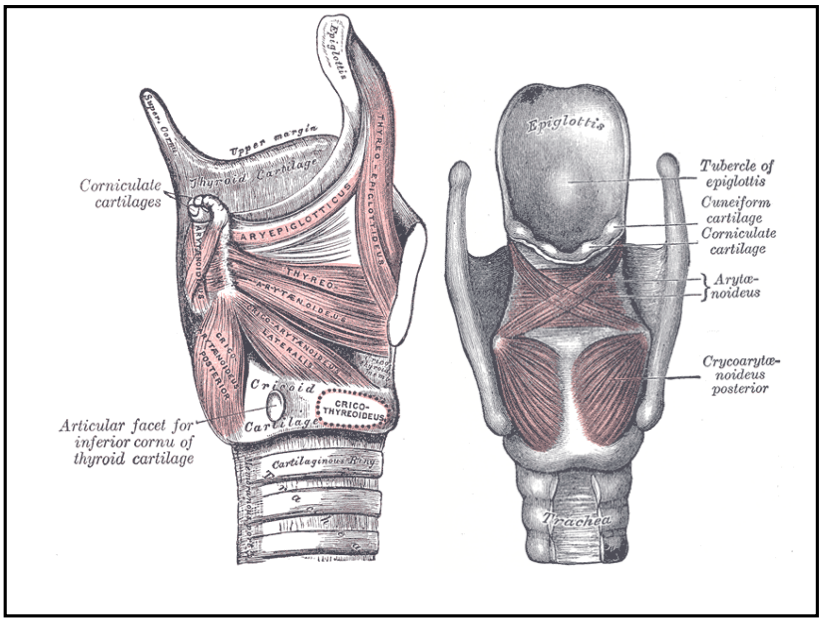

Figura 2.3 - *Musculatura da laringe – vista lateral – esquerda – e vista posterior – direita. De Williams P. et al [68].*

As acções dos músculos intrínsecos da laringe são de fácil compreensão quando considerados como um grupo funcional: esfíncterianos, adutores e abdutores, e tensores e relaxadores.

Os músculos adutores e abdutores são os que movem as pregas vocais para abrir e fechar a glote. Os principais adutores são os músculos cricoaritenóideus laterais que, provocando tracção nos processos musculares anteriormente, giram as cartilagens aritenóides de modo a que os seus processos se movam medialmente; quando esta acção é combinada com a dos músculos aritenóideus transversos, que produzem tracção em simultâneo nas cartilagens aritenóideias, o ar impelido através da glote produz vibrações das pregas vocais, traduzindo-se em fonação. Quando as pregas vocais são aduzidas, mas os músculos aritenóideus transversos não actuam, permitindo que as cartilagens aritenóides permaneçam algo separadas, o ar pode desviar-se das pregas – esta é a posição de murmúrio. Os únicos músculos abdutores são os músculos cricoaritenoideus posteriores que, ao produzirem tracção nos processos musculares posteriormente, giram os processos vocais lateralmente, ampliando assim a glote.

As acções combinadas da maioria dos músculos intrínsecos da laringe resultam numa acção esfincteriana que fecha o ádito da laringe como um mecanismo de protecção durante a deglutição. A contracção dos músculos cricoaritenoideus transversos e laterais**,** aritenoideus oblíquos e ariepiglóticos aduzem as pregas ariepiglóticas e produzem tracção nas cartilagens aritenóides em direcção à epiglote.

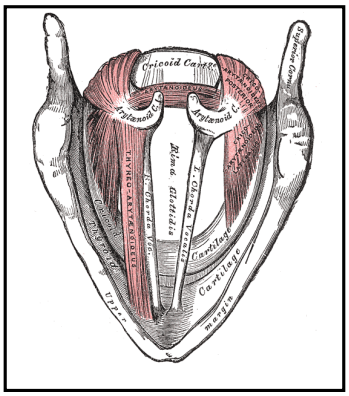

Figura 2.4 - *Músculos da laringe – vista interior. De Williams P. et al [68].*

Os principais *tensores* são os músculos *cricotiroideus*, que inclinam ou traccionam a proeminência ou ângulo da cartilagem tiróide para a frente e para baixo em direcção ao arco da cartilagem cricóide, aumentando a distância entre a proeminência laríngea e as cartilagens aritenóides. Como as extremidades anteriores dos ligamentos vocais se fixam na face posterior da proeminência tiroidea, os ligamentos vocais são alongados, aumentando a sua tensão, resultando numa elevação da frequência fundamental da voz.

Os principais *relaxadores* são os músculos *tiroaritenoideus*, que traccionam as cartilagens aritenóides para a frente, em direcção ao ângulo da cartilagem tiróide, relaxando assim os ligamentos vocais. Os músculos vocais produzem ajustamentos precisos dos ligamentos vocais, aumentando a tensão e relaxando, de modo selectivo, partes das pregas vocais durante a conversação e o canto.

2.1.3 MEMBRANA MUCOSA DA LARINGE

A membrana mucosa da laringe, é contínua com a da faringe e traqueia, estando frouxamente inserida na face anterior da epiglote e no revestimento das valéculas. Cobre as pregas ariepiglóticas e reveste a cavidade da laringe, formando as pregas vestibulares da laringe. Está firmemente inserida na face posterior da epiglote e nas faces laríngeas das cartilagens cuneiforme e aritenóide. A nível dos ligamentos vocais, a mucosa é fina e aderente.

De forma semelhante a algumas regiões localizadas acima da laringe, a face anterior da epiglote e a sua metade postero-superior, a região superior das pregas ari-epiglóticas e as pregas vocais, são recobertas por epitélio estratificado, escamoso, não queratinizante. Este tipo de epitélio confere uma maior resistência aos traumatismos mecânicos e térmicos. Noutros locais da laringe, o epitélio é do tipo pseudo-estratificado ciliado, também descrito como respiratório.

A mucosa da laringe possui numerosas glândulas mucosas, especialmente sobre a epiglote, onde elas escavam a cartilagem, e ao longo das margens das pregas ari-epiglóticas. As glândulas situadas no sáculo laríngeo, secretam periodicamente sobre as pregas vocais, durante a fonação, permitindo a sua lubrificação. As margens das pregas vocais estão desprovidas de glândulas, sendo assim vulneráveis aos efeitos da desidratação, requerendo, por isso, secreções de muco de glândulas vizinhas. As células epiteliais superficiais das pregas vocais apresentam microvilosidades e micropregas, consideradas como auxiliares na retenção das secreções superficiais.

Corpúsculos gustativos, semelhantes aqueles da língua, surgem sobre a face posterior da epiglote, nas pregas ariepiglóticas e, menos frequentemente, noutras regiões da laringe.

2.1.4 IRRIGAÇÃO DA LARINGE

As artérias da laringe são ramos das artérias tiroideias superior e inferior. A artéria laríngea superior acompanha o ramo interno do nervo laríngeo, através da membrana tiro-hioideia, e respectivos ramos, para suprir a face interna da laringe. O ramo crico-tiroideu é um pequeno ramo da artéria tiróideia superior que supre o músculo cricotiróideu. A artéria laríngea inferior acompanha o nervo laríngeo inferior, parte terminal do nervo laríngeo recorrente, e nutre a túnica mucosa e músculos situados na parte inferior da laringe.

As veias da laringe acompanham as artérias da laringe. De forma geral, a veia laríngea superior une-se à veia tiróideia superior, drenando para a veia jugular interna. A veia laríngea inferior une-se à veia tiroideia inferior ou ao plexo venoso de veias tiroideias, na face anterior da traqueia, que termina na veia braquiocefálica esquerda.

Os vasos linfáticos da laringe, acima das pregas vocais, acompanham a artéria laríngea superior através da membrana tiro-hioideia, e drenam para os gânglios linfáticos cervicais profundos superiores. Os vasos linfáticos abaixo das pregas vocais, drenam para os gânglios linfáticos cervicais profundos inferiores.

2.1.5 INERVAÇÃO DA LARINGE

Os nervos da laringe são os ramos laríngeos superior e inferior do nervo vago.

O nervo laríngeo superior tem origem no gânglio inferior do nervo vago, na extremidade superior do trígono carotídeo. Divide-se em dois ramos terminais: o ramo interno, sensitivo e autónomo, e o ramo externo, motor. O ramo interno é o maior dos ramos terminais, perfurando a membrana tiro-hioideia, com a artéria laríngea superior, e fornece fibras sensitivas para a mucosa da laringe acima das pregas vocais, incluindo a sua face superior. O ramo externo desce posteriormente ao músculo esterno-tiroideu,

acompanhando a artéria tiroideia superior. No início, o ramo externo situa-se no músculo constritor inferior da faringe e depois perfura e enerva este músculo assim como o músculo crico-tiroideu.

O nervo laríngeo inferior, a continuação do nervo laríngeo recorrente, entra na laringe profundamente à margem inferior do músculo constritor inferior da faringe. Divide-se em ramos anterior e posterior que acompanham a artéria laríngea inferior até à laringe. O ramo anterior enerva os músculos crico-tiroideu lateral, tiro-aritenoideu, ari-epiglotico e tiro-epigloticos. O ramo posterior enerva os músculos crico-aritenoideu posterior e aritenoideus transverso e oblíquo.

2.2 FALA DO PONTO DE VISTA ANATOMO-FUNCIONAL / MECÂNICA DA FONAÇÃO

Para a produção de fonação, a vibração das pregas vocais envolve a suspensão da respiração rítmica, a adução das cartilagens aritenóides, com ou sem reposicionamento da laringe, e a iniciação da de expiração voluntária. Assim, a fonação envolve três passos:

- Adução e aposição das cartilagens aritenóides na zona média e imobilização das porções terminais dos ligamentos vocais. A adução é conseguida primariamente pelos músculos inter-aritenoideus, enquanto que a imobilização se deve ao encerramento crico-aritenoideu, como consequência da geometria da articulação crico-aritenoideia.

- Alongamento e aumento de tensão nos ligamentos vocais. A regulação da tensão dos ligamentos vocais é o meio principal de controlo do tom e é conseguida por dois músculos antagonistas: o crico-tiroideu e o tiro-aritenoideu. A contracção do crico-tiroideu, enquanto o tiro-aritenoideu está inactivo, leva ao alongamento das pregas vocais, o que cursa com o aumento da frequência fundamental da voz.

- Passagem de ar expirado através da glote, produzindo vibração das pregas vocais. A passagem do ar é conseguida através dos músculos envolvidos na respiração.

Os músculos tiro-aritenoideus permitem ainda uma modulação isométrica mais fina na forma das pregas vocais.

2.3 CICLO GLOTAL

Para uma melhor compreensão do sistema de produção de fala é também importante conhecer o ciclo vibratório da laringe – a teoria mioelástica-aerodinâmica da fonação e a fala, do ponto de vista anatomo-funcional, já foram descritas anteriormente. Após o aumento de pressão subglótica, tem início a fase de abertura, que ocorre quando esta pressão é superior à resistência das pregas vocais fechadas, e dura até ao momento em que a resistência das pregas vocais ultrapassa a pressão do fluxo de ar. Neste momento inicia-se a fase de encerramento e as pregas vocais começam a mover-se uma em direcção à outra. Esta fase termina no momento em que a glote encerra, ou no máximo de oclusão possível, nos casos de uma fase de encerramento incompleta. Durante o resto do ciclo, na fase de oclusão, a glote encontra-se encerrada, enquanto a pressão subglótica aumenta progressivamente, para dar início a um novo ciclo.

São usualmente referidas duas razões para descrever de forma informativa o ciclo: o quociente de velocidade, SQ – *speed quotient,* que é a razão entre as durações da fase de abertura e a de encerramento; note-se que durante ambas as fases, a glote se encontra aberta; e o quociente de abertura, OQ – *open quotient*, que é a razão entre a duração do período em que a glote se encontra aberta e aquela do ciclo completo.

Numa voz normal, após o período em que a glote se encontra aberta, ocorre uma fase em que as pregas vocais se encontram unidas, tendo então lugar a fase de oclusão, durante a qual há aumento da pressão subglótica. Contudo, em alguns tipos de voz, a glote nunca oclui completamente, dando origem a uma fase de encerramento mais longa, nunca ocorrendo a fase de oclusão. Nestes casos, SQ é mais informativo do ciclo glótico que OQ.

A fase de oclusão é o período durante o qual as pregas vocais absorvem a força do impacto da sua colisão. Presume-se que a duração do período de oclusão depende da intensidade com que as pregas vocais colidem e da fracção desta força que é de facto absorvida pelas pregas, quanto mais elásticas, mais força absorvem e mais tempo permanecerão unidas.

As pregas vocais variam em elasticidade e em complexidade do padrão vibratório. A fase de abertura é iniciada de dentro para fora e de baixo para cima. A abertura inicia-se na sua base em direcção ao topo, e o encerramento ocorre no sentido inverso. Tal facto cria a chamada diferença de fase vertical. Tipicamente, as pregas vocais também abrem de trás para a frente e fecham no sentido inverso, havendo contudo casos de indivíduos em que este movimento se observa estar invertido.

Outro padrão vibratório complexo está relacionado com a membrana mucosa que cobre as verdadeiras pregas vocais. Esta membrana, frouxamente fixada, tende a ondular de forma oscilatória durante a fonação. De facto, crê-se que a mucosa vibra mais que o próprio ligamento vocal, podendo supor-se que esta é mais importante para a fonação que o próprio ligamento, que servirá apenas para ajustar a forma das pregas vocais. Lucero, J. [69]

2.4 TÉCNICAS DE ANÁLISE DA FUNÇÃO DA LARINGE

Estão disponíveis várias técnicas de análise da função laríngea, cada uma com vantagens e desvantagens, sendo, frequentemente, complementares. Segue-se um rápido resumo e descrição das principais técnicas. A única técnica descrita com maior detalhe é a electroglotografia, já que foi utilizada no âmbito desta dissertação.

2.4.1 LARINGOSCOPIA – ENDOSCOPIA

A laringoscopia é a técnica clinicamente mais eficiente para a observação das pregas vocais, Deliyski, D. e Petrushev, P. [70], e envolve uma medição óptica do movimento das mesmas. A laringoscopia directa envolve a observação por intermédio

de um fibroscópio flexível ou endoscópio rígido, enquanto que o método indirecto requer a introdução de um espelho pela cavidade oral, permitindo a avaliação da faringe e laringe. Como em qualquer técnica de endoscopia, podem adaptar-se câmaras e realizar a observação em monitores, passando a denominar-se nestes casos de videolaringoscopia.

2.4.2 ESTROBOSCOPIA E CINEMATOGRAFIA DE ALTA VELOCIDADE

A estroboscopia produz um vídeo composto a partir de várias capturas, com elevada frequência, de imagens do ciclo glótico, conseguidas pela iluminação da laringe por breves pulsos de luz, com uma frequência ligeiramente inferior à taxa de vibração – cerca de 2 Hz inferior – de forma a permitir observar o movimento com sub-amostragem, introduzindo aliasamento, e assim mostrando uma espécie de versão lenta do movimento. Na realidade, este tipo de observação não é uma versão lenta mas sim uma sub-amostragem, porque decorre ao mesmo tempo que a verdadeira vibração. Todavia, admitindo que a vibração é relativamente estacionária, pode considerar-se que a observação permitida é muito próxima da desejada versão lenta da vibração. A observação pode ser realizada por via trans-oral, com recurso a um laringoscópio rígido, ou através de um nasofaringoscópio flexível. Fornece uma imagem relativamente precisa do padrão de vibração regular. No entanto, na presença de uma vibração irregular, torna-se muito difícil o ajuste do tempo do flash estroboscópico ao período vibratório. Pulakka, H. [71]

A cinematografia de alta velocidade é um método muito próximo da estroboscopia, variando apenas no facto da laringe estar sempre iluminada por uma luz muito clara e as imagens serem capturadas a uma frequência 20 a 30 vezes superior à da frequência fundamental de vibração – contudo, câmaras modernas conseguem captar até 10000 *frames* por segundo. Tal facto torna esta técnica uma ferramenta mais fidedigna, comparativamente com a estroboscopia, para a análise de padrões vibratórios irregulares, tendo como desvantagens o seu custo muito elevado; e impossibilidade de realizar registos acústicos devido ao ruído produzido pelo sistema. Pulakka, H. [71]. Na Figura

2.5 é possível observar uma imagem da laringe obtida por cinematografia de alta velocidade.

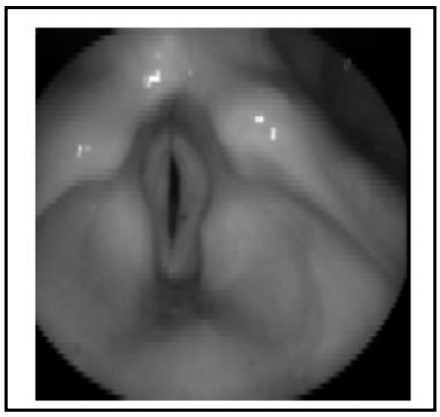

Figura 2.5 – *Imagem de laringe em fonação, obtida por cinematografia de alta velocidade. De Pulakka, H. [71].*

2.4.3 TRANSILUMINAÇÃO E FOTOCONDUÇÃO

Na iluminação trans-glótica, insere-se uma fonte luminosa através da fossa nasal e da nasofaringe. A luz obtida no outro lado das pregas vocais é medida por uma célula fotossensível colocado externamente na região cervical anterior. Assume que na fase de abertura do ciclo glótico, mais luz irá ser transmitida e tal se reflecte na forma de onda. Este método permite uma boa medição da actividade das pregas vocais mas é relativamente invasiva e torna-se difícil manter a fonte luminosa e a célula fotossensível em posição correcta durante um período de tempo adequado. Henrich, N. *et al* [72].

2.4.4 MEDIÇÃO DE FLUXO DE AR

O fluxo de ar trans-glótico pode ser estimado a partir de registos do fluxo de ar emitido pela boca, através de uma máscara de fluxo de ar especial com um transdutor de pressão diferencial integrado. Uma vantagem desta técnica é a obtenção de medidas absolutas do fluxo de ar glótico, não possíveis quando utilizada a forma de onda de pressão do sinal de fala; esta técnica será descrita mais detalhadamente no capítulo 4.3.1.

2.4.5 GLOTOGRAFIA ULTRASÓNICA

Na glotografia ultrasónica é utilizada a variação da frequência ultra sónica de Doppler como meio de monitorizar continuamente a velocidade de movimento das pregas vocais. Um feixe ultra sónico manter separado ou numa só palavra contínuo é direccionado para o pescoço e reflectido em vários tecidos e na interface tecido-ar da parede interna da laringe. Este processo torna-se possível, dado que a diferença de impedância acústica entre os tecidos e o ar circundante é tão grande que a transmissão de ultra sons do tecido para o ar é praticamente negligenciável. O sinal reflectido é comparado com o sinal transmitido de modo a obter a diferença frequencial, que surge no monitor, e é proporcional à velocidade da superfície reflectora. Depois de determinar a velocidade das pregas vocais em função do tempo, é possível integrar a área abaixo da curva e determinar o deslocamento da superfície reflectora como uma função do tempo. Mitra, P. [73]

2.4.6 ELECTROGLOTOGRAFIA – GLOTOGRAFIA DE IMPEDÂNCIA

Inicialmente desenvolvida por Fabre [74], em 1956, a electroglotografia, EGG, é uma técnica utilizada para registar o comportamento laríngeo de forma indirecta através da medição da impedância eléctrica, ao longo da região cervical, durante a fonação.

Aparelhos comercialmente disponíveis são produzidos por Laryngograph, Reino Unido, Glottal Enterprises, EUA, Tiger DRS, Inc., EUA e F-J Electronics, Dinamarca. A Figura 2.6 mostra um exemplo de um Electroglotografo – tendo sido este o aparelho utilizado nos procedimentos experimentais, descritos no capítulo 5.7.1.

Figura 2.6 - *Electroglotografo EG-PC3 da Tiger DRS, Inc. Unidade e eléctrodos em banda flexível.*

Masarek, K. [75] providencia uma descrição detalhada do funcionamento do aparelho de EGG bem como da informação que é possível recolher a partir deste. A base de funcionamento do electroglotógrafo é a medição da impedância do trajecto definido entre os dois eléctrodos situados na superfície da região anterior do pescoço ao nível da cartilagem tiróide através de uma corrente eléctrica de alta frequência e mas de baixas tensão e intensidade, que passa entre os eléctrodos. Os eléctrodos podem ser de cobre, prata e ouro e são geralmente em forma de anel, cobrindo uma área de 3 a 9 cm^2. Um terceiro eléctrodo pode ser utilizado como referência. Os eléctrodos são usualmente montados numa banda flexível cujo comprimento pode ser ajustado de modo a fixa-los numa posição estável, permitindo simultaneamente a fala e a respiração, de forma normal e confortável. Um gerador de sinal alimenta os eléctrodos com corrente sinusoidal com uma frequência geralmente na gama de 1 a 5 MHz. Esta frequência é suficientemente alta para que a corrente ultrapasse capacitivamente a camada menos condutora da pele sem necessidade de pasta ou gel condutor. A corrente aplicada varia em função do aparelho em particular, mas não é superior a alguns miliampere. A tensão entre os eléctrodos dependerá da impedância do tecido, mas o valor típico é de cerca de 0.5 volts e a dissipação de potência que ocorre ao nível das pregas vocais do sujeito não ultrapassa alguns microwatts.

A percentagem de modulação de amplitude do sinal recebido reflecte a percentagem da variação da impedância do tecido na passagem da corrente. O sinal recebido é então desmodulado por um detector existente no circuito. A forma de onda desmodulada sofre uma conversão A/D e é transmitida digitalmente e armazenada num computador. Adições frequentes à configuração padrão consistem na utilização de

instrumentos para medição da amplitude do sinal, por exemplo, um LED, ou para a medição da simetria do sinal, mostrando o equilíbrio entre os sinais dos dois eléctrodos. Tal é bastante conveniente para o correcto posicionamento dos eléctrodos.

A rápida variação da condutância é causada principalmente pelo movimento das pregas vocais. À medida que se separam, a impedância eléctrica transversal é aumentada devido ao facto da impedância do ar ser muito maior que a impedância do tecido. À medida que as pregas vocais se aproximam e o contacto entre as duas aumenta, a impedância vai diminuindo, o que resulta num fluxo de corrente relativamente crescente através das estruturas da laringe. No momento de máximo contacto, a diminuição é de cerca de 1%, até 2%, do total da condutância da laringe. Trata-se de uma modulação pequena e por esse facto relativamente difícil de detectar, tanto mais que existe também ruído eléctrico que se pode confundir com modulação. De acordo com Childers, D e Larar J. [76], a razão para o efeito de modulação da corrente deve-se ao facto de existir uma passagem de tecido mais longa para a corrente quando a glote está aberta, já que a impedância total do tecido é uma função do comprimento do tecido de passagem, a resistência paralela combinada é menor que a resistência de qualquer outro caminho, logo é razoável postular que a impedância do tecido observada no EGG é inversamente proporcional à área de contacto lateral entre as pregas vocais. No entanto, não é possível afirmar que o EGG meça a área de abertura glótica mas que apenas se encontra relacionado com esta abertura.

A amplitude do sinal varia devido à permanente variação do contacto das pregas vocais, e depende: configuração e posicionamento dos eléctrodos; contacto eléctrico entre os eléctrodos; posição da laringe e das pregas vocais na região cervical; estrutura da cartilagem tiróide; quantidade e proporção de tecido muscular, glandular e adiposo que rodeia a laringe; assim como a distância entre os eléctrodos.

Segundo Masarek, K. [75] pode ocorrer que a flutuação da impedância causada pelos movimentos das pregas vocais seja muito fraca para ser registado e que se deve ter em conta que os sinais de EEG, de qualidade aceitável, são mais difíceis de obter em mulheres e crianças do que nos homens. Este facto está relacionado com a menor massa das pregas vocais, do ângulo mais amplo da cartilagem tiróide, e da diferente proporção entre diferentes tipos de tecido.

Apesar dos problemas acima descritos, a electroglotografia estabeleceu-se como um método valioso para a avaliação da função laríngea. Em comparação com outros métodos glotográficos, Frokjaer-Jensen, B. [77] observou que a electroglotografia permite uma melhor representação das fases de encerramento e de abertura do movimento das pregas vocais, especialmente da área de contacto vertical. A fotoglotografia parece ser particularmente vantajosa no que diz respeito à representação da fase abertura. A EGG é superior a todos os outros métodos relativamente ao conforto do sujeito analisado, já que é um método não-invasivo, não exercendo, em princípio, qualquer influência no modo de articulação e produção de som.

2.4.7 GLOTOGRAFIA ELECTROMAGNÉTICA

A glotografia electromagnética, EMGG, é um método próximo da electroglotografia, em que se utiliza a propagação transversal de uma onda electromagnética, na gama dos GHz, através do pescoço, que é detectada, para obtenção de informação, sobre a interface dos tecidos da laringe. A antena emissora é geralmente posicionada em frente à proeminência tiroideia. Mitra, P. [73]

2.4.8 OUTRAS TÉCNICAS

Outras técnicas de estudo da função e anatomia da laringe incluem as técnicas de imagiologia como a radiografia, a Tomografia Axial Computorizada, TC, e a Ressonância Magnética, RM.

Outros métodos complementares de estudo são ainda a Electromiografia, EMG, e a análise comparativa da pressão sub e supraglótica.

2.5 MODELAÇÃO DA FONTE GLOTAL

De acordo com Fujisaki, H. e Ljungqvist, M. [78], podem dividir-se os modelos da fonte vocal em dois grandes grupos: interactivos e não-interactivos. Um modelo interactivo é aquele onde a forma de onda é computada indirectamente a partir de: área glótica, com recurso a dados provenientes da fotoglotografia; condutância eléctrica via electroglotografia; e modelos mecânicos e aerodinâmicos de vibração das pregas vocais. Os modelos interactivos requerem um conhecimento da fisiologia da glote e das interacções entre a fonte e o tracto vocal. Um modelo não-interactivo é aquele onde o fluxo glótico, ou a derivada do fluxo, como função do tempo, é parametrizado directamente – modelos acústicos. Estes modelos podem ou não estar relacionados directamente com a fisiologia da laringe e podem ou não conter informação da interacção entre a fonte e o tracto vocal.

2.5.1 MODELOS INTERACTIVOS

Como as técnicas de Electroglotografia e Fotoglotografia já foram descritas brevemente, focam-se aqui os modelos aerodinâmicos-mecânicos.

2.5.1.1 MODELOS AERODINÂMICOS-MECÂNICOS

Childers, D. [79] apresenta uma descrição detalhada de vários modelos aerodinâmicos-mecânicos.

Os modelos mecânicos são considerados como tal já que representam a fonte glótica como um conjunto de osciladores mecânicos agregados, destacando-se os modelos "uma massa", *one-mass*, "duas massas", *two-mass*, e "massas múltiplas", *multiple-mass*. No modelo mecânico unificado, o sistema subglótica é representado por um reservatório de ar com pressão, P_g, que cria um fluxo de ar com uma velocidade volumétrica, U_g. As pregas vocais são mecanicamente modeladas como um sistema oscilatório de massas com atenuação viscosa e molas.

2.5.1.1.1 MODELO DE UMA MASSA (ONE MASS)

O modelo de uma massa foi o primeiro modelo matemático da dinâmica oscilatória das pregas vocais, e foi proposto por Flanagan, J. e Landgraf, L. [80] em 1968. Nele, cada prega vocal é representada por um sistema massa-atenuador-mola – tal como se pode ver na Figura 2.7. Assume-se que as pregas vocais são simétricas e que apenas existe movimento na direcção horizontal. A equação do sistema pode ser descrita como:

$$m\ddot{x} + r\dot{x} + kx = dl_g P_g(x) \tag{2.1}$$

Onde m representa a massa, r o coeficiente de atenuação, k o coeficiente de rigidez, x o deslocamento, d_g e l_g a largura e comprimento da prega vocal respectivamente, e $P_g(x)$ a pressão de ar glótica. A pressão depende da área seccional da glote e é portanto uma função do seu deslocamento x.

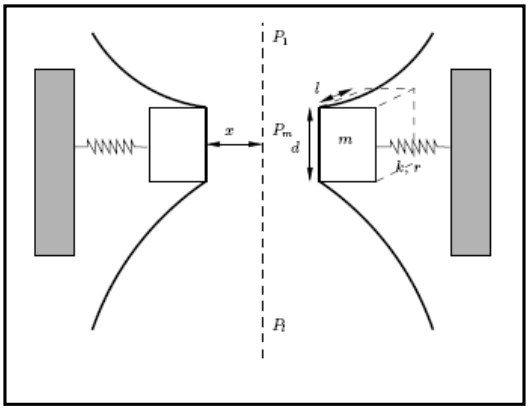

Figura 2.7 - *Modelo de Uma Massa. De Drioli, C. [81].*

De acordo com Childers D. [79] as vantagens deste modelo são: simplicidade; leveza computacional; poder contemplar a interacção entre a fonte e o tracto vocal. Apesar de poder ser utilizado como um modelo simples da oscilação da fonte em sintetizadores simplificados, não capta o principal mecanismo de oscilação das pregas vocais nem considera a diferença de fase entre os movimentos das extremidades da prega vocal.

2.5.1.1.2 MODELO DE DUAS MASSAS (TWO MASS)

O modelo de duas massas foi proposto por Ishizaka, K. e Flanagan, J. [82], também designado de modelo IF, e descreve cada prega vocal recorrendo a uma aproximação a duas massas – tal como se vê na Figura 2.8. O modelo assume que as pregas vocais são simétricas, sendo apenas necessário modelar uma delas, e que as massas têm movimento lateral apenas. Ao longo da direcção lateral assume-se também que as massas se comportam como osciladores mecânicos de segunda-ordem sujeitas a forças elásticas e de dissipação. Para uma modelação correcta das propriedades elásticas da prega vocal, as molas são não-lineares, havendo ainda uma terceira mola que ligas as duas massas. A colisão entre as pregas é modelada pela consideração nas equações de uma força de retorno adicional, representada por uma outra mola não-linear equivalente, assim quando uma das massas colide a sua rigidez aumenta. As equações mecânicas do sistema são dadas por:

$$\begin{cases} m_1\ddot{x}_1(t) + r_1\dot{x}_1(t) + k_1(x_1)\left[x_1(t) - x_{01}\right] + k_{12}\left[x_1(t) - x_2(t)\right] = l_g d_1 p_{m1}(t) \\ m_2\ddot{x}_2(t) + r_2\dot{x}_2(t) + k_2(x_2)\left[x_2(t) - x_{02}\right] + k_{12}\left[x_1(t) - x_2(t)\right] = l_g d_2 p_{m2}(t) \end{cases} \quad (2.2)$$

Onde x_i representa o deslocamento lateral das duas massas, m_i as massas, r_i as atenuações para cada massa ,neste caso a resistência provém de perdas por viscosidade e fricção, k_i as constantes das molas, $l_g d_i$ são o comprimento e largura das superfícies sobre as quais actuam as pressões p_{m1} e p_{m2}. Os autores afirmam que a utilização de duas massas possibilita um grande detalhe, incluindo a modelação das diferenças de fase no movimento das porções terminais superior e inferior das pregas vocais.

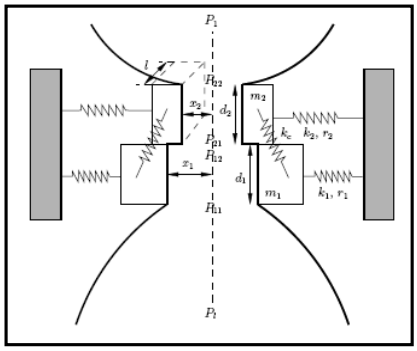

Figura 2.8 - *Modelo de Duas Massas. De Drioli, C. [80].*

De acordo com Childers, D. [79] o modelo de duas massas tem os seguintes méritos: é uma simulação mais realística das propriedades glotais; a diferença de fase entre o movimento das extremidades da prega vocal é considerada – a onda da mucosa não é considerada, mas é-o no modelo modificado de Koizumi, T. et al. [83]; e permite produzir uma fala natural com um custo computacional razoável.

Há contudo limitações ao modelo. Por exemplo, as variações das dinâmicas das pregas vocais ao longo do comprimento da glote não podem ser simuladas pois o modelo só tem em consideração as diferenças verticais.

2.5.1.1.3 OUTROS MODELOS DE MASSAS MÚLTIPLAS

Childers, D. [79] nota que apesar do modelo IF ser um marco na quantificação da vibração das pregas vocais, modela apenas as pregas como um estrutura mecânica mínima capaz de responder a forças aerodinâmicas e sustentar a oscilação, não permitindo a exibição vários modos de vibração longitudinal exibidos na fonação humana normal.

Hirano, M. [84] introduziu o conceito de corpo-revestimento para descrever a estrutura laminar das pregas vocais que teria grande influência sobre todos os modelos mecânicos propostos posteriormente. O autor sugere que as pregas vocais podem ser divididas em duas camadas de tecido com diferentes propriedades mecânicas. A camada do corpo consiste nas fibras musculares e algumas fibras de colagénio fortemente ligadas ao ligamento vocal, e a camada de revestimento consiste em tecido dúctil, não-contráctil que actua como uma cobertura flexível à volta da camada do corpo. O revestimento está tipicamente ligado de forma frouxa ao corpo durante a vibração. O movimento da camada de revestimento é geralmente observado como uma onda de superfície que se propaga do fundo das pregas vocais para o topo, evidenciando assim um movimento em ambas as direcções lateral e vertical. A oscilação auto-sustentada das pregas vocais é altamente dependente deste comportamento de onda de superfície, também referido como a diferença de fase vertical, e pensa-se que será o mecanismo primário de transferência de energia do fluxo glótico para o tecido, alimentando a vibração. A camada do corpo está primariamente envolvida no movimento lateral.

Baseado nas suas descobertas, Hirano, M. sugeriu que as pregas vocais deveriam ser tratadas como um vibrador duplamente estruturado com parâmetros de rigidez que deveria ser baseado nas acções relativas dos músculos tiroaritenóideu e cricotiróideu – tal como se pode ver na Figura 2.9. Assim, a vibração resultante das pregas vocais é composta por oscilações agrupadas das camadas do corpo e do revestimento.

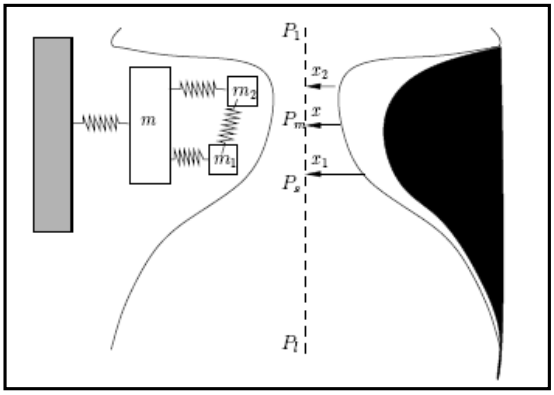

Figura 2.9 - *Modelo de Corpo-Revestimento. De Drioli, C. [80].*

Titze, I. [85], numa tentativa de alargar os graus de liberdade horizontais, propôs um modelo de 16 massas, *16 mass-model*, composto por duas fileiras de 8 massas cada. A fileira de massas superior representa primariamente a membrana mucosa e a fileira inferior representa primariamente o ligamento e o músculo vocal. As forças T_m e T_v representam as tensões longitudinais como determinadas pelo equilíbrio das forças entre os músculos cricotiróideu e tiroaritenóideu. Especificamente, as constantes das molas das fileiras superiores e inferiores aumentam não linearmente com o alongamento das pregas vocais. O modelo de 16 massas tem as seguintes características, segundo Childers, D. [79]: é um modelo complexo com alto custo computacional; a onda de superfície da mucosa pode ser simulada; tem a capacidade de ser regulado através de parâmetros que possuem correlação directa com parâmetros fisiológicos; verifica-se um aumento de naturalidade nas elocuções, e a fonação é possível em pelo menos dois registos diferentes.

Num outro modelo, Titze, I. e Strong, W. [86] inovaram em relação aos outros modelos mecânicos até então propostos ao representarem as pregas vocais, não como um conjunto discreto de massas agregadas mas como um meio deformável contínuo. A incompressibilidade das pregas vocais dita um acoplamento entre o movimento

horizontal e vertical. Uma consequência importante da incompressibilidade das pregas vocais é que o modo vibratório mais facilmente excitado parece envolver diferenças de fase verticais, já que este modo tende a preservar o volume das pregas vocais. O estudo mostrou também que a estrutura em camadas das pregas vocais está idealmente adaptada ao suporte da vibração das pregas. A estrutura fibrosa longitudinal é mais solta na direcção vertical do que na longitudinal. Tal facto permite que ocorram as diferenças de fase verticais. O modelo contínuo é altamente informativo no que diz respeito às relações entre a estrutura das pregas vocais e os modos vibratórios das mesmas. No entanto, a forma das pregas vocais neste modelo está restrita à forma rectangular. Adicionalmente, está ausente do modelo uma representação completa das interacções entre o fluxo de ar aerodinâmico e o tecido elástico das pregas vocais devido à derivação dos modos de vibração das pregas baseados em análise de valores próprios do tecido das pregas.

Titze I., [87] propôs ainda o modelo da onda da mucosa onde melhora o sistema massa-atenuação-mola adicionando uma onda superficial que se propaga na direcção do fluxo do ar. Esta onda reproduz observações da oscilação das pregas vocais referidas atrás. O modelo pode ser descrito como:

$$m\ddot{x} + r\dot{x} + kx = dl_g \frac{2\tau P_s \dot{x}}{x_0 + x + \tau\dot{x}} \tag{2.3}$$

Onde τ é o atraso no tempo da onda da mucosa que viaja ao longo da glote. Este modelo permitiu perceber que é esta onda que permite às pregas vocais absorver energia do fluxo do ar, permitindo depois a sua oscilação.

De modo a representar mais realisticamente a estrutura corpo-revestimento das pregas vocais, Story, B. e Titze, I. [88] estenderam o modelo de duas massas a um modelo de três massas. O modelo de três massas consiste em duas massas de revestimentos acopladas lateralmente a uma massa de corpo por molas não-lineares e elementos de atenuação viscosa. A massa do corpo, que representa tecido muscular, está também acoplada lateralmente a uma parede rígida, que se assume representar a cartilagem tiróide, por uma mola não linear e um elemento de atenuação. As duas molas do revestimento pretendem representar as propriedades elásticas do epitélio e da lâmina

própria, enquanto que a mola do corpo simula a tensão produzida pela contracção do músculo tiroaritenóideu. Assim, as contracções dos músculos cricotiróideu e tiroaritenóideu são incorporadas nos valores utilizados para os parâmetros de rigidez das molas do corpo e do revestimento. As duas massas do revestimento estão acopladas uma à outra através de uma mola linear, que pode representar a propagação da onda vertical da mucosa.

De uma forma generalizada, e tal como definido para os modelos interactivos, os modelos referidos acima são controlados por parâmetros que estão relacionados com os parâmetros de controlo fisiológico de produção da fala, com as devidas afinações, de modo a produzirem formas de onda glotais realistas. Estes estudos, associados a outros da medicina, permitiram perceber que o principal movimento de abertura e fecho das pregas vocais acontece na direcção horizontal, perpendicular ao fluxo de ar, mas não é suficiente para descrever a dinâmica oscilatória das mesmas. Para obter modelos realistas é necessário incluir a onda da mucosa que é responsável pela transferência da energia do fluxo de ar para as pregas vocais alimentando a sua oscilação.

Note-se também que as condições para iniciar a oscilação são mais restritas que as condições para a manter após inicialização.

2.5.2 MODELOS NÃO INTERACTIVOS

De modo a facilitar a análise do sinal fonte e permitir uma eficiente caracterização com recurso a um conjunto pequeno de parâmetros, os modelos acústicos provaram ser mais vantajosos. Estes modelos caracterizam o impulso glotal em termos de sinal de fluxo de ar ou em termos da sua derivada. Os impulsos modelados são gerados com base em funções paramétricas para cada fase diferente do ciclo glotal.

O modelo mais simples do fluxo glotal foi proposto por Markel, J. e Gray A. [89] onde a forma de onda da velocidade volumétrica glotal era a saída de um filtro passa-baixo com dois pólos e uma frequência de corte próxima dos 100 Hz. Apesar de simplista, o modelo consegue aproximar a velocidade volumétrica real no domínio das frequências, e permite a integração num modelo só com pólos do tracto vocal tornando

facilitado o processo de estimação de parâmetros pelos métodos de predição linear, ver abaixo. Por outro lado, a aproximação à velocidade volumétrica real no domínio do tempo é relativamente fraca.

Segue-se a descrição de alguns dos modelos mais importantes, tendo também em vista uma perspectiva histórica.

2.5.2.1 MODELO DE ROSENBERG

Rosenberg, A. [90] propôs vários modelos para o impulso glotal com vários parâmetros ajustáveis. Os impulsos deste modelo são constituídos por dois segmentos trigonométricos com um declive descontínuo na fase de fecho. Os modelos de Rosenberg possuem três parâmetros: a, amplitude; T_P, duração da fase de abertura; e T_N, duração da fase de fecho.

A equação de um dos modelos de Rosenberg (B):

$$u^{R(B)} = \begin{cases} a\left[3\left(\frac{t}{T_P}\right)^2 - 2\left(\frac{t}{T_P}\right)^3\right] & 0 \le t \le T_p \\ a\left[1-\left(\frac{t-T_P}{T_N}\right)^2\right] & T_P \le t \le T_P + T_N \end{cases} \tag{2.4}$$

A Figura que se segue mostra a representação gráfica de vários modelos de Rosenberg.

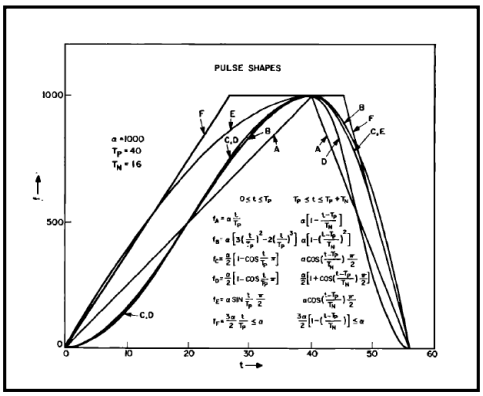

Figura 2.10 - *Modelos de Rosenberg. De Rosenberg, A. [90].*

2.5.2.2 MODELO DE FANT (OU MODELO F)

Fant, G. [91, 92] propôs um modelo do impulso glotal também com três parâmetros: U_0, pico do fluxo; F_g frequência glotal; e k, factor de assimetria.

$$u^L = \begin{cases} U = \frac{1}{2}U_0\left[1-\cos w_g (t-T_1)\right], & 0 < t < T_2 \\ U = U_0[k\cos w_g (t-T_2) - k + 1], & T_2 < t < T_3 \end{cases} \tag{2.5}$$

Onde

$$F_g = \frac{w_g}{2\pi}$$

O impulso inicia-se em $t = T_1$ atingindo o valor máximo, U_0, em $t = T_2$. A fase de retorno, descendente, inicia-se em $t = T_2$ e atinge o valor 0 após o período $T_3 - T_2 = \frac{1}{w_g} ar\cos(\frac{k-1}{k})$. Se $k > 0.5$ a terminação é abrupta sendo o declive da fase de retorno igual a $U_3^{'} = \frac{dU}{dt}_{(t=T_3)} = -U_0 w_g \sqrt{2k-1}$. Se $k = 0.5$ a fase de retorno é

simétrica em relação à fase ascendente, sendo valor mínimo para este modelo, já que um valor inferior a 0.5 implicaria uma fase de retorno inferior à fase ascendente.

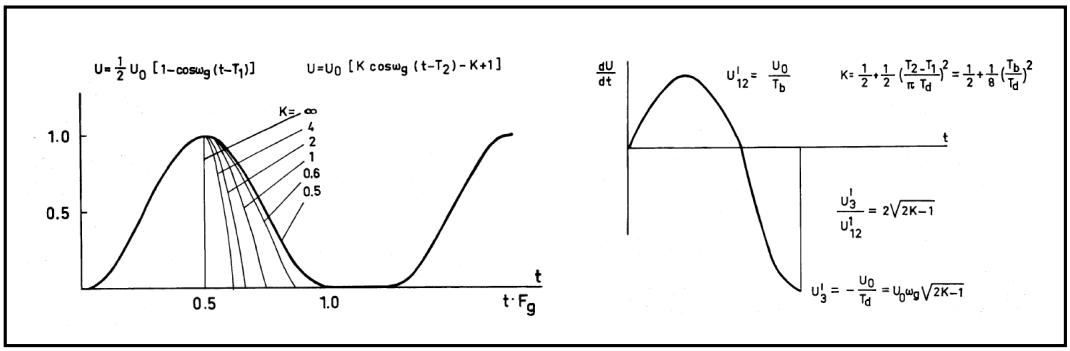

Figura 2.11 - *Modelo F – esquerda – e derivada do Modelo F – direita. De Fant, G. [91]*

Um dos principais defeitos deste modelo de três parâmetros é a descontinuidade abrupta na fase de oclusão glotal, não capturando a possibilidade de um oclusão incompleta ou uma fase residual que prossegue em direcção ao fecho após a descontinuidade. Por razões semelhantes aos modelos de Rosenberg, o modelo F tem pouca aplicação no processamento da fala.

2.5.2.3 MODELO DE ANANTHAPADMANABHA

Ananthapadmanabha, T. [93] propôs um modelo com cinco parâmetros variáveis que modela directamente a saída do filtro inverso – a derivada do impulso glotal – e não o seu integral – o impulso glotal. Outro facto importante é a parameterização do movimento de fecho glotal não abrupto, modelado aqui como uma função parabólica de retorno terminal, que permite uma monitorização indirecta do grau de abdução. Os parâmetros do modelo são *TP* , intervalo entre *T*0 estimado e *TP* como percentagem do período fundamental; *TN* **,** intervalo entre *TP* e *TE* como percentagem de do período fundamental; *TL* , base temporal da parábola como percentagem do período fundamental; *EI* , amplitude de uma meia sinusóide modelada entre *T*0 e *TP* ; *EE* , amplitude absoluta em *TE* .

$$du^A = \begin{cases} EI \sin(\frac{\pi}{2}\frac{t}{TP}) & 0 \le t \le TP/2, \qquad T0 < t < TI \text{ (ou } TP/2) \\ (EI + EE)\cos\left[\frac{\pi}{4}\frac{\hat{t}}{TN}\right] - EE, & TI < t < TE \\ -EE\left[\frac{\tilde{t}}{TL}\right]^2, & TE < t < TC \end{cases} \tag{2.6}$$

Onde $\hat{t} = t - TP/2$ e $\tilde{t} = [TP + TN + TL - T]$

A Figura 2.12 mostra a representação gráfica do modelo de Ananthapadmanabha.

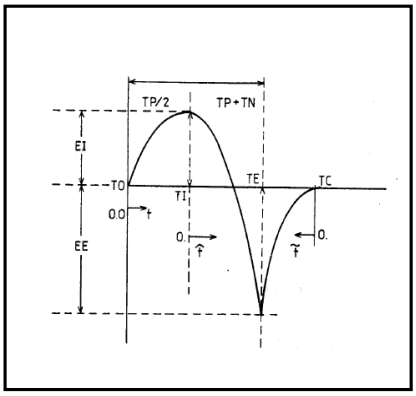

Figura 2.12 *Modelo de Ananthapadmanabha. De Ananthapadmanabha, T. [93]*

Este modelo não apresenta as desvantagens dos anteriores relacionadas com a descontinuidade na fase de oclusão. No entanto a forma do pulso não tem um pico marcado como no caso dos resíduos de AR, tornando-o menos desejável para representar a entrada de um modelo AR.

2.5.2.4 MODELO DE LILJENCRANTS-FANT (MODELO LF)

Fant, G. *et al* [94] propuseram aquele se tornou o mais popular e utilizado modelo da fonte vocal. No domínio do tempo, o modelo LF é constituído por duas secções: uma sinusóide com crescimento exponencial modela a fase aberta até ao primeiro contacto de colisão entre as pregas vocais, seguido por uma fase de retorno com decaimento exponencial. O modelo possui apenas 4 parâmetros: t_p, instante de máximo impulso glotal; t_e, instante do máximo valor negativo da derivada do impulso glotal, que

corresponde ao instante de maior fecho glotal; t_a, duração efectiva da fase exponencial de retorno; e E_e, amplitude máxima do instante de maior fecho glotal. Por conveniência, o parâmetro t_c é igualado ao período fundamental. O modelo LF é então dado por:

$$du^{LF}(t) = \begin{cases} E_0 e^{\alpha t} \sin w_g t, & t \le t_e \\ \dfrac{-E_e}{\beta t_a}\left[e^{-\beta(t-t_e)} - e^{-\beta(t_c-t_e)}\right], & t_e \le t_c \end{cases} \qquad (2.7)$$

Onde α e β satisfazem as seguintes equações:

Para t_a pequeno $\beta = 1/t_a$; senão $\beta t_a = \left[1 - e^{-\beta(t_c-t_e)}\right]$

A Figura 2.13 mostra a representação gráfica do modelo LF.

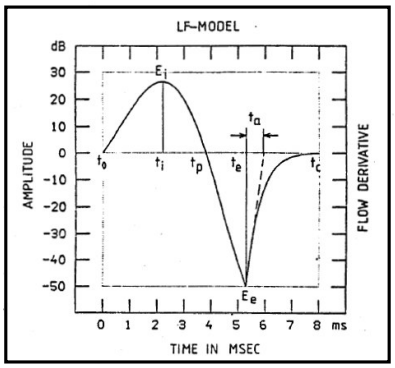

Figura 2.13 – *Modelo LF. De Fant, G. et al [94]*

Assumindo que o período fundamental é T_0 e que a taxa de amostragem do sistema é F_s, o tamanho de amostras de um período fundamental é $N_0 = \text{int}(T_0.F_s)$. Assim, os parâmetros correspondentes no domínio do tempo discreto são: $N_p = \text{int}(T_p.F_s)$, $N_e = \text{int}(T_e.F_s)$, $N_a = \text{int}(T_a.F_s)$ e $N_c = \text{int}(T_c.F_s)$, respectivamente. A versão discreta do modelo LF pode ser definida como:

$$du^{LF}(n)=\begin{cases} E_0 e^{\alpha n}\sin(w_g n), & 0\le n\le N_e \\ -\dfrac{E_e}{\varepsilon N_a}\left[e^{-\varepsilon(n-N_e)}-e^{-\varepsilon(N_c-N_e)}\right], & N_e\le n\le N_c \\ 0, & N_c\le n\le N_0-1 \end{cases} \tag{2.8}$$

Conjuntamente com a magnitude da excitação do fecho glotal E_e, o modelo LF pode ser especificado por dois conjuntos de parâmetros: os parâmetros de síntese directa, E_0, α, ω, ε, e os parâmetros temporais, N_p, N_e, N_a, N_c. Como os parâmetros temporais e E_e são facilmente identificados a partir da forma de onda glotal, obtêm-se primeiro os parâmetros temporais e E_e, derivando-se posteriormente os parâmetros de síntese directa com as seguintes restrições:

$$\begin{cases} \displaystyle\int_0^{T_0} v_g^{LF}(t)dt = 0 \\ \omega_g = \dfrac{\pi}{N_P} \\ \varepsilon N_a = 1-e^{-\varepsilon(N_c-N_e)} \\ E_0 = \dfrac{E_e}{e^{\alpha N_e}\sin(\omega_g N_e)} \end{cases} \tag{2.9}$$

A razão da popularidade do modelo LF reside na capacidade deste aproximar a forma de vários impulsos glotais utilizados em síntese e análise de fala através de um número mínimo de parâmetros, sendo também capaz de modelar variabilidades extremas de fonação, Fant, G. *et al* [94].

2.5.2.5 MODELO DE FUJISAKI-LJUNGQVIST (MODELO FL)

Fujisaki, H. e Ljungqvist, M. [78] analisaram alguns dos modelos anteriormente propostos que consideraram mais significativos, nomeadamente Rosenberg, Hedelin, Fant, Ananthapadmanabha, LF, e propuseram um novo modelo que pretendia conjugar as melhores características dos modelos analisados. O modelo proposto modela a derivada do impulso glotal sendo composto por segmentos polinomiais. A escolha de um modelo polinomial deve-se à facilidade de variação do número de parâmetros, e

consequentemente o nível de detalhe do modelo. Na sua forma mais elaborada, o modelo apresenta três parâmetros limitadores: W, duração da fase aberta que é igual $(R+F)$; S, pendente do impulso que é igual $(R+F)/(R-F)$; D, intervalo de tempo desde o fecho glotal até ao máximo fluxo negativo. Mais três parâmetros de amplitude: A, declive na abertura glotal; B, declive antes do fecho; C, declive posterior ao fecho. O modelo pode ser descrito como:

$$du^{FL}(t)=\begin{cases} A-\dfrac{2A+R\alpha}{R}t+\dfrac{2A+R\alpha}{R}t^2, & 0<t\le R \\ \alpha(t-R)+\dfrac{3B-2F\alpha}{F^2}(t-R)^2-\dfrac{2B-F\alpha}{F^3}(t-R)^3, & R<t\le W \\ C-\dfrac{2(C-\beta)}{D}(t-W)+\dfrac{C+\beta}{D^2}(t-W)^2, & W<t\le W+D \\ \beta, & W+D<t\le T \end{cases} \tag{2.10}$$

Onde: $\alpha=\dfrac{4AR-6FB}{F^2-2R^2}$, $\beta=\dfrac{CD}{D-3(T-W)}$ e T é igual ao período fundamental.

Na Figura 2.14 é possível ver a representação gráfica do modelo F.

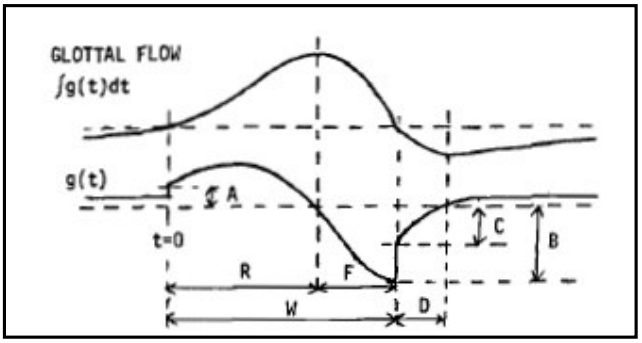

Figura 2.14 - *Modelo F. De Fujisaki, H. e Ljungqvist, M. [78]*

No mesmo artigo, e na consequência da análise dos modelos anteriormente propostos, os autores concluem que o modelo LF e o seu modelo são os que apresentam o melhor desempenho em termos de qualidade e naturalidade da voz produzida. Propõem ainda um método para a estimação automática e simultânea dos parâmetros do tracto vocal e daqueles do novo modelo descrito – descrito no capítulo 4.2.

2.5.2.6 MODELO KLGLOTT88 (MODELO ROSENBERG-KLATT ou MODELO RK)

Klatt, D. e Klatt, L. [95] propuseram um modelo, que também se tornou popular em processamento da fala. Neste é parametrizada a derivada do impulso glotal e a forma de onda é descrita por parâmetros convencionais como T_0, período fundamental; AV, amplitude máxima do impulso glotal; OQ, quociente de abertura; e TL, pendente espectral; permitindo ainda a determinação de outros parâmetros como: FL, flutuação período-a-período – flutuações quase aleatórias; e AH, amplitude do ruído de aspiração. A forma de onda da velocidade volumétrica é parametrizada de forma a obedecer à relação proposta por Rosenberg [90].

O modelo não considera uma fase de retorno na derivada assumindo que fecho glotal é abrupto, sendo portanto zero na fase de oclusão. Na fase aberta, a onda da derivada do impulso glotal é modelada por um polinómio de segunda ordem:

$$du^{RK} = \begin{cases} 2an - 3bn^2 & 0 \le n \le T_0 * OQ \\ 0 & T_0 * OQ < n \le T_0 \end{cases} \tag{2.11}$$

Onde, $a = \dfrac{27 * AV}{4 * (OQ^2 * T_0)}$ e $b = \dfrac{27 * AV}{4 * (OQ^3 * T_0^{\,2})}$

O parâmetro OQ varia entre 0 e 1, onde 0 representa a glote sempre fechada e 1 o oposto.

Apesar da forma de onda considerar um fecho abrupto no momento de fecho glotal, o parâmetro TL acaba por suavizar a fase de retorno – tal como se pode ver na Figura 2.15. O parâmetro *TL* pode ser aproximado por um filtro passa-baixo de primeira-ordem.

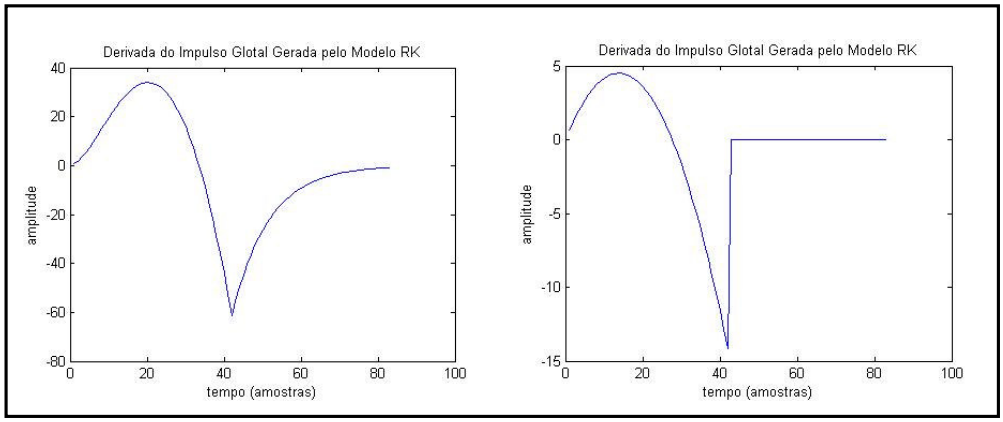

Figura 2.15 – *Modelo RK – direita – e Modelo RK após filtragem com parâmetro TL.*

Comparativamente com o modelo LF, e apesar de algumas semelhanças com o mesmo, o modelo RK é matematicamente mais simples e computacionalmente menos custoso, tornando-se um candidato ideal para processos de optimização que envolvam a identificação da fonte ou a aproximação a um sinal filtrado inversamente.

Em 1995, Fant, G. apresentou uma versão revista do modelo FL, Fant, G. [96].

2.5.2.7 OUTROS MODELOS

Outros modelos interessantes incluem: Rothenberg, M. [97]; Hedelin, P. [98]; Milenkovic, P. [99]; Childers, D.[100]; Rosenberg++ [101].

3 CARACTERIZAÇÃO ACÚSTICO-ANATÓMICA DO TRACTO VOCAL

O estudo desta secção justifica-se na medida em que é necessário explorar mais a possibilidade de a electrolaringe utilizar as propriedades acústicas do tracto vocal, especificamente, no caso de uma prótese de electrolaringe implantável.

3.1 ANATOMIA DO TRACTO VOCAL

A anatomia do tracto vocal encontra-se descrita de forma extensiva em vários livros de referência de Anatomia Médica, como por exemplo em Williams, P. [68].

3.1.1 FARINGE

A *faringe* localiza-se posteriormente às cavidades do nariz e da boca e estende-se da base do crânio até à margem inferior da cartilagem cricóide, anteriormente, e a margem inferior da vértebra C6, posteriormente. É mais larga, aproximadamente 5 cm, oposta ao osso hióide e mais estreita, cerca de 1,5 cm, na sua extremidade inferior, onde é contínua com o esófago. A parte posterior da faringe situa-se contra a lâmina pré-vertebral da fáscia cervical.

A faringe é dividida em três partes: a nasofaringe, com localização posterior ao nariz e acima do palato mole; a orofaringe, situada posteriormente à cavidade oral; e a hipoafaringe, posicionada na região posterior da laringe.

A *nasofaringe* tem uma função respiratória. Situa-se acima do palato mole e é a extensão posterior das cavidades do nariz, as coanas. O tecto e a parede posterior da nasofaringe formam uma superfície contínua que se situa abaixo do corpo do esfenóide e da parte basilar do occipital. O tecido linfóide abundante na faringe forma um anel incompleto junto da parte superior da faringe – o *anel linfático da faringe*. O tecido linfóide agrega-se em determinadas regiões. As adenoides encontram-se na túnica

mucosa do tecto da parede posterior da nasofaringe. Estendendo-se inferiormente a partir da extremidade medial da tuba de Eustáquio encontra-se uma prega vertical de túnica mucosa, a *prega salpingofaríngea*. Esta recobre o músculo salpingofaríngeo que abre o óstio faríngeo da tuba de Eustáquio durante a deglutição. A colecção de tecido linfóide na túnica submucosa da faringe próxima do óstio faríngeo da tuba de Eustáquio é a *amígdala tubária*. Atrás do toro tubário e da prega salpingofaríngea encontra-se uma projecção lateral da faringe em forma de fenda, o *recesso faríngeo*, que se estende lateral e posteriormente.

A *orofaringe* tem uma função digestiva. É limitada pelo palato mole, superiormente, a base da língua, inferiormente, e os arcos palatoglóssico e palatofaríngeo, lateralmente. Estende-se do palato mole até à margem superior da epiglote. As amígdalas palatinas são agregados de tecido linfóide localizadas lateralmente na orofaringe, sobre o leito amigdalino, entre os arcos palatinos. O leio amigdalino é formado pelo músculo constritor superior da faringe e pela lâmina fibrosa da fáscia faringobasilar. Esta fáscia funde-se com o periósteo da base do crânio e define os limites da parede faríngea na sua parte posterior. Nos adultos, a amígdala não preenche a fissura amigdalina, entre os arcos palatoglóssico e palatofaríngeo.

A *hipofaringe* situa-se atrás da laringe, estendendo-se da margem superior da epiglote e das pregas faringoepiglóticas até à margem inferior da cartilagem cricóide, onde se estreita, tornando-se contínua com o esófago. Posteriormente, a parte laríngea da faringe está relacionada com os corpos das vértebras C4 até C6. As suas paredes posterior e lateral são formadas pelos músculos constritores médio e inferior, e, internamente, a parede é formada pelos músculos palatofaríngeo e estilofaríngeo. A hipofaringe comunica com a laringe através do adito da laringe, na sua parede anterior. O *recesso piriforme* é uma pequena depressão da hipofarínge de ambos os lados do ádito da faringe. Este recesso é revestido por mucosa e está separado do ádito da laringe pela *prega ariepiglótica*. Lateralmente, o recesso piriforme é limitado pelas faces mediais da cartilagem tiróide e pela membrana tiro-hióideia. Os ramos internos do nervo laríngeo superior e os ramos do nervo laríngeo recorrente situam-se profundamente à túnica mucosa do recesso piriforme.

A parede da faringe é composta principalmente por uma camada de músculo circular externa e uma camada de músculo longitudinal interna. Em algumas partes, a disposição é oposta. A camada circular externa de músculos da faringe é composta pelos três músculos constritores. A camada interna de músculos, essencialmente longitudinais, consiste no palatofaríngeo, no estilofaríngeo e no salpingofaríngeo. Estes músculos elevam a laringe e encurtam a faringe durante a deglutição e a fonação. Os constritores da faringe têm um forte revestimento fascial interno, a fáscia faringobasilar, e um revestimento externo fino, a fáscia bucofaríngea. Os constritores da faringe contraem voluntariamente e de forma sequencial, da extremidade superior da faringe para a inferior, impelindo os alimentos para o esófago. Todos os nervos constritores são supridos pelo plexo nervoso faríngeo que é formado pelos ramos faríngeos dos nervos vago e glossofaríngeo, e pelos ramos simpáticos do gânglio cervical superior. O plexo faríngeo situa-se na parede lateral da faringe, principalmente no músculo constritor médio.

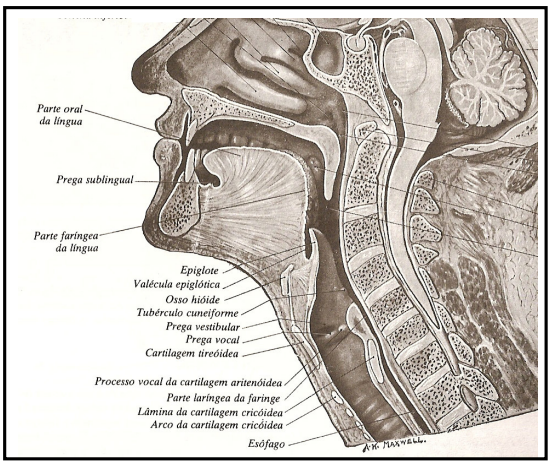

Figura 3.1 – *Tracto vocal – vista em orientação sagital. De [102]*

3.1.2 REGIÃO ORAL

A região oral inclui a cavidade oral, onde se incluem os dentes, os arcos alveolares recobertos por mucosa, a língua, o pavimento da boca, o palato e a região das amígdalas palatinas.

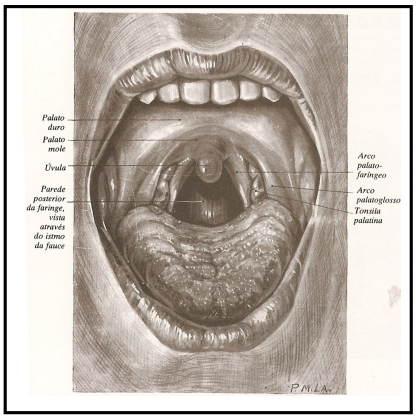

Figura 3.2 - – *Região Oral – vista frontal. De [102]*

A *cavidade oral* consiste em duas partes: o vestíbulo e a cavidade própria oral. O *vestíbulo* é o espaço semelhante a uma fenda entre os arcos alveolares e os lábios e as bochechas. O vestíbulo comunica com o exterior e o tamanho deste orifício é controlado pelos músculos peri-orais. A *cavidade própria oral descreve-se como* o espaço entre os arcos dentais superior e inferior. É limitada lateral e anteriormente pelos arcos alveolares maxilares e mandibulares onde se inserem os dentes. O tecto da cavidade oral é formado pelo palato. Posteriormente, esta cavidade comunica com a orofaringe. Quando a cavidade oral está fechada e em repouso, é completamente ocupada pela língua.

3.1.3 LÁBIOS

Os lábios são as pregas musculares móveis que circundam a cavidade oral e são compostos pelos músculo orbicular dos lábios, vasos e nervos labiais superiores e inferiores. Os lábios são recobertos externamente por pele e internamente por mucosa.

3.1.4 DENTES

Os dentes são estruturas cónicas, duras, fixadas nos alvéolos da mandíbula e maxila. Os adultos possuem normalmente 32 dentes. Podem ser divididos segundo as suas características: os incisivos, que apresentam uma margem cortante fina; os caninos, que são cones proeminentes simples; os pré-molares, com duas cúspides divididas por

um sulco sagital; e os molares, compostos por três ou mais cúspides. A face vestibular de cada dente é direccionada para fora ou superficialmente e a face oral é direccionada para dentro ou profundamente.

3.1.5 PALATO

O palato forma o tecto arqueado da cavidade oral e a base das fossas nasais e da nasofaringe, separando estas estruturas anatómicas. O palato consiste em duas regiões: o palato duro, de estrutura óssea, anteriormente, e o palato mole, composto por diversos músculos, posteriormente.

Os dois terços anteriores do palato possuem um esqueleto ósseo formado pelos processos palatinos dos maxilares e lâminas horizontais dos palatinos.

O terço posterior, móvel, do palato, continua-se na margem posterior do palato duro. O palato mole inclui uma lâmina aponevrótica, membranácea, que se fixa na margem posterior do palato duro e uma lâmina fibromuscular, posterior. O palato mole estende-se postero-inferiormente como margem curva livre a partir da qual pende um processo cónico, a *úvula*. O palato mole é reforçado pela aponevrose palatina formada pelo tendão expandido do músculo tensor do véu palatino. Lateralmente, o palato mole é contínuo com a parede da faringe unindo-se à língua e à faringe por meio dos arcos palatoglóssico e palatofaríngeo respectivamente.

A mucosa do palato duro está firmemente ligada ao osso subjacente. A mucosa que recobre a face lingual dos dentes e o processo alveolar é contínua com a mucosa do palato. Profundamente à mucosa encontram-se as glândulas palatinas que produzem muco. Os orifícios dos ductos destas glândulas conferem à túnica mucosa palatina uma aparência esburacada. Na linha mediana, posteriormente aos dentes incisivos maxilares, encontra-se a papila incisiva. Posteriormente à linha mediana do palato a partir da papila incisiva encontra-se uma faixa esbranquiçada estreita, a rafe palatina. Esta pode apresentar-se como uma crista anteriormente e um sulco posteriormente.

Os cinco músculos do palato mole originam-se na base do crânio e descem até ao palato e são: o levantador do véu palatino, o tensor do véu palatino, o palatoglosso, o palatofaríngeo e o músculo da úvula.

3.1.6 LÍNGUA

A língua é um órgão móvel que pode assumir uma variedade de formas e posições. A língua localiza-se parcialmente na cavidade da boca e na faringe. Em repouso, ocupa essencialmente toda a cavidade própria da boca. A língua possui uma raiz, um ápice, uma face dorsal curva e uma face inferior. A *raiz* da língua é a parte postero-inferior da mesma, sendo relativamente fixa ao osso hióide e à mandíbula, e aos músculos génio-hióideu e milo-hióideu. O *corpo* da língua é a restante parte. O *ápice* da língua é a parte anterior pontiaguda do corpo. O corpo e o ápice da língua são extremamente móveis. O *dorso* da língua é a face póstero-superior da língua que inclui um sulco em forma de V, o sulco terminal, cujo ápice aponta posteriormente para o orifício cego, uma pequena depressão. O sulco terminal divide o dorso da língua em parte anterior, que se situa na cavidade da boca, e na parte posterior, que se situa na orofaringe.

A mucosa da parte anterior da língua é rugosa devido à presença de numerosas pequenas papilas linguais, nomeadamente as papilas circunvaladas, as folhadas, as filiformes e as fungiformes. A túnica mucosa do dorso é fina sobre a parte anterior da língua e está intimamemente fixada no músculo subjacente. O sulco mediano é uma depressão na face dorsal, que se continua profundamente no septo da língua; estas estruturas dividem a língua em duas metades.

A parte posterior da língua é aquela localizada atrás do sulco terminal e dos arcos palatoglossos. Não possui papilas, mas apresenta folículos linguais. A face inferior da língua é recoberta com uma fina túnica mucosa transparente. Com a língua elevada observa-se o freio da língua, uma grande prega mediana da túnica mucosa e a carúncula sublingual, uma papila de cada lado do freio.

Os músculos da língua podem ser divididos em extrínsecos, aqueles que alteram a posição da língua, e intrínsecos, aqueles que alteram a sua forma. Os quatro músculos

extrínsecos e os quatro músculos intrínsecos, em cada metade da língua, são separados pelo septo da língua que se funde posteriormente com a aponevrose da língua. Os músculos extrínsecos são: o genioglosso, o hioglosso, o estiloglosso e o palatoglosso. Os músculos intrínsecos são: o longitudinal superior, o longitudinal inferior, o transverso e o vertical.

3.1.7 FOSSAS NASAIS

As fossas nasais estendem-se desde as narinas, anteriormente, e continuam-se posteriormente, na nasofaringe, através das coanas. São revestidas por mucosa, excepto a zona do vestíbulo, que é revestido por pele, a partir da qual crescem as vibrissas. A mucosa do nariz está firmemente aderente ao periósteo e ao pericôndrio dos ossos e das cartilagens do nariz. A mucosa é contínua com o revestimento de todas as câmaras com as quais comunicam as cavidades nasais, incluindo: a nasofaringe, posteriormente; os seios perinasais, o frontal, os etmoidais, o esfenoidal e o maxilar superior e lateralmente; o saco lacrimal e túnica conjuntiva, superiormente.

Os dois terços inferiores da mucosa do nariz formam a área respiratória e o terço superior a área olfactiva.

Cada cavidade nasal é limitada por: o tecto, que é curvo e estreito excepto na sua extremidade posterior, sendo dividido em três partes - frontonasal, etmoidal e esfenoidal; a base, que é mais larga que o tecto, sendo formada pelo processo palatino do maxilar e pela lâmina horizontal do palatino; a parede medial, formada pelo septo nasal; a parede lateral, que é irregular devido a três elevações, os cornetos, que se curvam infero-medialmente, cada um formando um meato.

Os cornetos nasais – superior, médio e inferior – dividem a cavidade nasal em quatro passagens: o recesso esfenoetmoidal, o meato superior, o meato médio, e o meato inferior.

3.2 FONÉTICA ARTICULATÓRIA

A unidade básica da linguística é o fone: um som articulado. A articulação é o processo de produção de sons por manipulação dos articuladores que alteram a configuração do tracto vocal. Uma colecção de fones é designada de fonema. Os fones são sons e os fonemas elementos do discurso. Os fonemas podem ser combinados em supra-unidades designadas por sílabas. A combinação de sílabas forma palavras – geralmente compostas por dois a cinco fonemas. As frases são unidades linguísticas superiores que requerem regras gramaticais ou regras que determinem a forma como é organizada a ordem das palavras.

Os sons que produzimos são influenciados e alterados pelos sons vizinhos. Uma dessas influências é a adaptação fonética. Tais adaptações são o resultado de variações na forma como o orador move os articuladores, causando alterações nas cavidades do tracto vocal. A forma e extensão como essas cavidades são alteradas dependem dos fonemas passados, presentes e seguintes. A posição dos articuladores e a forma das cavidades para um fonema influenciam os movimentos dos articuladores na produção dos fonemas vizinhos. Um exemplo de adaptação surge quando se fala mais rapidamente. Quanto mais rapidamente se fala menos probabilidade tem a língua de atingir certas posições específicas para certos fonemas específicos.

A adaptação é o processo de variação, ou alteração, de um fonema pela influência da forma do tracto vocal para fonemas vizinhos. A adaptação surge como resultado da alteração do movimento dos articuladores devido ao contexto fonémico. Assim, a produção de um fonema é influenciado pela forma do tracto vocal vizinha. Contudo, a manifestação acústica do fonema permanece inalterada.

A assimilação é uma forma excessiva de adaptação. Quando um fonema se altera o suficiente para se tornar noutro fonema vizinho, diz-se que ocorreu um fenómeno de assimilação.

A co-articulação é definida como o movimento simultâneo de dois articuladores para fonemas distintos. A co-articulação pode ocorrer com ou sem alteração da

produção do som. Este processo pode também levar ao "esbatimento" entre fronteiras fonémicas, tal como na assimilação, levando à alteração das características de dado fonema.

O discurso não é uma linha ordenada e precisa de fonemas. Aproxima-se mais a uma série de fonemas com curvas de *onset* e *ofset* de amplitude variável, contribuindo para a transição entre fonemas. Assim, o discurso pode ser visto como uma sequência de fonemas "esbatidos", usualmente não pronunciados correctamente, tornando a sua identificação a partir de formas de onda ou espectrogramas uma tarefa nem sempre simples de efectuar.

3.2.1 CLASSIFICAÇÃO DE FONEMAS

Uma classificação tradicional dos fonemas divide-os em duas classes: a classe das consoantes e a classe das vogais e semivogais.

As vogais são menos bem definidas que as consoantes pelo facto de a língua não tocar outro órgão durante a sua produção. Podem ser descritas pelas seguintes variáveis: relação da altura da língua com o palato – superior e inferior; posição mais anterior ou posterior da língua; arredondamento dos lábios; sons nasais ou não-nasais, de acordo com o encerramento da continuidade entre a faringe e a nasofaringe, pelo palato. As semivogais têm características articulatórias idênticas às vogais mas apresentam uma duração muito menor e não constituem o núcleo da sílaba. Um ditongo é um fonema monossilábico *glissandi* que tem início numa posição articulatório para uma vogal e move-se na direcção da posição de uma outra vogal, ocorrendo também uma alteração na ressonância da vogal.

As consoantes distinguem-se de acordo com o seu modo de articulação e pelo seu ponto de articulação. O modo de articulação considera a forma como o fluxo de ar é modulado pelo tracto vocal durante a realização da consoante. De acordo com este critério definem-se as seguintes classes:

- Plosivas ou Oclusivas Orais – são consoantes que necessitam de um fecho completo do tracto vocal com acumulação de pressão que é depois subitamente libertada. Se o véu palatino levantar, impedindo o acoplamento da cavidade nasal, trata-se de uma plosiva oral. Em português existem oclusivas orais vozeadas, [b], [d] e [g] e não vozeadas [p], [t] e [k].

- Nasais ou Oclusivas Nasais – se a oclusão se der na cavidade bucal e o véu palatino estiver descido, abrindo a cavidade nasal, obtém-se uma plosiva nasal. Em português apenas existem plosivas nasais vozeadas [m] e [n].

- Fricativas – há uma constrição do tracto vocal que provoca um comportamento turbulento do fluxo de ar. Exemplos de fricativas vozeadas do português são: [v] e [z], e as não vozeadas: [f] e [s].

- Laterais – há obstrução parcial do fluxo de ar provocado pela língua em contacto com o palato ou alvéolos, deixando aberturas laterais para a passagem do ar. Em português as consoantes laterais são vozeadas, por exemplo: [l] e [lh].

- Vibrante – a oscilação da intensidade do fluxo de ar é provocada pela repetida movimentação de um articulador. No português há três consoantes vibrantes: [R] velar com vibração da língua junto do véu, como em "carro"; [□] alveolar onde há apenas uma obstrução da língua com os alvéolos, com em "caro"; e [r] alveolar múltiplo, ou r múltiplo onde existe vibração da ponta da língua junto aos alvéolos.

- Africadas – ocorre uma obstrução completa do tracto vocal seguida de contrição de tipo fricativo. Pode ser vozeada ou não vozeada. Só ocorre em alguns dialectos do português.

As consoantes laterais e vibrantes também podem ser designadas como líquidas devido à sensação de fluidez da sua articulação.

Quanto ao ponto de articulação as consoantes podem ser definidas como:

- Bilabiais – entre os dois lábios; exemplos, [b], [p] e [m].

- Labiodentais – entre o lábio inferior e os incisivos; exmplos, [v] e [f].

- Dentais – entre a ponta da língua e os incisivos; exemplos, [d], [t], [z].

- Alveolares – entre a ponta da língua e a ruga palatiana; exemplos, [l] e [n].

- Pré-palatais – entre a lâmina da língua e o pré-palato; exemplos, [z] e [x].

- Palatais – entre a lâmina da língua e o palato; exemplos, [lh] e [nh].

- Velares – entre a parte detrás da língua e o véu palatiano; exemplos, [g] e [k].

3.2.2 ALFABETOS FONÉTICOS

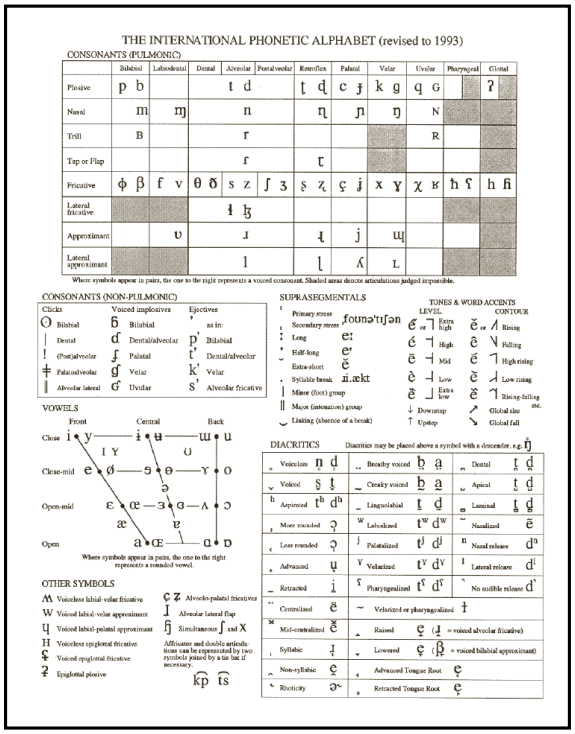

Figura 3.3 - *Alfabeto Fonético Internacional. De Roach, P. [103]*

Os alfabetos fonéticos nasceram da necessidade de se representar graficamente os sons da fala, o fonema.

O Alfabeto Fonético Internacional, IPA, foi desenvolvido por foneticistas britânico e franceses sob a alçada da Associação Fonética Internacional estabelecida em Paris em 1886. O IPA pretende ser uma notação padrão para a representação de todas as linguagens. Já sofreu várias revisões durante a história, tendo sido a mais recente em 1996. A Figura 3.3 mostra a versão mais recente do IPA.

O SAMPA, *Speech Assessment Methods Phonetic Alphabet*, é um sistema de escrita fonético legível por computadores. É um mapeamento do IPA para ASCII. Foi originalmente desenvolvido no final da década de 1980, no âmbito do projecto ESPIRIT da Comunidade Económica Europeia. Este alfabeto só é válido para a língua específica a que foi adaptado e, actualmente, já existem adaptações para mais de 30 línguas, incluindo o português europeu.

3.3 MODELO FÍSICO ACÚSTICO DO TRACTO VOCAL

Childers, D. [79] e Deller, J. *et al.* [2] descrevem com algum pormenor a modelação física do tracto vocal. Segue-se uma abordagem mais sucinta dos principais pontos de interesse a reter para esta dissertação.

O tracto vocal é uma cavidade com perdas, tridimensional, composta por uma secção não uniforme e paredes não rígidas, Sondhi, M. [104]. Apesar de ser possível uma aproximação matemática a tal sistema, nunca é possível obter valores precisos para a forma do tracto vocal ou para as propriedades físicas das paredes de modo a estabelecer um modelo fiável. Tais limitações sugerem a necessidade de uma versão simplificada do modelo acústico do tracto vocal.

Uma primeira simplificação é assumir propagação planar de ondas no tracto vocal, ao contrário da propagação em espaço livre onde esta é radial em três dimensões. Tal é razoável já que, em primeiro lugar, o tecido mole ao longo do tracto vocal impede a propagação radial da onda sonora, e, em segundo lugar, a dimensão média lateral,

seccional do tracto vocal é bastante inferior aos comprimentos de onda das frequências da fala. Tendo em conta que o diâmetro médio do tracto vocal é cerca de 2 cm verifica-se que este é muito inferior ao comprimento de uma onda sonora a 4 kHz:

$$\lambda_{4kHz} = \frac{c}{F} = \frac{340 \quad m/seg}{4000 \quad ciclos/seg} = 8{,}5 \quad cm \tag{3.1}$$

Onde c é a velocidade do som no ar.

Estritamente falando, esta assumpção é válida apenas para frequências inferiores a 4 kHz. Mas para a fala, onde se considera 5 kHz como uma largura de banda apropriada, a propagação planar é uma assumpção adequada.

Assumindo a propagação planar no tracto vocal, apenas a área seccional e o perímetro ao longo do tracto vocal determinam as características acústicas do mesmo. Assim, as equações acústicas podem ser descritas aproximadamente em uma dimensão em vez de três. Tendo isto em consideração é possível afirmar que todas as partículas, para cada área seccional, num dado deslocamento x, terão a mesma velocidade independentemente da sua localização. Fala-se então de uma velocidade volumétrica e não de uma velocidade particular.

Para cada secção do tracto vocal, o modelo acústico foi demonstrado por Portnoff, M. [105] onde as ondas sonoras em tubos sem perdas satisfazem as seguintes equações:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial (u/A)}{\partial t} \tag{3.2}$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2}\frac{\partial (pA)}{\partial t} + \frac{\partial (A)}{\partial t} \tag{3.3}$$

Onde a primeira equação caracteriza a velocidade volumétrica, $u = u(x,t)$, e a segunda a pressão, $p = p(x,t)$, ao longo do tracto vocal desde a glote, $x = 0$, até aos lábios, $x = l$. Quanto aos parâmetros: ρ é a densidade do ar, c é a velocidade do som, e

$A = A(x,t)$ é a função de área do tubo. Aplicando as equações anteriores à secção especificada pela área seccional de A obtém-se:

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A}\frac{\partial (u)}{\partial t} \tag{3.4}$$

$$-\frac{\partial u}{\partial x} = \frac{A}{\rho c^2}\frac{\partial (p)}{\partial t} \tag{3.5}$$

Este modelo matemático permite idealizar o tracto vocal como um tubo de comprimento igual ao do tracto vocal e de área constante. Mesmo dando a esse tubo uma forma curvada "em L", mais próxima do tracto vocal real, parecem manter-se as propriedades acústicas. Shondi, M. [106] mostrou que a curvatura de tal tubo afecta apenas os pontos de ressonância em pequena percentagem quando comparado com o tubo a direito.

A modelação em forma de tubo é importante para se perceber de forma simplista os efeitos de ressonância que ocorrem no tracto vocal. Os objectos possuem frequências nas quais vibram preferencialmente. Se se tentar fazer vibrar o objecto a uma frequência diferente de uma preferencial, os sons resultantes serão mais débeis ou abafados e acabam por desaparecer mais rapidamente. Se se tentar fazer vibrar o objecto numa sua frequência preferencial, as vibrações serão reforçadas, exibindo-se ressonância. Este mesmo fenómeno ocorre com o tracto vocal. Cada uma das frequências de ressonância do tracto vocal é chamada de formante, sendo frequentemente numeradas desde a de mais baixa frequência até à de mais alta frequência. Por vezes refere-se a frequência formante como "um formante", sendo esta terminologia aparentemente derivada da observação de espectrogramas de fala vozeada, nos quais, os formantes serão os elementos mais escuros dispostos horizontalmente num espectrograma de banda larga.

Para melhor compreender o processo de ressonância, pode modelar-se um tubo que vibra com uma extremidade fechada e outra aberta – de modo análogo à abertura nos lábios. É possível obter uma onda estacionária num meio tubo se a área de pressão elevada atingir a extremidade aberta exactamente ao mesmo tempo que a extremidade fechada retorna à pressão normal. Quando tal acontece, as ondas reflectidas que viajam para trás da extremidade aberta irão coincidir exactamente com as ondas que viajam

para a frente da extremidade fechada e serão reforçadas.O o tubo exibirá ressonância. Tal como mostra a Figura 3.4, as frequências de ressonância do tubo serão todas as frequências, F_i, ais que: o comprimento do tubo equivale a ¼ do comprimento de onda de F, ou que o comprimento tubo equivale a ¾ do comprimento de onda de F, ou que o comprimento do tubo equivale a 5/4 do comprimento de onda de F, e assim em diante. Isto significa que a segunda frequência de ressonância será 3 vezes maior que a primeira, a segunda cinco vezes maior e em diante.

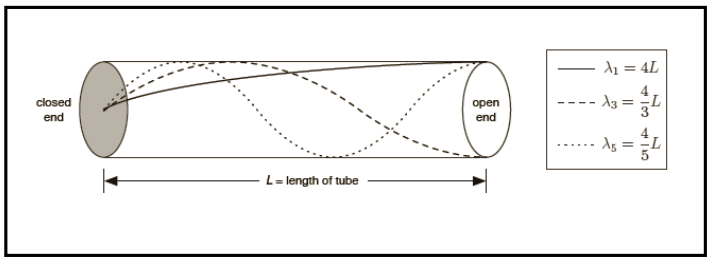

Figura 3.4 – *Tubo acústico sem perdas. De Kim, Y. [107]*

A resposta em frequência de tal tubo de comprimento l, no caso de uma excitação exponencial, é igual a:

$$H = \frac{1}{\cos(2\pi f l / c)} \tag{3.6}$$

com pólos em:

$$F_i = \frac{2i+1}{4l} \tag{3.7}$$

Os pólos representam as frequências naturais do modelo e correspondem às formantes do tracto vocal. Estas formantes ocorrem independentemente da forma do tracto vocal mas a sua localização varia com a forma, alterando o seu valor. Note-se que este modelo só com pólos não representa as anti-ressonâncias, zeros, introduzidos pelo tracto nasal – considera-se aqui que o tracto nasal não contribui para o sinal de fala produzido, o que ocorre por exemplo na produção de vogais orais do português europeu.

Assim, para o tracto vocal é possível determinar uma curva de resposta indicando as frequências de ressonância preferenciais do mesmo. Estas frequências são determinadas pelos parâmetros da configuração do tracto vocal como o comprimento do tracto vocal e a posição dos articuladores. A localização e largura de banda de cada formante dependem portanto desta configuração, alterando o modelo ideal acima descrito.

As regiões formantes também não estão directamente relacionadas com a frequência fundamental de excitação e podem manter-se mais ou menos constantes à medida que esta varia.

Contudo, como a produção de fala é caracterizada por uma forma do tracto vocal variante no tempo, espera-se que um modelo mais realista do tracto vocal consista num tubo cuja área varie em função do tempo e ao longo do eixo de propagação do som. Uma forma de o conseguir consiste em modelar o tracto vocal como uma série de tubos de área variável concatenados. Se se utilizar um grande número de tubos com pequenos comprimentos, espera-se que a estrutura de formantes dos tubos concatenados se aproxime daquela de um tubo com área seccional continuamente variante e também daquela do tracto vocal humano real.

Depois de estabelecido este modelo genérico que permite de facto aproximar o tracto vocal real, convém relembrar que desde o início, se introduziram alguns pressupostos irrealistas no modelo. Em particular, assumiu-se que o tracto vocal poderia ser aproximado por uma série de tubos de paredes rígidas, uniformes e sem perdas. Mas de facto ocorrem perdas de energia durante a fala devido a fricção viscosa entre o fluxo de ar e as paredes, a vibração e a condução térmica ao longo das paredes. As perdas mais significativas ocorrem devido à vibração das paredes. Como estas são estruturas relativamente maciças respondem preferencialmente às baixas frequências, aumentando as larguras de banda dos formantes mais baixos. Por outro lado, as perdas térmicas e por viscosidade são mais influentes nas altas frequências, levando a um aumento das larguras de banda dos formantes mais elevados. As perdas por vibração também levam a um pequeno aumento da frequência central dos formantes, enquanto que as restantes perdas levam a uma pequena diminuição. O resultado líquido é um pequeno aumento

em relação ao modelo do tubo sem perdas. O conhecimento destes factores permite uma implementação que os compense espectralmente.

Matematicamente também é possível adicionar os efeitos de vibração da parede do tracto vocal ao modelo de tubo. As variações de pressão no interior do tracto vocal irão causar variações na área seccional, já que exercem uma força nas paredes elásticas do tracto. Assumindo que as paredes estão sujeitas a reacções locais, isto é, o movimento de uma porção da parede está dependente apenas da pressão acústica dessa porção e independente da movimentação de qualquer outra parte da parede, a área $A(x,t)$ será uma função da pressão $p(x,t)$. Como as variações da pressão são pequenas, as variações na área seccional resultantes podem ser tratadas como pequenas perturbações:

$$A(x,t) = A_0 + \Delta A = A_0 + yS_0 \tag{3.8}$$

Onde A_0 é a área nominal, ΔA é a pequena perturbação, S_0 é a circunferência do tubo, e y é o deslocamento das paredes devido à pressão sonora no interior dos tubos. A vibração das paredes é modelada por um modelo mecânico de massa-elasticidade-viscosidade e governado pela lei de Newton. Representando m, b, e k como a massa, a resistência mecânica e a rigidez da parede por unidade de comprimento do tubo, respectivamente, de acordo com a lei de Newton:

$$m\frac{\partial^2 y}{\partial^2 t} + b\frac{\partial y}{\partial t} + ky = pS_0 \tag{3.9}$$

Definindo a velocidade volumétrica gerada pela vibração da parede como:

$$u_w = \frac{\partial(yS_0 l)}{\partial t} \tag{3.10}$$

As duas equações anteriores podem ser combinadas para se obter:

$$p = \frac{m}{S_0^2 l}\frac{\partial u_m}{\partial t} + \frac{b}{S_0^2 l}u_m + \frac{k}{S_0^2 l}\int u_m dt \tag{3.11}$$

A impedância de vibração da parede pode ser incluída em cada secção elementar do tracto vocal como um elemento distribuído ou inserida como um elemento *shunt* agrupado.

Os efeitos da fricção viscosa e condução térmica locais na parede são muito menos pronunciados que aqueles da vibração das paredes. Childers, D. [79] descreve essas perdas em detalhe e mostra que os efeitos de fricção viscosa podem ser considerados pela introdução de uma resistência dependente da frequência, R, em série com uma bobine, L. Os efeitos da condução térmica através das paredes do tracto vocal podem também ser considerados pela introdução de uma resistência dependente da frequência em paralelo com um condensador, C. A resistência, R, é significativa em simulações no domínio do tempo; quando ocorre uma constrição, a resistência muito elevada e o fluxo de ar é bloqueado. Como resultado, uma secção do tracto vocal pode ser representada por um número finito de elementos de uma linha de transmissão.

3.4 TÉCNICAS DE ANÁLISE DO TRACTO VOCAL

De modo a obter o comprimento, área e forma do tracto vocal, bem como a posição dos articuladores, é possível recorrer às seguintes técnicas indirectas: Raios-X, Tomografia Computorizada, TC, Ressonância Magnética, RM, Ultrasonografia, Articulografia Electromagnética, EMA, e Electropalatografia / Optopalatografia.

3.4.1 IMAGIOLOGIA

3.4.1.1 RAIOS-X

A imagiologia do tracto vocal por Raios-X foi realizada pela primeira vez nos anos 20, mas a sua utilização generalizada para esse fim teve início apenas nos anos 50. A principal limitação da utilização de Raios-X para estes estudos prende-se com a exposição à radiação por parte dos pacientes. Outra limitação está relacionada com o facto de as imagens de Raios-X não possuírem qualquer informação volumétrica das estruturas captadas uma vez que se tratam de imagens obtidas por transparência. A

captação de imagens com informação tridimensional pode ser conseguida através de RM e TC. Ridouane, R. [108].

3.4.1.2 TOMOGRAFIA COMPUTORIZADA

A Tomografia Computorizada, TC, permite a obtenção de imagens de alta resolução do tracto vocal, com informação volumétrica do mesmo. Relativamente às estruturas ósseas, a resolução das imagens obtidas é superior à conseguida pela RM e estruturas como os dentes são claramente demonstradas. O tempo de aquisição também é baixo – mas mais rápido que em RM –, o que reduz o potencial de fadiga de articulação do sujeito e surgimento de artefactos devidos à sua eventual movimentação. A principal desvantagem da Tomografia Computorizada é, tal como o Raio-X, a exposição dos sujeitos à radiação ionizante. Rua, S. [109]

3.4.1.3 RESSONÂNCIA MAGNÉTICA

A RM explora as características de ressonância magnética diferenciais dos tecidos e do ar para obter imagens tridimensionais da forma do tracto vocal.

Desenvolvido com maior detalhe no capítulo 3.5.1.1

3.4.1.4 ULTRASONOGRAFIA

A ultrasonografia, quando aplicada ao estudo do tracto vocal, é utilizada principalmente para a observação da língua, da sua posição e movimentação em tempo real – o que é permitido pela reflexão das ondas ultrassonoras na interface tecido-ar da superfície da língua. Esta técnica possibilita ainda a obtenção da posição do palato duro: o sujeito tem de encher a boca com água e forçá-la para cima, garantindo o contacto com o palato – deste modo, as ondas acústicas não são impedidas pela superfície da língua. As principais vantagens da ultrasonografia são a sua inocuidade para o sujeito, as boas resoluções temporal e espacial, a portabilidade do equipamento e o baixo custo.

As desvantagens incluem a incapacidade de analisar o tracto vocal como um todo, a obtenção, por vezes, de imagens muito ruidosas, e dificuldade em localizar o ápice da língua de forma sistemática – esta dificuldade pode, contudo, ser ultrapassada por alterações ao processo tradicional de obtenção da imagem. Ridouane, R. [108].

3.4.2 OUTRAS TÉCNICAS

3.4.2.1 ARTICULOGRAFIA ELECTROMAGNÉTICA

A técnica de articulografia electromagnética, EMA, baseia-se na lei de indução de Faraday que formula a indução de uma força electromotriz numa bobine posicionada num campo magnético variante no tempo. Num sistema de EMA, o campo magnético é criado pela aplicação de uma corrente eléctrica a um outro tipo de bobina. Tal bobina é denominada de transmissora, enquanto a bobina onde a corrente é induzida é referida como a bobina receptora. A intensidade do campo magnético no local da bobine receptora decresce em função do aumento da distância entre o transmissor e o receptor. A distância pode então ser determinada pela medição da corrente induzida na bobina receptora

O posicionamento de múltiplas bobinas transmissoras em diferentes posições permite determinar a localização relativa das transmissoras à receptora. Cada bobina transmissora gera um campo magnético particular pela aplicação de corrente com diferentes frequências às diversas bobinas. Assim, a partir da corrente na bobina receptora é possível separar a componente oriunda de cada bobina transmissora e medir as distâncias de cada bobina transmissora à bobina receptora simultaneamente. Maeda, S. *et al* [110].

3.4.2.2 ELECTROPALATOGRAFIA / OPTOPALATOGRAFIA

A electropalatografia, EPG, utiliza uma rede de sensores colocados no palato. Cada sensor mede electricamente se a língua entra em contacto com o palato em determinado ponto. A partir da localização desse ponto ou de vários pontos é possível

deduzir a área de constrição provocada pela língua. Este método só é relevante para o estudo de consoantes dado que só nestas articulações a língua entra em contacto com as superfícies superior e laterais do tracto vocal. Apesar de ser um bom método para estudo das constrições no tracto vocal, não dá qualquer informação sobre as restantes áreas não constritas.

A optopalotografia, OPG, apresenta uma configuração semelhante à EPG mas utiliza fibra óptica para condução de luz no sentido do palato e no sentido inverso, de modo a determinar a distância pela quantidade de luz reflectida na superfície da língua. A OPG permite a determinação da posição da língua, mesmo que esta não esteja em contacto com o palato. Wrench, A. *et al* [111].

3.5 MODELAÇÃO DO TRACTO VOCAL

Existem duas abordagens principais para modelar o tracto vocal: modelação articulatória e modelação acústica. A modelação articulatória tem como objectivo representar o tracto vocal e o movimento dos articuladores com o máximo detalhe fisiológico possível, assumindo que um sistema semelhante irá produzir uma saída semelhante. Os modelos articulatórios têm o potencial de conseguir boas representações através de sinais de controlo simples e são capazes de reproduzir todos os efeitos perceptualmente relevantes do sinal de fala real. No entanto, para estes modelos é necessário conhecer as dimensões do tracto vocal e levar a cabo uma análise detalhada do movimento dos articuladores. Tal informação é difícil de obter e por vezes requer técnicas de medição intrusivas.

As técnicas de modelação acústica do tracto vocal modelam a forma de onda directamente, quer no domínio do tempo quer no das frequências. Os modelos são de construção simples, requerendo apenas a forma de onda de fala, facilmente obtida por meio de um microfone. Estes modelos também não requerem uma semelhança exacta da forma de onda ou do espectro para a produção de uma síntese perceptualmente robusta, não necessitando de modelar eventos perceptualmente não importantes. A técnica mais popular é o método de modelação acústica no tempo conhecida por predição linear. Vaseghi, S. [112].

3.5.1 MODELOS ARTICULATÓRIOS DO TRACTO VOCAL

Os modelos articulatórios interpretam e simulam directamente o sistema de produção de fala através de parâmetros como a posição e movimento dos articuladores ou a área do tracto vocal – informação que pode ser obtida pelos métodos de análise descritos atrás. Os métodos são diversos e incluem técnicas que envolvem a solução de equações diferenciais acústicas do tracto vocal, o uso de aproximações a tubos acústicos – introdução a este método no capítulo 3.3 –, filtros de onda digitais e circuitos eléctricos análogos. No que diz respeito à síntese articulatória é, contudo, possível dividi-la em dois grandes grupos: a baseada em modelos físicos precisos derivados de medidas do tracto vocal; ou a baseada no modelo articulatório da produção de fala, incluindo modelos simples dos articuladores.

3.5.1.1 IMAGIOLOGIA DE RESSONÂNCIA MAGNÉTICA DO TRACTO VOCAL

Tal como referido anteriormente, a determinação precisa da forma, do comprimento e do volume do tracto vocal são de extrema importância para a modelação articulatória do tracto vocal. A RM apresenta-se como uma técnica que permite realizar esse objectivo de forma segura, mas acima de tudo com detalhe e qualidade.

As principais vantagens da RM para o estudo da produção da fala são: a aparente inocuidade para o sujeito; a alta qualidade da imagem dos tecidos moles; a capacidade de selecção de sequências de imagem relativamente pequenas a partir de qualquer ângulo; e a produção de imagens passíveis de tratamento tridimensional.

As principais desvantagens prendem-se com o longo tempo de exposição, comparativamente com TC; com a influência da gravidade nas partes moles do tracto vocal devido à postura corporal que é necessária adoptar e com a razão sinal-ruído que depende directamente do tempo de aquisição e deverá ser relativamente elevada para permitir uma boa diferenciação estrutural; e com a ausência da representação de

estruturas com baixa ou nenhuma densidade de hidrogénio, como é o caso dos dentes. Ridouane, R. [108]

O primeiro artigo publicado sobre o estudo tridimensional da forma do tracto vocal a partir de imagens de RM data de 1986, Rokkaku, M *et al* [113]. Rokkaku, M. *et al.* apresentaram, pela primeira vez, a RM como uma alternativa às técnicas de raios-X, para o estudo da forma, com elevado detalhe, do tracto vocal. Rapidamente surgiram mais estudos, incluindo o trabalho pioneiro de Baer, T. *et al* [114], entre outros, Matsumura, M. e Sugira, A. [115], Baer, T. *et al* [116] e Lakshiminarayanan, S. *et al* [117], orientados para a análise do processo de fala, nos quais se aprofundou o conhecimento das características articulatórias para a vocalização sustentada de vogais e outros sons. Se as primeiras abordagens focaram a aquisição de imagens estáticas, rapidamente se percebeu a capacidade da RM para o estudo dinâmico do tracto vocal durante a produção de fala – através da aquisição de uma sequência de sons, sincronizados com o equipamento, durante um certo período de tempo e posterior reconstrução dos movimentos articulatórios. O primeiro estudo a explorar as potencialidades desta técnica data de 1990, Foldvik, A. *et al* [118].

A capacidade da RM para proporcionar uma forma detalhada do tracto vocal permitiu a análise mais precisa da função de área do tracto vocal – variação da área seccional ao longo do tubo acústico aéreo –, das suas propriedades acústicas e uma maior precisão na sua aplicação à síntese articulatória de fala. Matsumura, M. e Sugira A. [115], também em 1990, conceptualizaram esta aplicação. Sulter, A. *et al* [119] e Moore, C. [120] estudaram a relação entre a forma e volume do tracto vocal e as suas propriedades acústicas. Os primeiros trabalhos orientados para a obtenção precisa da área, comparação com características ressonantes, e aplicação em sintetizadores surgiram com Greenwood, A. *et al* [121], Tiede, M. *et al* [122], Story, B. *et al* [123] e Soquet, A. *et al* [124].

Outros estudos, começaram a focar-se em zonas específicas do tracto vocal, como a língua, a cavidade nasal, os dentes, e a sua articulação para grupos específicos de fonemas, nomeadamente vogais, fricativas, e líquidas, como por exemplo Narayanan, S. *et al* [125] e Matsumura, M. *et al* [126]. Tais trabalhos aumentaram o conhecimento

articulatório que estudos posteriores vieram aprofundar, nomeadamente para diferentes línguas.

3.5.1.1.1 MODELOS TRIDIMENSIONAIS DO TRACTO VOCAL

Um modelo tridimensional do tracto vocal, a partir de imagens de RM, pode ser obtido pelo alinhamento de cortes contíguos, obtidos por segmentação da área do tracto vocal – fronteira tecido-ar – e sua reconstrução em 3D.

Quase que simultaneamente com o despertar para as potencialidades da RM para análise tridimensional do tracto vocal, surgiu também a necessidade de segmentação da área do tracto vocal. Baer, T. *et al* [116] utilizaram técnicas de segmentação por limiar, *thresholding*, semi-automáticas, tendo sido posteriormente propostas variações, da técnica por Matsumura, M. *et al* [126] e Moore, C. [127]; Demolin, D. *et al* [128] utilizaram um método manual de traçado; Soquet, A. *et al* [129] descrevem ainda um método de crescimento de curvas elásticas; Behrends, J. e Wismüller [130] propõem um método de crescimento de regiões tridimensional.

É possível encontrar na literatura científica vários métodos de reconstrução 3D do tracto vocal a partir de imagens de RM, destacando-se os trabalhos de: Badin, P. *et al* [131] cujo modelo parte do apresentado em Badin, P. *et al* [132] e pretende refinar a posição da língua, dos dentes e da forma da face, recorrendo para tal à análise de componentes principais para identificação das coordenadas dos diferentes órgãos – uma versão melhorada do método proposto foi apresentada 6 anos depois em Badin, P. e Serrurier, A. [120]; Kröger, P. *et al* [133] utilizaram um sistema de rede de mapeamento da linha médio-sagital e posterior reconstrução com recurso a vários algoritmos automáticos de segmentação e referenciação tridimensional; Xiao, B. *et al* [134] com um modelo de elementos finitos, cuja movimentação dos articuladores é permitida para síntese articulatória em tempo real; Engwall, O. [135] recorre a uma rede poligonal tridimensional dividida nas seguintes zonas: paredes do tracto vocal e nasal, língua, lábios e dentes.

A Figura 3.5 apresenta exemplos de alguns dos modelos tridimensionais do tracto vocal acima descritos.

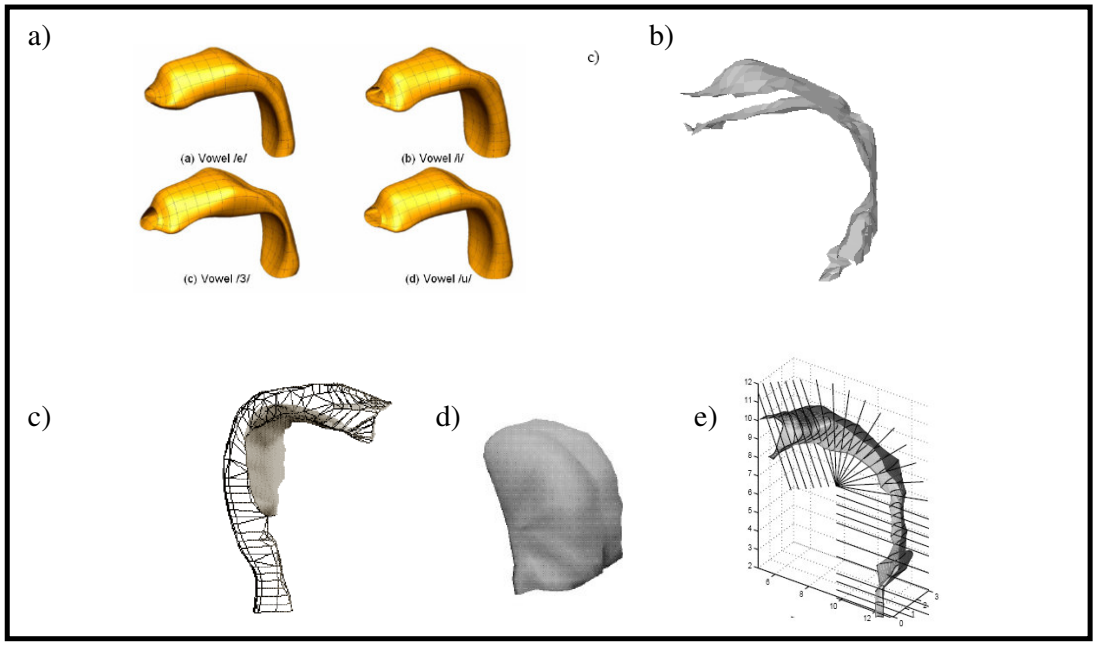

Figura 3.5 - *Exemplos de modelos tridimensionais do tracto vocal. a) modelos de Xiao, B. et al [133] para as vogais /e/ /i/ /y/ /u/; b) modelo de Kröger, P. et al [132] para a vogal /a/; c) e d) modelos de Engwall, O. [135] com detalhe da língua; e) modelo de Badin, P. et al [130] para a vogal /a/.*

3.5.1.1.2 MODELOS SÓLIDOS DO TRACTO VOCAL

A criação de modelos sólidos do tracto vocal, a partir dos dados recolhidos através das várias técnicas do mesmo, teve início no trabalho pioneiro de Chiba, T. e Kajima, M. [136], do ano de 1941. Os autores recolheram dados fisiológicos e medições tridimensionais do tracto vocal – área – através da utilização das técnicas mais avançadas da época, incluindo imagiologia por raios-X. A partir destes dados calcularam os espectros e frequências ressonantes para todas as vogais, concluindo que a natureza acústica das mesmas era determinada pela forma do tracto vocal. Com esta informação construíram modelos sólidos para cada vogal, descrito no capítulo *Artificial Vowels* do seu livro, e compararam os resultados sintéticos com aqueles das vogais naturais. Em 2001, como homenagem mas também com objectivos pedagógicos, Arai, T. [137] recriou os modelos originais de Chiba, T e Kajima, M. – tal como se podem ver na Figura 3.6 – mostrando serem extremamente efectivos na demonstração das características acústicas das vogais bem como da teoria fonte-filtro – descrita no

capítulo 3.5.2.1. Arai, T. estendeu, ainda, os modelos físicos a algumas consoantes baseado em literatura moderna de fonética acústica sobre as mesmas.

Figura 3.6 - *Modelos de Chiba, T. e Kajima, M. reproduzidos por Arai, T. [137].*

Mais recentemente, Fujita *et al* [138] desenvolveram modelos sólidos por via de prototipagem rápida – estereolitografia – para todas as vogais japonesas a partir de imagens de RM, com o intuito de identificar características anatómicas que contribuem para a identidade vocal de cada indivíduo.

Noutro trabalho semelhante, Pavel, R. *et al* [139] desenvolveram um modelo sólido, também por via de prototipagem rápida – Modelação de Deposição Fundida – *Fused Deposition Modelling* – da vogal checa /a/ a partir de imagens de RM. Este estudo pretendia validar a construção destes modelos sólidos para o estudo das propriedades acústicas do tracto vocal.[3]

Apesar da construção de modelos sólidos do tracto vocal ser uma área de interesse para a acústica fonética e para o processamento da fala, bem como para a medicina, são escassas as menções, na literatura científica internacional, a experiências deste tipo.

[3] O autor da presente dissertação só teve conhecimento destes dois trabalhos numa data próxima da conclusão da sua escrita, pelo que se ainda surgiram atempadamente para que pudessem ser mencionados, não o fizeram para que os resultados neles apresentados fossem tidos em conta na realização das actividades experimentais.

3.5.2 MODELOS ACÚSTICOS DO TRACTO VOCAL

3.5.2.1 MODELO FONTE-FILTRO

O modelo fonte filtro foi originalmente proposto por Fant, G. [140] e é a concepção teórica do processo de produção de fala mais utilizada em análise, síntese e processamento da fala. O modelo produz um sinal de fala a partir de um sinal excitatório gerado na fonte – laringe – que passa por um filtro variante no tempo – o filtro do tracto vocal, que pode ou não incluir o filtro do tracto nasal – e um filtro de radiação. A fonte vocal é gerada por uma vibração periódica das pregas vocais, e o filtro do trato vocal é caracterizado pelas ressonâncias das suas cavidades. O filtro de radiação é geralmente tido como uma característica constante que aproxima os efeitos da radiação que ocorre nos lábios.

O modelo assume que o filtro e a fonte são linearmente separáveis, e que a resposta em frequência do sinal de fala, à saída, é dado pelo produto das respostas em frequência da fonte e do filtro –tal como se observa na Figura 3.7.

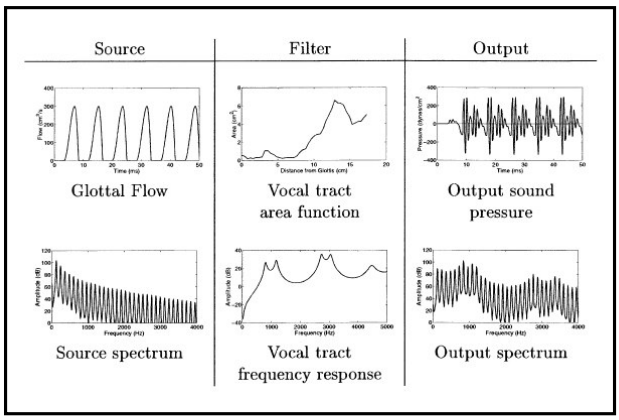

Figura 3.7 – *Ilustração da representação fonte-filtro para vogais, do domínio do tempo – superior – e no domínio das frequências – inferior. De Story. B. [141].*

3.5.2.2 PREDIÇÃO LINEAR – LPC

Vaseghi, S. [112] providencia uma introdução e análise bastante intuitiva aos métodos de predição linear, tal como aplicados ao processamento da fala pela primeira vez por Atal, B. e Hanauer, S. [142]. De uma forma genérica, os modelos de predição

linear, LP, predizem os valores de um sinal a partir da combinação linear dos seus valores passados. Considerando que os sinais de fala são parcialmente aleatórios e parcialmente predizíveis, estes podem ser modelados como a saída de um filtro linear excitado por um sinal aleatório não correlacionado – modelo fonte filtro descrito na secção acima. A entrada aleatória constitui a parte imprevisível do sinal, enquanto que o filtro modela a estrutura previsível. O objectivo da predição linear é modelar os mecanismos que introduzem correlação num sinal.

No modelo de predição linear, a fala é modelada como a saída de um filtro só com pólos, excitado por um trem de impulsos separados por um período fundamental para sons vozeados e ruído aleatório para sons não vozeados.

Tais pressupostos implicam que em cada janela de análise, e para cada instante n, a equação de saída do modelo é dada por:

$$s(n) = \sum_{k=1}^{P} a_k s(n-k) + G u_n \tag{3.12}$$

Onde P é o número de pólos, u_n é uma excitação de entrada apropriada, G é o ganho do filtro e a_k são os coeficientes que caracterizam o filtro – coeficientes de predição linear.

A formulação da LPC tem dois campos separados: análise e síntese. A análise de fala por predição linear baseia-se no pressuposto que a forma de onda pode ser representada por um número fixo de pólos. Este pressuposto é bastante preciso para vogais e sons próximos de vogais. Já para outros sons o modelo é apenas uma aproximação. Expansões do modelo LP introduziram também zeros para além de pólos no sistema.

O erro de predição $e(n)$, definido como a diferença entre o real valor da amostra $s(n)$ e o valor predito $\hat{s}(n)$, é dado por:

$$e(n) = s(n) - \hat{s}(n)$$

$$= s(n) - \sum_{k=1}^{P} a_k s(n-k) \tag{3.13}$$

Para sinais de fala, ou outros transportadores de informação, o erro de predição $e(n)$ pode ser considerado como a informação da amostra $s(n)$. A partir da equação anterior, um sinal gerado, ou modelado, por um preditor linear pode ser descrito pela seguinte equação:

$$s(n) = \sum_{k=1}^{P} a_k s(n-k) + e(n) \tag{3.14}$$

A transformada em z de uma equação LP mostra que o modelo LP é um filtro digital só com pólos com a seguinte função de transferência:

$$H(z) = \frac{X(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{P} a_k s(n-k)} = G \frac{1}{\prod_{k=1}^{N} (1 - r_k z^{-1})} \frac{1}{\prod_{k=1}^{M} (1 - 2r_k \cos\varphi_k z^{-1} + r_k^2 z^{-2})} \tag{3.15}$$

Nesta equação assume-se que existem M pares de pólos complexos e N pólos reais com $P = N + 2M$; e r_k e φ_k como sendo o raio e o ângulo do pólo k. A resposta em frequência do modelo LP é dada por:

$$H(f) = \frac{G}{1 - \sum_{k=1}^{P} a_k e^{-j2\pi f}} = G \frac{1}{\prod_{K=1}^{N} (1 - r_k e^{-j2\pi f})} \frac{1}{\prod_{k=1}^{M} (1 - 2r_k \cos\varphi_k e^{-j2\pi f} + r_k^2 e^{-j4\pi f})} \tag{3.16}$$

As principais características da ressonância espectral de um pólo são: frequência, largura de banda e magnitude da ressonância. As raízes de um par complexo de pólos podem ser escrita em termos de raio r_k e o ângulo φ_k como:

$$z_k = r_k e^{\pm j\varphi_k} \tag{3.17}$$

E a frequência de ressonância de um par de pólos é:

$$F(\varphi_k) = \frac{F_s}{2\pi}\varphi_k \tag{3.18}$$

Onde F_s é a frequência de amostragem. A largura de banda de um pólo relaciona-se com o raio r_k como:

$$B_k = (-\log r_k)(F_s / \pi) \tag{3.19}$$

A magnitude da ressonância espectral é dada por $H(\varphi)$.

Na aplicação da predição linear à fala, os pólos do modelo de predição linear modelam as ressonâncias nas formantes da fala.

3.5.2.2.1 CÁLCULO DOS COEFICIENTES DO PREDITOR

Os coeficientes de predição linear são obtidos pela minimização da média quadrática do erro de predição, tal como:

$$E\left[e^2(n)\right] = E\left[\left[s(n) - \sum_{i=1}^{P} a_i s(n-i)\right]^2\right] \tag{3.20}$$

O intervalo do sinal utilizado para cálculo dos coeficientes que minimizam o erro é o que distingue os dois métodos disponíveis para tal. No método de autocorrelação assume-se que o sinal é janelado de forma a que fora da janela todo o sinal seja zero. O método de autocorrelação garante um filtro LP estável, isto é, para o qual todos os pólos se encontram no interior do círculo unitário. O método de covariância não assume que o sinal é janelado e que para N amostras disponíveis, o erro é janelado de forma a que $N - p$ amostras estejam disponíveis. O método de covariância pode gerar filtros instáveis.

Notas adicionais sobre a estimação dos coeficientes pelos dois métodos podem ser encontradas em vários livros, incluindo Childers, D. [79].

3.5.2.2.2 O FILTRO INVERSO

O modelo de predição linear só com pólos modela o espectro de um sinal de entrada transformando um sinal de excitação não-correlacionado, $u(n)$, num sinal de saída correlacionado, $s(n)$. No domínio das frequências, a relação entrada-saída do filtro só com pólos é dado por:

$$S(f) = \frac{GU(f)}{A(f)} = \frac{E(f)}{1 - \sum_{k=1}^{P} a_k e^{-j2\pi fk}} \tag{3.21}$$

Onde $S(f)$, $E(f)$ e $U(f)$ são os espectros de $s(n)$, $e(n)$ e $u(n)$ respectivamente; G é o factor de ganho da entrada; e $A(f)$ é a resposta em frequência do preditor inverso. Como o sinal de excitação, $e(n)$, é assumido como tendo um espectro plano, a forma do espectro do sinal, $S(f)$, deve-se à resposta em frequência, $1/A(f)$, do modelo de predição só com pólos. O preditor linear inverso, tal como o nome implica, transforma o sinal correlacionado, $s(n)$, no sinal não correlacionado de espectro plano, $e(n)$. O filtro inverso, também conhecido como filtro do erro de predição, é um filtro de resposta impulsional finita, FIR, só com zeros, definido como:

$$\begin{aligned} e(n) &= s(b) - \hat{\mathrm{s}}(\mathrm{n}) \\ &= s(n) - \sum_{i=1}^{P} a_i s(n-k) \\ &= (a^{inv})^T s \end{aligned} \tag{3.22}$$

Onde o filtro inverso é $(a^{inv})^T = [1, -a_1, ..., -a_P]$ e $s^T = [s(n), ..., s(n-P)]$. A função de transferência, no domínio z, do modelo de predição inverso é dado por:

$$A(z) = 1 - \sum_{i=1}^{P} a_i z^{-i} \tag{3.23}$$

O modelo de predição linear é um filtro só com pólos, onde os pólos modelam as ressonâncias do espectro do sinal. O inverso do filtro só com pólos é um filtro só com zeros, com os zeros situados nas mesmas posições que os pólos do filtro só com pólos. Consequentemente, os zeros do filtro inverso introduzem anti-ressonâncias que cancelam as ressonâncias dos pólos do preditor. O filtro inverso tem o efeito de tornar plano o espectro do sinal de entrada, sendo tal operação também conhecida como *spectral whitening*.

3.5.2.2.3 O SINAL DE ERRO DE PREDIÇÃO

Em geral, o sinal do erro de predição é composto por três componentes: o sinal de entrada, também chamado de sinal de excitação; os erros devido a imperfeições da modelação; o ruído.

O erro de predição no sentido dos mínimos quadrados só se torna zero se as três condições seguintes forem satisfeitas: o sinal é determinístico; o sinal é correctamente modelado por um preditor de ordem P; e o sinal não apresenta ruído.

O erro quadrático mínimo de predição pode ser obtido por:

$$E^{(P)} = E\left[e^2(m)\right] = r_{ss}(0) - \sum_{K=1}^{P} a_k r_{ss}(k) \tag{3.24}$$

Onde $E^{(P)}$ denota o erro de predição para um preditor de ordem P. O erro de predição decresce com o aumento da ordem do preditor, inicialmente muito rapidamente e depois mais lentamente, até à ordem correcta do modelo.

3.5.2.2.4 SELECÇÃO DA ORDEM DO MODELO DE PREDIÇÃO

Um procedimento para determinar a ordem correcta do modelo é aumentar a ordem do modelo e monitorizar a variação diferencial na potência do erro até que a

variação estabilize. A variação incremental na potência do erro com o aumento da ordem do modelo de $i-1$ até i é dada por:

$$\Delta E^{(i)} = E^{(i-1)} - E^{(i)} \tag{3.25}$$

A ordem P para a qual a diminuição da potência do erro $\Delta E^{(P)}$ se torna inferior a um dado limiar é tomada como a ordem do modelo.

Em predição linear, são necessários dois coeficientes para modelar cada um dos picos espectrais do espectro do sinal. Assim, a modelação de um sinal com K ressonâncias dominantes no espectro necessita de $P = 2K$ coeficientes. Deste modo, uma forma de seleccionar a ordem do modelo é examinar a magnitude do espectro do sinal e determinar a ordem do modelo como duas vezes o número de picos espectrais significativos encontrados no espectro.

Quando a ordem é inferior à ordem correcta do modelo, o sinal fica sub-modelado. Neste caso, o erro de predição não fica correctamente descorrelacionado e será superior ao mínimo óptimo. Uma outra consequência da sub-modelação é o decréscimo da resolução espectral do modelo: picos espectrais adjacentes do sinal podem fundir-se e surgir como um único pico. Quando a ordem do modelo é superior à ordem correcta, o sinal é sobre-modelado. Tal pode resultar numa equação matricial sobre-determinada, surgindo soluções numéricas pouco fiáveis e picos espectrais não genuínos.

3.5.3 MODELO DE RADIAÇÃO NOS LÁBIOS

A energia acústica escapa do tracto vocal através dos lábios. Considerando as analogias com as linhas de transmissão, os lábios são tratados como impedância de radiação que carrega o tracto vocal. A impedância de radiação contém uma parte resistiva que representa a perda de energia acústica, e uma parte de reactância que representa a inércia da massa de ar nos lábios.

A conversão da velocidade volumétrica na boca para pressão de radiação nos lábios é frequentemente modelada como uma simples operação de diferenciação, podendo ser representada pela seguinte função de transferência:

$$L(z) = 1 - z^{-1} \tag{3.26}$$

Tal facto indica que, quando a fonte para a filtragem inversa é o sinal de pressão acústica, o efeito de radiação nos lábios tem de ser cancelado pelo seu filtro inverso, de modo a ser obtida a onda glotal. Em termos de processamento de sinal, o filtro inverso de $L(z)$ reduz-se a uma simples operação de integração. No entanto, como tal operação implicaria uma ganho infinito na componente contínua e grande amplificação das frequências mais baixas, a implementação prática é dada por:

$$inv(L(z)) = \frac{1}{1 - \mu z^{-1}} \tag{3.27}$$

Onde μ é um parâmetro variável cujo valor se encontra ligeiramente abaixo da unidade.

4 FILTRAGEM INVERSA

A filtragem inversa, tal como geralmente aplicada à fala, é essencialmente uma tentativa de reverter o processo de produção de fala, pela aplicação de um filtro ao sinal de fala com uma função de transferência inversa à função de transferência do tracto vocal. Assim, o efeito de filtragem do tracto vocal é cancelado, e o resultado do filtro inverso é uma estimativa do sinal fonte da fala. Tipicamente, a filtragem inversa aplica-se a formas de onda de pressão sonora da fala, mas também pode ser aplicada a gravações do sinal de fluxo de ar oral.

Estimativas robustas e precisas da função de transferência do tracto vocal são essenciais para uma filtragem inversa eficiente. A forma do impulso glotal é muito sensível a erros de estimação das frequências e larguras de banda da primeira formante, especialmente quando F1 é muito baixa. Pequenos erros nas formantes superiores pouco efeito têm na forma do impulso glotal principal e correspondente espectro frequencial.

Existem vários métodos desenvolvidos para a estimação das frequências formantes e respectivas larguras de banda, podendo operar de forma completamente automática, métodos automáticos, ou, alternativamente, permitir ao utilizador a afinação fina do filtro inverso de uma forma interactiva, pulso a pulso se necessário, métodos interactivos.

A grande maioria dos métodos de filtragem inversa assume o modelo fonte-filtro e baseia-se na análise de predição linear, LPC, do sinal de fala, atrás descrito. Existem inúmeras abordagens, pelo que se apresentam aqui, de forma breve, os trabalhos mais influentes e a partir dos quais foram derivados outros métodos.

A primeira técnica de desconvolução do sinal de fala, não-síncrona com o perído fundamental, baseada no modelo de predição linear foi proposto por Atal, B. e Hanauer, S. [142]

Wong *et al* [143] propuseram um método síncrono com o período, de modo a considerar a natureza variante no tempo do tracto vocal, estimando a função de transferência deste ao longo da fase de oclusão para cada período do ciclo glotal, recorrendo a um critério do erro de predição.

Alku, P. [144], baseado no seu próprio trabalho, Alku, P. e Laine, U. [145], propôs uma abordagem iterativa onde o sinal glotal é modelado por um formante glotal – um pólo de primeira ou segunda ordem. Este método pretende retirar a contribuição do sinal glotal do sinal de fala antes de se proceder à análise de predição linear para estimar as formantes do tracto vocal. Nesse mesmo trabalho, o autor notou também que a análise de predição linear é um factor limitador na estimação dos parâmetros do tracto vocal, especialmente quando se analisam sinais de fala cuja frequência fundamental é alta e a primeira formante é baixa.

Foram desenvolvidas variações ao método de predição linear convencional para a produção de melhores estimativas do filtro do tracto vocal – e consequentemente uma mais eficaz desconvolução – muitas deles baseadas em pressupostos perceptuais. Destaca-se o trabalho de Backstom, T. *et al* [146] que aplicou o método DAP, *Discrete All-Pole Modelling*, [147]; e de Strube, H. [148] que aplica o método de predição linear com eixo de frequências deformado, *Warped Linear Prediction*.

4.1 DIFICULDADES DA FILTRAGEM INVERSA

A filtragem inversa baseada em LPC, apesar de útil e engenhosa, assenta em vários pressupostos que podem não ser realistas, introduzindo por vezes erros e dificultando o processo de estimação precisa do tracto vocal, e por conseguinte da fonte. Consideram-se aqui apenas as dificuldades associadas à filtragem inversa de sons sustentados, sendo feita uma pequena menção a algumas dificuldades associadas com a filtragem de fala contínua.

O facto da excitação para sons vozeados ser um trem de impulsos cria um pressuposto irrealista já que a excitação de entrada é produzida pela vibração das pregas vocais, e nunca por um trem de impulsos. Assim, nos momentos em que o sinal de

entrada está presente, os coeficientes LPC obtidos não permitem uma representação precisa do tracto vocal. De modo a ultrapassar este problema, os parâmetros LPC são muitas vezes calculados apenas na fase de oclusão do ciclo glotal, onde não existe excitação, e o sinal de fala é apenas resultante das ressonâncias livres do tracto vocal. De notar contudo que, por vezes, a fase de oclusão é muito curta ou inexistente, podendo levar a erros grosseiros. Para ajudar na identificação da fase de oclusão e alinhamento de períodos, bem como para uma melhor determinação da ocorrência de fases de oclusão incompletas ou inexistentes, Krishnamurthy, A. e Childers, D. [149] propuseram um método de filtragem inversa assistida por meio de registos de EGG simultâneos – filtragem inversa de canal duplo.

Do mesmo modo, ao assumir-se que o sinal de excitação é um conjunto de impulsos para sons vozeados e ruído branco para sons não-vozeados, assume-se também que o sinal de excitação tem um espectro plano, o que está longe da realidade. De modo a ultrapassar esta dificuldade, foram propostos vários métodos onde a equação de LPC é alterada de forma a incluir coeficientes que correspondam à fonte glotal. Minimizando o erro quadrático médio obtém-se um conjunto diferente de equações que incluem a estimação conjunta dos parâmetros da fonte vocal bem como aqueles do tracto vocal – ver no próximo capítulo uma descrição de vários métodos que recorrem a tal abordagem.

O critério de nivelamento do declive espectral da fonte, cancelando os picos das formantes o mais efectivamente possível, pode sofrer erros devido à interacção da fonte com o filtro, não considerada no modelo clássico fonte-filtro. A modelação do fenómeno de interacção fonte-filtro mostrou que muitos picos e vales espectrais podem ter origem nos efeitos de interacção e não na função de transferência supraglótica, Bickley, C. e Stevens, K. [150]. Por exemplo, pode existir uma variação da largura de banda dos formantes, principalmente do primeiro formante, devido à variação da impedância glotal durante a fase aberta do ciclo glotal; ou a possibilidade de excitação adicional na abertura glotal, causando oscilações que interagem com as oscilações dos formantes na excitação principal. Adicionalmente, podem ainda existir oscilações provenientes do sistema subglótico.

Relativamente à fala contínua, os métodos baseados em LPC apresentam dificuldade em lidar com as transições encontradas na fala natural, onde a configuração do tracto vocal se altera rapidamente ao longo de uma elocução bem como com transições de sons vozeados para não-vozeados. Sistemas baseados no filtro de Kalman oferecem uma solução robusta e satisfatória para tal questão – ver capítulo 5.5.1 para descrição do filtro de Kalman e sua aplicação à obtenção do sinal glotal.

Os problemas descritos tornam a filtragem inversa uma tarefa desafiadora. Os pontos individuais podem ter pouco efeito – muitos dos quais podem nem ter efeitos perceptuais – mas considerados conjuntamente podem resultar potencialmente em grandes erros. No entanto, o aumento do conhecimento sobre as características dos tracto e fonte vocal, bem como a interacção entre ambos, irá facilitar melhores estimativas do filtro inverso, que por sua vez irão levar a melhores descrições da fonte vocal.

4.2 ESTIMAÇÃO CONJUNTA DA FONTE E DO TRACTO VOCAL / OPTIMIZAÇÃO PARAMÉTRICA DE MODELOS DA FONTE GLOTAL

Como o ajuste do filtro inverso se baseia em apreciações qualitativas quanto à forma da onda glotal, é necessário estabelecer critérios objectivos e quantitativos, para proceder à modelação da onda glotal. Assim, assumem especial importância os modelos matemáticos capazes de descrever as principais características da forma de onda glotal – ver descrição de modelos no capítulo 2.5. Uma vez obtidos os modelos da fonte, é possível incluir os mesmos em processos de estimação conjunta da fonte e do tracto vocal que levem a estimativas óptimas de ambos.

A ideia de estimar conjuntamente os parâmetros da fonte e do tracto vocal surgiu com Fujisaki, H. e Ljungqvist M. [78, 151] que introduziram um método iterativo baseado num sistema de produção de fala autorregressivo, AR, que vieram a denominar de GAR, *Glottal AR*. Os autores adoptaram uma abordagem análise-por-síntese, na qual os parâmetros do tracto vocal eram identificados por uma simples análise de predição

linear e o modelo glotal era identificado por um método de descida do gradiente. Os resultados mostraram que a introdução de um modelo glotal reduzia o erro de predição em 4 dB comparativamente com o LPC convencional só com pólos. No mesmo trabalho, compararam ainda vários modelos da fonte glotal no processo iterativo e concluíram que o modelo que produz melhores resultados é aquele que providencia uma modelação detalhada do período de encerramento glotal, enquanto que as descontinuidades da abertura glotal parecem ser de menor importância.

Fujisaki, H. e Ljungqvist M. [152], propuseram, um ano mais tarde, uma extensão do método anterior permitindo a modelação com pólos e zeros da função de transferência do tracto vocal, denonimando o método de GARMA – *glottal ARMA*.

Ding, W. e Kasuya, H. [153] e colaboradores, Ding, W. *et al* [154], Kasuya, H. *et al* [155] e Ding, W. *et al* [156] apresentaram um método de análise adaptativo onde utilizam um modelo ARX – ver capítulo 5.2 para descrição de um modelo ARX – com pólos e zeros, do processo de produção de fala, e a entrada exógena é gerada pelo modelo RK. Os parâmetros da fonte e do tracto vocal são identificados conjuntamente de modo a minimizar o erro quadrático médio do modelo ARX. Os coeficientes, do tracto vocal, variantes no tempo, são estimados através de um algoritmo adaptativo baseado no filtro de Kalman, enquanto que os parâmetros do modelo RK são estimados num procedimento de optimização com recurso a um método de solidificação simulada – ver capítulo 5.6.1 para descrição da técnica de solidificação simulada. Contudo cada artigo apresenta pequenas variações relativamente a que parâmetros da fonte estimar de facto e a forma como são estimados. Adicionalmente, em Kasuya, H. *et al* [155] são estimadas as características do ruído incluído no sinal de fala através da determinação das componentes aperiódicas, por intermédio de uma função de autocorrelação alterada, determinadas em várias bandas frequenciais do sinal. Em Ding, W. *et al* [156] a ordem do sistema é determinada de forma semi-automática pela análise de vários candidatos e o seu efeito na largura de banda das formantes inferiores a 3 KHz. No mesmo artigo também é proposto um critério de erro de modo a optimizar automaticamente o instante de abertura glotal.

Fu, Q. e Murphy, P. [157] propuseram também um método baseado na modelação do processo de fala pelo modelo ARX, cujo processo de identificação da fonte e dos coeficientes ARX é em todo semelhante ao proposto por Ding, W. e Kasuya, H. [151]. No entanto, o modelo escolhido para representar a fonte foi o modelo LF.

Os mesmos autores, já em 2006, apresentaram uma versão revista do procedimento anterior, Fu, Q. e Murphy, P. [158], naquele que será provavelmente o método mais completo e complexo até agora proposto, e que pretende ultrapassar, de forma objectiva, as dificuldades de métodos anteriores, nomeadamente no que diz respeito à obtenção de uma solução global óptima. Dada essa dificuldade, os autores separaram o processo de optimização em duas fases: uma primeira fase de inicialização, na qual se obtém uma solução exacta e robusta para um subproblema aproximado, que providenciará os parâmetros iniciais à segunda fase; e uma segunda fase onde ocorre a optimização final dos parâmetros através da minimização do erro quadrático médio, na qual está embebido um filtro de Kalman para estimação adaptativa dos parâmetros do tracto vocal. Na fase de inicialização, o tracto vocal é assumido como um sistema linear invariante no tempo só com pólos cuja entrada é a forma de onda gerada pelo modelo RK. Os parâmetros do tracto vocal são estimados através de uma simples análise LPC de covariância – estes resultados servirão como inicialização da segunda fase – e os parâmetros da fonte são estimados através de um processo de minimização de mínimos quadrados entre o sinal obtido por filtragem inversa a partir dos coeficientes LPC e o sinal gerado pelo modelo RK. Depois de convertidos os parâmetros do modelo RK estimados em parâmetros LF, através de funções definidas pelos autores, procede-se ao processo de estimação conjunta. Neste, os coeficientes do tracto vocal são identificados por um filtro de Kalman e os parâmetros da fonte, agora modelo LF, são optimizados por um método de descida de gradiente com recurso à matriz de Hessian e matriz Jacobiana, com vista à minimização do erro quadrático médio do modelo ARX.

Funaki, K. *et al*. [159] apresentaram o modelo GARMAX, *Glottal-ARMAX*, com compensação de fase. O processo de produção de fala é descrito como um modelo ARMAX cuja entrada é um modelo RK e ruído branco gaussiano para sons não vozeados. Os parâmetros ARMAX são determinados por um Sistema de Identificação de Modelo, MIS – *Model Identification System*, e os parâmetros da fonte por um método híbrido que combina solidificação simulada e algoritmos genéticos.

Lu, H. e Smith III, J. [160] propuseram um método de optimização convexa para estimar conjuntamente o filtro do tracto vocal e a forma de onda da fonte vocal, com o objectivo de melhorar a eficiência computacional e optimização global. O modelo glotal utilizado é o RK e o tracto vocal é modelado como um filtro IIR só com pólos. O processo de optimização proposto pretende minimizar o erro entre a forma de onda glotal estimada por filtragem inversa e a forma de onda real que se assume como a gerada pelo modelo. A resolução é conseguida através de programação linear utilizando a norma L2 via um método de minimização sequencial sem restrições, SUMT. Ao modelo do tracto vocal é adicionado mais um pólo, real, para determinação do pendente espectral do modelo RK.

Fröhlich, M. *et al.* [161, 162] apresentaram um método para filtragem inversa automática derivado do algoritmo DAP, *Discrete All-pole Modeling*, e optimização simultânea de um modelo LF ajustado à forma de onda resultante da filtragem inversa. O processo de utilização emprega um método de descida Simplex, DSM, a seguir ao qual é aplicado um método de Powell, PM.

Kim, Y. [107] também optimiza os parâmetros da fonte de um modelo RK de modo a ajustar este ao sinal obtido por filtragem inversa, mas através de um algoritmo de método de predição linear com eixo de frequências deformado*, Warped Linear Predicton*. A optimização dos parâmetros RK é realizada com recurso a programação quadrática.

Shiga, Y. e King, S. [163] propuseram um método de separação fonte-filtro, sendo esta realizada através de uma aproximação iterativa conjuntamente com um método de análise Multi-Frame. De modo a resolver o problema convencional onde a estimação do envelope espectral está seriamente afectada pela estrutura harmónica do som vozeado, os autores aplicam a vários segmentos de fala a mesma forma do tracto vocal. Considerando que cada um desses segmentos possui geralmente diferente F0 é possível obter um número suficiente de harmónicos em várias frequências para formar o envelope espectral que corresponde à função de transferência real do tracto vocal. O envelope é estimado pela aproximação de todo o espectro dos harmónicos de todas os segmentos utilizando um critério de mínimos quadrados baseados no cepstro. Como se

utilizam *frames* múltiplos na análise do envelope espectral, os autores denominam a abordagem de Análise Multi-Frame, MFA – Multi-frame Analysis.

Jinachitra, P. e Smith III, J. [164] propuseram um método que tem também em conta o ruído de aspiração e o ruído de observação. É utilizado o modelo glotal RK em conjunto com um filtro só com pólos como modelo do processo de produção de fala. É aplicado um algoritmo de Maximização da Expectância, EM, para estimar de forma iterativa os parâmetros do modelo no sentido de máxima verosimilhança, utilizando um filtro de Kalman com função suavizadora no passo de avaliação.

4.3 OUTROS MÉTODOS PARA OBTENÇÃO DO SINAL GLOTAL

4.3.1 MÁSCARA DE FLUXO DE AR

O fluxo glótico também pode ser estimado a partir de registos do fluxo de ar oral, através de uma máscara de fluxo de ar especial com um transdutor de pressão diferencial integrado. Uma vantagem desta técnica é a obtenção de medidas absolutas do fluxo de ar glótico, não possíveis quando utilizada a forma de onda de pressão do sinal de fala. Como o sinal proveniente da máscara é uma estimativa do fluxo de ar oral, a saída do filtro inverso é uma estimativa do verdadeiro fluxo glótico. O diferencial do fluxo glótico pode ser obtido pela adição de um diferenciador – um zero real na frequência zero – ao filtro inverso. Uma das maiores desvantagens deste método é a limitação da resposta em frequência da máscara, tornando-o ineficaz para a análise espectral da fonte vocal. Mesmo com máscaras de pneumatografia circunferencialmente ventiladas, especialmente desenvolvidas para uma resposta frequencial linear estendida, foi mostrado que possuem uma frequência limite superior de cerca de 1.6 KHz. Gobl, C. [165].

4.3.2 TRANSDUTORES DE PRESSÃO MINIATURIZADOS

Estimativas do fluxo de ar glótico também podem ser obtidas a partir de medições de pressão na traqueia e faringe. Medidas precisas da pressão subglótica são tradicionalmente difíceis de obter, podendo ser útil, para tal fim, utilizar a pressão esofágica como uma aproximação. A pressão esofágica pode ser medida com recurso a um pequeno balão insuflável colocado no esófago, directamente anterior e inferior às pregas vocais. Uma técnica que providencia resultados mais fiáveis é a punção traqueal directa, não sendo contudo utilizada extensivamente, talvez devido ao potencial risco médico e desconforto para o sujeito. Medidas subglotais directas também foram obtidas através de um cateter oco introduzido através da glote na traqueia. Nesta técnica, as variações de pressão são conduzidas por tubos estreitos até ao transdutor externo, com consequente perda de resolução frequencial. Contudo, com os métodos descritos obtém-se uma resolução espectral muito baixa.

Esta dificuldade particular pode ser ultrapassada com recurso a transdutores de pressão em forma de células de carga semicondutoras miniaturizados, que possuem uma resposta em frequência mais alargada até os 20 KHz. Estes transdutores podem ser fixados a um cateter fino e introduzidos na traqueia através da glote de tal modo que o cateter se posicione na comissura posterior das pregas vocais.

Esta técnica apresenta a vantagem de obter medidas directas das pressões subglótica, supraglótica e trans-glótica. Uma das maiores limitações da técnica é, contudo, o facto de, mesmo que o transdutor apresente uma excelente resposta em frequência, a fórmula para aproximação do fluxo de ar só é válida para frequências abaixo de 1.5 KHz. Outra dificuldade relaciona-se com a praticabilidade da calibração e registo. Contudo, a maior dificuldade surge do facto dos transdutores de pressão serem extremamente sensíveis a variações de temperatura, devido ao material semicondutor que os constitui. A colocação dos cateteres, e anestesia do sujeito, podem afectar a fonação normal do mesmo. Gobl, C. [165].

4.3.3 TUBO DE SONDHI

O método descrito por Sondhi, M. [166] recorre a um longo tubo de metal para dentro do qual o sujeito procede à fonação. Este tubo actua como uma terminação pseudo-infinita do tracto vocal. Um microfone dentro do tubo regista o sinal, que é uma aproximação do fluxo glótico já que a terminação sem reflexão do tubo provoca uma considerável redução das ressonâncias do tracto vocal.

A principal vantagem deste método é a ausência de necessidade de uma filtragem inversa, não sendo necessários quaisquer pressupostos a priori sobre a fonte vocal. Mais, permite também operar numa gama alargada de frequências. A resposta em fase, no entanto, não é linear para as frequências mais baixas, o que levará a distorção da forma do impulso glotal se não for incluído um filtro de compensação de fase. Uma desvantagem principal desta técnica é o facto de, em princípio, só se aplicar à análise de vogais sustentadas. Também, a fonação para o interior de um longo tubo é uma tarefa não natural, sendo o *feedback* auditivo obrigatoriamente distorcido. Ambos factores podem interferir com o comportamento normal do sujeito. Gobl, C. [165].

5 IDENTIFICAÇÃO E PARAMETRIZAÇÃO DO SINAL GLOTAL

Com o objectivo de obter uma boa estimativa do sinal glotal, numa forma parametrizada, apresenta-se a implementação de um método semi-automático de estimação do sinal glotal assente numa optimização conjunta da fonte e do filtro vocais, baseado nos artigos de Lu, H. e Smith III, J. [160], Fu, Q. e Murphy, P. [158] e Ding, W. *et al* [167].

No método implementado, o modelo RK, que modela a derivada da velocidade volumétrica glotal, é integrado num modelo ARX, *AutoRegressive with eXogenous input*, só com pólos, variante no tempo, do processo de produção de fala. Os dois modelos, RK e ARX, são estimados num processo de estimação conjunta no qual um filtro de Kalman identifica os parâmetros do tracto vocal e um algoritmo de solidificação simulada, SA – S*imulated Annealing*, identifica os parâmetros da fonte glotal.

5.1 MATERIAL

Os sinais acústicos e electroglotográficos, utilizados experimentalmente, foram gravados simultaneamente num computador Pentium 4 a 2.66 GHz e registados no programa livre de gravação e de edição áudio Audacity, versão 1.2.6. A placa de som utilizada foi a GINA-24 da Echo Audio. Cada sinal foi gravado num canal separado, com um frequência de amostragem de 44.1 KHz e 16 bits. Posteriormente ao seu registo, ambos foram decimados para 10 KHz.

A gravação acústica foi realizada utilizando um microfone supercardióide dinâmico Beta 58A da marca Shure, ligado a um pré-amplificador Pro Mike – Servo Drive Mic Preamp, da marca SPL, a uma distância de 20 cm da boca, fora do fluxo de ar. O electroglotógrafo utilizado foi o EG-PC3 da Tiger DRS, Inc. As gravações ocorreram numa sala normal mas isenta de ruído de fundo.

O processamento de sinal foi realizado em Matlab [168] e Praat [169].

5.2 MODELO ARX DO PROCESSO DE PRODUÇÃO DE FALA

O processo de produção de fala pode ser modelado apropriadamente como um sistema de resposta impulsional infinita, IIR, com uma equação de erro dada por:

$$s(n)+\sum_{k=1}^{p}a_k(n)s(n-k)=\sum_{j=1}^{q}b_l(n)u(n-j)+u(n)+e(n) \tag{5.1}$$

Onde $s(n)$ é o sinal de fala observado; $u(n)$ é a fonte glotal desconhecida; a_k e b_j são os coeficientes do filtro IIR variantes no tempo; p e q são as ordens do filtro; e $e(n)$ é a equação de erro. Neste caso, a função de transferência do tracto vocal é definida por:

$$H(z)=\frac{B(z)}{A(z)}=\frac{1+b_1z^{-1}+...+b_jz^{-j}}{1+a_1z^{-1}+...+a_kz^{-k}} \tag{5.2}$$

Contudo, o modelo ARX pode ser definido apenas como um modelo só com pólos, onde $b=0$, dado pelo sistema IIR variante no tempo:

$$s(n)-\sum_{k=1}^{p}a_k(n)s(n-k)=u(n)+e(n) \tag{5.3}$$

Onde $s(n)$ é o sinal de fala observado; $u(n)$ é a fonte glotal desconhecida; a_k são os coeficientes do filtro IIR variantes no tempo; p é a ordem do filtro; e $e(n)$ é a equação de erro. A função de transferência do tracto vocal correspondente é igual a:

$$H(z)=\frac{1}{A(z)}=\frac{1}{1+a_1z^{-1}+...+a_kz^{-k}} \tag{5.4}$$

Como descrito no capítulo relativo à LPC, esta assumpção é válida quando for considerado o modelo fonte-filtro para a produção de vogais, no qual se pressupõe que não existe qualquer contribuição do tracto nasal, e onde também se assume que a interacção entre a fonte e o filtro é negligenciável.

5.3 MODELO RK

Como a entrada do sistema $u(n)$ é, à partida, desconhecida, é necessário encontrar um modelo que a aproxime, de modo a permitir a sua identificação paramétrica, mas também a sua integração no modelo ARX, permitindo a convergência.

Vários modelos foram apresentados no capítulo 2.5.2 desta dissertação. O modelo RK foi o escolhido para implementação no algoritmo dada à sua facilidade de implementação – polinómio de segundo grau na fase de abertura e zero na fase de oclusão – reduzido número de parâmetros a estimar – AV, OQ e TL – e capacidade para aproximar o sinal fonte de uma vocalização normal.

5.4 DESCRIÇÃO DO ALGORITMO

Resumidamente, o algoritmo implementado pode ser descrito da forma que se segue – as questões particulares da sua implementação são discutidas em capítulos subsequentes. Regista-se simultaneamente o sinal de fala e de EGG para a vocalização de uma vogal do português europeu – nos os procedimentos experimentais, a vogal registada foi o /a/. O sinal EGG, a partir da sua derivada, permite a identificação do instante de abertura glotal e a duração do período fundamental. Com o período fundamental determinado, gera-se uma onda da derivada do impulso glotal através do modelo RK, que sofre posteriormente uma operação de pré-ênfase. O sinal de fala registado no início, bem como a onda gerada pelo modelo RK, são integrados num filtro de Kalman para identificação dos parâmetros do tracto vocal. A partir destes coeficientes estimados e da onda RK é reconstruída uma estimativa do sinal de fala. O erro de estimação, isto é, a diferença entre o sinal de fala real registado e a estimativa reconstruída, é utilizado num procedimento de optimização não-linear multidimensional

baseado no método de solidificação simulada para actualização dos coeficientes do modelo RK. Do procedimento de optimização da fonte surgem dois caminhos possíveis: um que repete o procedimento dentro de cada período, até a estabilização do processo de optimização; e outro que avança o procedimento para o período seguinte, após a estabilização do processo de optimização. Como resultado final, obtém-se uma estimativa conjunta dos parâmetros do tracto e da fonte vocais, para cada período, ao longo de uma elocução de duração seleccionada. A Figura 5.1 mostra o fluxograma do algoritmo implementado.

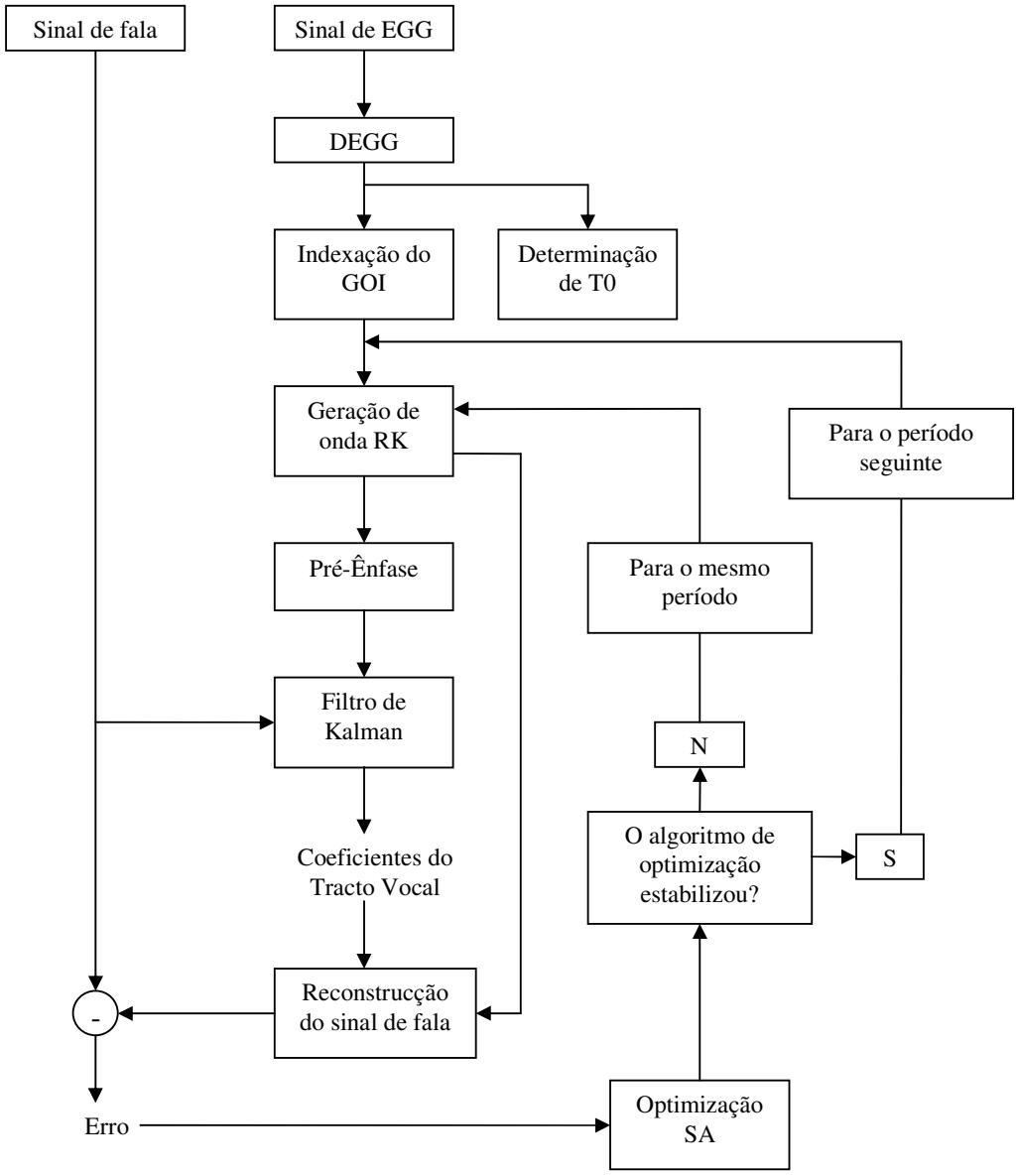

Figura 5.1 - *Fluxograma do algoritmo de identificação conjunta dos parâmetros do tracto e fonte vocais.*

5.5 IDENTIFICAÇÃO DO MODELO ARX

5.5.1 FILTRO DE KALMAN

O filtro de Kalman, Kalman, R.[170], é um método linear de mínimos quadrados recursivo que permite a estimação de um vector de estado, que contém os parâmetros de interesse, utilizando para tal toda a informação disponível até aquele instante. O filtro de Kalman pode ser utilizado em processos variantes ou invariantes no tempo, sendo especialmente robusto em sistemas dinâmicos estocásticos de dimensão linear finita.

A teoria do filtro de Kalman baseia-se na abordagem de espaços de estado onde uma equação de estado modela a dinâmica do processo de geração do sinal e uma equação de observação modela a observação do sinal ruidoso e distorcido. Para um sinal ou estado $X(n)$ e a observação em ruído $y(n)$, as equações que descrevem o modelo de estado e o modelo de observação são definidas como:

$$X(n) = A.X(n-1) + q(n) \tag{5.5}$$

$$y(n) = C.X(n) + r(n) \tag{5.6}$$

Onde $X(n)$ é um vector do sinal ou parâmetros de estado no instante n; A é a matriz de transição de estado, que relaciona o estado presente $X(n)$ com o estado anterior $X(n-1)$; $q(n)$ é o ruído do sistema, definido como um processo Gaussiano com $p[q(n)] \sim N(0,Q)$, onde Q é a matriz de covariância de $q(n)$; $y(n)$ é o vector de observação ruidosa e distorcida; C é a matriz de observação, que relaciona o estado presente $s(n)$ com a observação presente $y(n)$; e $r(n)$ é o ruído de medição, definido como um processo Gaussiano com $p[r(n)] \sim N(0,R)$ onde R é a matriz de covariância de $r(n)$.

O algoritmo do filtro de Kalman geral pode dividir-se em dois passos: o passo de predição e o passo de actualização. No passo de predição, imediatamente antes de uma medição, ocorre a predição do estado presente com base no estado passado, e obtém-se a matriz de covariância do erro de predição. No passo de actualização, após a medição

da observação presente, o estado é estimado, actualizado, com base na inovação, isto é, na diferença entre a predição e a observação. Neste passo também é calculado o ganho de Kalman e a matriz de covariância do erro de estimação.

Em mais detalhe, para a o vector de observação $y(n)$, o estado $X(n)$, e as condições iniciais: predição inicial, $\hat{s}(0|-1)=0$, e matriz de covariância do erro de predição inicial, $P(0|-1)=I$; para $n=0,1,...,N-1$:

PASSO DE PREDIÇÃO

Predição de estado:

$$X(n|n-1)=A.X(n-1) \tag{5.7}$$

Covariância do erro de predição:

$$P(n|n-1)=A.P(n-1).A^T+Q \tag{5.8}$$

PASSO DE ACTUALIZAÇÃO

Ganho de Kalman:

$$K(n)=P(n|n-1).C^T\left[C.P(n|n-1).C^T+R\right]^{-1} \tag{5.9}$$

Actualização de estado:

$$X(n)=X(n|n-1)+K(n)\left[y(n)-C.X(n|n-1)\right] \tag{5.10}$$

Matriz de covariância do erro de estimação:

$$P(n)=\left[I-K.C\right]P(n|n-1) \tag{5.11}$$

5.5.2 APLICAÇÃO AO PROBLEMA

Para o problema em questão, o filtro de Kalman é uma solução ideal para identificar os coeficientes do modelo ARX, dada a sua capacidade de aproximar a

solução óptima em sistemas lineares não-estacionários na presença de ruído de observação.

Considerando que o vector de estado $X(n)$ no instante n é igual aos coeficientes do modelo ARX a estimar nesse dado instante, tal como:

$$X(n) = \{a_1(n), a_2(n), ..., a_p(n)\}$$

E assumindo que a transição é lenta, o que se verifica na maioria dos casos para a vocalização de vogais, a equação de estado é dada por:

$$X(n) = X(n-1)$$

Considerando o modelo ARX:

$$s(n) - \sum_{k=1}^{p} a_k(n) s(n-k) = u(n) + e(n) \tag{5.12}$$

Mostra-se que a observação se relaciona com o estado – os coeficientes – as amostras passadas do sinal e o erro de equação da seguinte forma:

$$s(n) - u(n) = \sum_{k=1}^{p} a_k(n) s(n-k) + e(n) \tag{5.13}$$

Dando origem à seguinte equação de observação:

$$y(n) = C^T(n).X(n) + r(n) \tag{5.14}$$

Onde,

$$y(n) = s(n) - dug^{RK}(n) \tag{5.15}$$

Onde $X(n)$ representa o estado que contém os coeficientes do modelo ARX, $X(n) = \{a_1(n), a_2(n),..., a_p(n)\}$; $s(n)$ o sinal de fala observado; $dug^{RK}(n)$ o sinal de entrada gerado pelo modelo RK; $C(n)$ a matriz de observação definida por $C(n) = \{-s(n-1),...,-s(n-p)\}^T$; e p a ordem do sistema.

5.5.3 ALGORITMO KALMAN

Vector de Estado $X(n) = \{\hat{a}_1(n), \hat{a}_2(n),..., \hat{a}_p(n)\}$

Vector de Observação: $y(n) = \{y(0), y(1),..., y(T_0 - 1)\}^T$ onde $y(n) = s(n) - dug^{RK}(n)$

Matriz de Observação: $C^T(n) = \{-s(n-1),...,-s(n-p)\}^T$

INICIALIZAÇÃO

Vector de Estado Inicial $X(0) = \{\hat{a}_0, \hat{a}_1,..., \hat{a}_p\}$

Matriz de Covariância do erro de predição Inicial: $P(0) = diag(X(0)^2)$

Matriz de Transição de Estado: $A = 1$

Matriz de Correlação do Ruído do Sistema: $Q = 0$

Matriz de Correlação do Ruído de Observação: $R = 0.001$

ALGORITMO

Em cada período de $n = 1,2,...,T_0 - 1$:

(1) $K(n) = P(n-1).C(n)/((C(n).P(n-1).C^T(n)) + R)$

(2) $X(n) = X(n-1) + K(n).(y(n) - C(n).X(n-1)$

(3) $P(n) = P(n-1) - K(n).C(n).P(n-1)$

(4) No final de cada período, verifica-se se existem pólos instáveis, que, caso se verifiquem, são reflectidos para o interior do círculo unitário.

(5) Avança para nova iteração do processo de optimização com novos valores RK de entrada.

5.6 IDENTIFICAÇÃO DA FONTE GLOTAL

5.6.1 SOLIDIFICAÇÃO SIMULADA (SIMULATED ANNEALING)

Van Laarhoven, P. e Aarts, E. [171] providenciam uma visão detalhada sobre a teoria e aplicações do método de solidificação simulada.

O método de solidificação simulada é uma técnica de optimização geral para a resolução de problemas de optimização combinatória, incorporando características de algoritmos de melhoria iterativa, *iterative improvement*. A aplicação de um algoritmo de melhoria iterativa pressupõe: a definição de estados, uma função de custo, e um mecanismo de geração de perturbações que permita gerar a transição de um estado para outro. O mecanismo de geração define uma vizinhança para cada estado, consistindo de todos os estados que podem ser atingidos a partir do estado existente no início de cada iteração. Devido a tal, este tipo de algoritmos também são designados de algoritmos de procura de vizinho ou procura local.

O algoritmo geral de melhoria iterativa pode ser descrito da forma seguinte. Tem início com um dado estado a partir da qual se gera uma sequência de iterações, e onde cada iteração consiste numa possível transição do estado presente para um vizinho seleccionado desse estado. Se o vizinho desse estado apresentar um custo inferior, o estado actual é substituído pelo o estado vizinho. Se o custo for superior ou igual é seleccionado e comparado outro estado vizinho. O algoritmo termina assim que se obtenha um estado cujo custo não seja superior aquele dos seus vizinhos.

As principais desvantagens dos algoritmos de melhoria iterativa são: possibilidade de terminação num mínimo local, não providenciando informação sobre o desvio entre esse mínimo local e o mínimo global; o mínimo obtido depende do estado inicial para o qual não se espera que exista, à partida, conhecimento dos seus valores; em geral não é possível definir um limite superior para o tempo de computação.

Uma forma de resolver os problemas anteriores pode passar pela aceitação de transições que correspondem a um aumento da função de custo de forma limitada. Um algoritmo que segue esta abordagem é o algoritmo desenvolvido por Kirkpatrick, D. *et al.* [172] e Černy, V. [173] denominado de solidificação simulada. De uma forma geral, as soluções obtidas por este algoritmo não dependem das configurações iniciais e apresentam um custo próximo do custo mínimo. Por outro lado, apresenta um tempo computacional superior aos outros métodos de melhoria iterativa.

O algoritmo de solidificação simulada obteve o seu nome devido à analogia entre a simulação da solidificação de líquidos e a resolução de problemas de optimização combinatórios de grandes dimensões. Na física dos materiais, solidificação denomina um processo no qual um sólido num banho térmico é aquecido pelo aumento da temperatura do banho até um máximo no qual todas as partículas do sólido se arranjam aleatoriamente na fase líquida, seguido por um arrefecimento pelo abaixamento lento da temperatura do banho. Deste modo, todas as partículas arranjam-se de forma a obter o mais baixo estado de energia, assumindo que a temperatura máxima é suficientemente alta e o arrefecimento é suficientemente lento.

Começando no valor máximo de temperatura, a fase de arrefecimento do processo de solidificação pode ser descrito da seguinte maneira. A cada valor de temperatura T, é permitido ao sólido atingir o equilíbrio térmico, caracterizado pela probabilidade de se encontrar num estado com energia E dada a distribuição de Boltzmann:

$$P(E = E) = \frac{1}{Z(T)}.\exp\left(-\frac{E}{k_b T}\right) \tag{5.16}$$

Onde $Z(T)$ é um factor de normalização dependente da temperatura T, e k_b é a constante de Boltzmann. O factor $\exp\left(-\frac{E}{k_b T}\right)$ é conhecido como o factor de Boltzmann.

À medida que a temperatura diminui, a distribuição de Boltzmann concentra-se nos estados com menor energia e finalmente, quando a temperatura se aproxima de zero, apenas os estados de energia mínima apresentam uma probabilidade de ocorrência

diferente de zero. Contudo, se o arrefecimento for muito rápido, não é permitido ao sólido atingir o equilíbrio térmico para cada valor de temperatura, e os defeitos podem fixar-se no sólido produzindo uma estrutura amorfa muito diferente da estrutura cristalina de baixa energia. De modo a simular a evolução do equilíbrio térmico de um sólido para um valor fixo de temperatura T, Metropolis N. *et al* propuseram o método de Monte Carlo, no qual são geradas sequências de estados de sólidos da seguinte forma. Dado o estado actual, caracterizado pela posição das partículas, é aplicada uma pequena perturbação aleatoriamente gerada através de um pequeno deslocamento de uma partícula aleatoriamente escolhida. Se a diferença na energia, ΔE, entre o estado actual e o estado ligeiramente perturbado for negativa, ou seja, se a perturbação resultar num estado com energia mais baixa, então o processo é continuado com o novo estado. Se ΔE for maior que zero, então a probabilidade de aceitar o novo estado perturbado é dado pelo factor de Boltzmann. Esta regra de aceitação de estados perturbados é conhecida como o critério de Metropolis. Seguindo este critério, o sistema eventualmente evolui para o equilíbrio térmico, isto é, após um número grande de perturbações, utilizando o critério atrás definido, a probabilidade de distribuição dos estados aproxima-se da distribuição de Boltzmann.

O algoritmo de Metropolis também pode ser utilizado para gerar uma sequência de estados em problemas de optimização combinatória. Neste caso, os estados assumem o papel dos estados do sólido, enquanto que a função de custo C e o parâmetro de controlo c assumem os papéis da energia e da temperatura, respectivamente. Desta forma, é possível ver o método de solidificação simulada como uma sequência de algoritmos de Metropolis avaliados em cada sequência de valores decrescentes da temperatura de controlo. Pode ser descrito da seguinte forma. Inicialmente, ao parâmetro de controlo é atribuído um valor elevado e é gerada uma sequência de estados do problema de optimização combinatória, da seguinte forma. Primeiro define-se um mecanismo de geração de modo a que dado um estado i seja possível obter um outro estado j pela escolha aleatória de um elemento na vizinhança de i. Este passo corresponde à perturbação no algoritmo de Metropolis. Se $\Delta C_{ij} = C(j) - C(i)$, então a probabilidade do estado j ser o próximo estado na sequência é dada por 1 se $\Delta C_{ij} \leq 0$, ou $\exp\left(-\frac{\Delta C_{ij}}{c}\right)$ se $\Delta C_{ij} > 0$, o critério de Metropolis. Assim, existe uma probabilidade

diferente de zero de continuar com um estado com um custo superior ao do estado actual. Este processo é continuado até ser atingido o equilíbrio, isto é, até que a distribuição de probabilidade dos estados se aproxime da distribuição de Boltzmann, agora dada por:

$$P(estado = i) = \frac{1}{Q(c)}.\exp\left(-\frac{C(i)}{c}\right) \tag{5.17}$$

Onde $Q(c)$ é uma constante de normalização dependente do parâmetro de controlo c. O parâmetro de controlo é depois baixado sequencialmente, permitindo ao sistema aproximar-se do equilíbrio, em cada abaixamento, pela geração de uma sequência de estados pelo método já descrito. O algoritmo termina depois para um valor baixo de c, para o qual já não se aceitam quaisquer deteriorações. O estado final fixado é tido como a solução para o problema. Note-se que o critério de aceitação é implementado pela geração de número aleatórios de uma distribuição normal entre [0,1] e posterior comparação com $\exp\left(-\frac{\Delta C_{ij}}{c}\right)$.

O método de solidificação simulada converge assimptoticamente para um estado mínimo global do problema, permitindo, contudo, durante o processo, deteriorações diferentes de zero na função de custo – ou seja, admite a existência de uma probabilidade diferente de zero de continuar com um estado com um custo superior ao de dado estado actual. Note-se, no entanto, que esta probabilidade decresce gradualmente, e que o processo é continuado até ser atingido o equilíbrio que deverá ser igual ao mínimo global do problema.

5.6.2 ALGORITMO DE SOLIDIFICAÇÃO SIMULADA

INICIALIZAÇÃO

Sinal de fala $s(n) = \{s(1), s(2),..., s(T_0 - 1)\}$

Parâmetros da fonte:

$AV = 0.005$

$OQ = 0.65$

Parâmetros de solidificação simulada:

Temperatura Inicial: $Temp = 4.0$

Temperatura Final: $Temp\_Min = 4.0$

Iteração Inicial: $I = 1$

Número de Iterações: $I\_Max = 50$

Função de custo: $E = 1000000$

ALGORITMO

(1) Gerar um deslocamento aleatório para um vizinho de AV e OQ:

$AV' = AV + \Delta AV$

$OQ' = OQ + \Delta OQ$

(2) Gerar a onda RK com AV' e OQ'

(3) Implementar o Filtro de Kalman para identificação dos coeficientes ARX e reconstrução do sinal de fala. Definir o erro como E' e calcular a diferença entre o erro e a função de custo: $\Delta E = E' - E$

4. Se $\Delta E \leq 0$, aceita-se AV' e OQ'. Ir para 5.

Se $\Delta E > 0$, gerar número aleatório ξ de uma distribuição normal

entre [0,1], e:

Se $\exp\left(-\frac{\Delta E}{Temp}\right) > \xi$, aceitar AV' e OQ'.

Se $\exp\left(-\frac{\Delta E}{Temp}\right) < \xi$, rejeitar AV' e OQ'.

(5) Fazer $E = E'$, $I = I + 1$;

Se $I < I\_Max$, ir para 2.

Se $I \geq I\_Max$, fazer $Temp = 0.89 * Temp$.

(6) Se $Temp > Temp\_Min$, fazer $I = 1$, ir para 2.

Se $Temp \leq Temp\_Min$, gravar resultados.

(7) Após gravação de resultados, terminar processo e avançar para o período seguinte.

A Figura 5.2 mostra o fluxograma do algoritmo de solidificação simulada implementado.

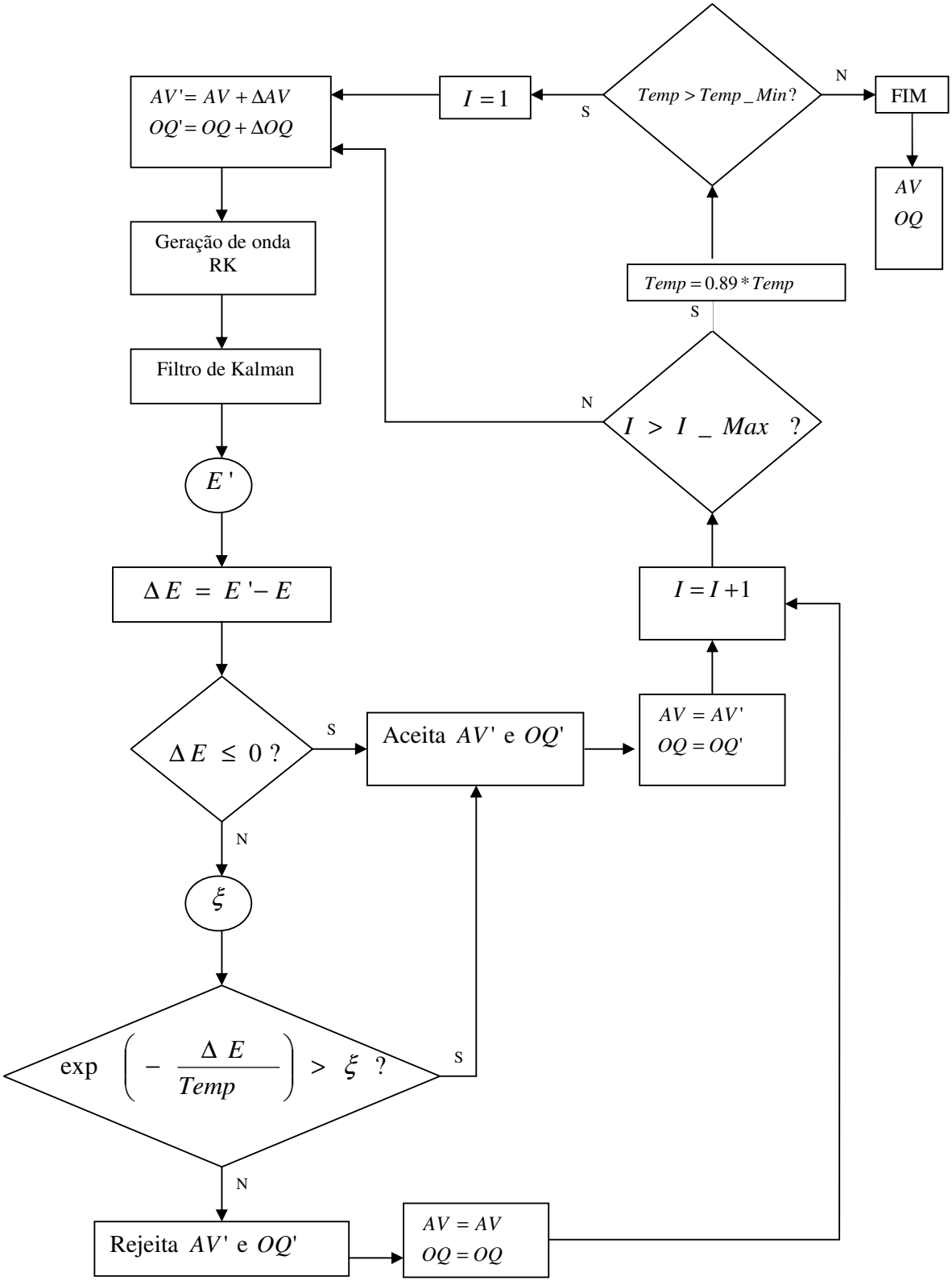

Figura 5.2 - *Fluxograma de algoritmo de solidificação simulada*

5.7 QUESTÕES DE IMPLEMENTAÇÃO

5.7.1 IDENTIFICAÇÃO DO INSTANTE DE ABERTURA E DA DURAÇÃO DO PERÍODO FUNDAMENTAL

O filtro de Kalman e o procedimento de optimização por SA são altamente sensíveis ao correcto alinhamento do período fundamental dos sinais da fala e da fonte glotal, gerada pelo modelo RK. A estimação do parâmetro OQ por SA é especialmente sensível. Assim, é fundamental identificar de forma precisa os instantes de abertura e encerramento glotais do sinal de fala, para um correcto alinhamento e garantia de convergência óptima do algoritmo.

Krishnamurthy, A. e Childers, D. [149] demonstraram que, para vozes masculinas normais, os picos encontrados na derivada do sinal de EGG – DEGG – estão relacionados com os instantes glotais de abertura, GOI – *Glottal Opening Instant*, e encerramento, GCI – *Glottal Closure Instant*, definidos como os instantes de inicialização e terminação da variação da área glotal –tal como se observa na Figura 5.3.

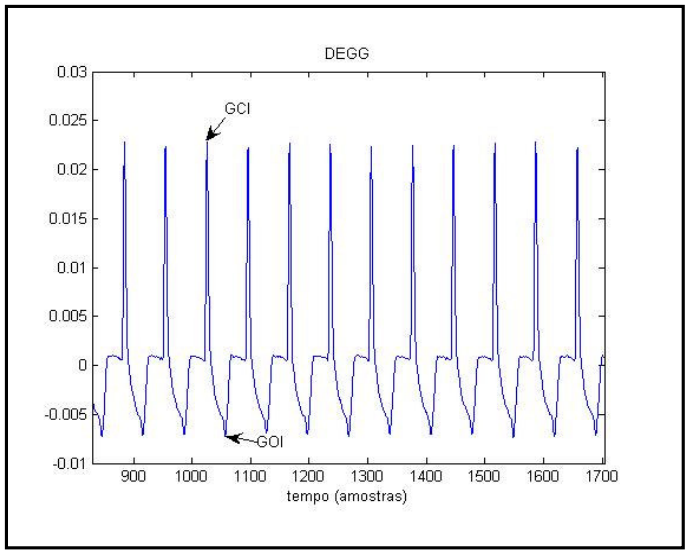

Figura 5.3 - *Derivada de EGG.*

Os autores alertam, contudo, que no caso de existir muco que crie uma ponte entre as pregas vocais durante a fase de abertura, o pico de abertura ocorre no instante em que o muco quebra e não no instante exacto de abertura glótica.

Tendo como base os vários estudos realizados, os picos de DEGG podem ser considerados como indicadores razoavelmente fiáveis dos instantes de abertura e encerramento glótico. Como o encerramento glótico é geralmente abrupto, no caso de uma fase de encerramento completa, o pico de encerramento é bem definido e apresenta uma amplitude elevada. Já a abertura pode ser menos precisa, o que se reflecte num pico de abertura de mais baixa amplitude. Tal facto associado a outros, como o da ocorrência de muco, ou a existência de excitação secundária, que pode levar à ocorrência de um falso pico de abertura, com uma amplitude superior ao real pico de abertura, levam a que o pico de abertura não seja considerado um indicador tão fiável da abertura glótica.

Contudo, pode aceitar-se que o pico de abertura glótico de DEGG ocorrerá muito proximamente do verdadeiro instante de abertura glótica. Assim, para determinar o instante de abertura glótico com maior fiabilidade, após identificação do ponto em que ocorre o pico de abertura glótica – através de um algoritmo de detecção de mínimos de um dado sinal – foram testados os 5 pontos para a frente e para trás dessa localização no que diz respeito ao erro quadrático médio – e por consequência a convergência dos parâmetros do modelo – e o ajuste entre formas de onda. Deste modo, foi possível determinar o ponto de abertura glótica que corresponde ao menor erro e ao melhor ajuste da forma de onda e que, para efeitos do algoritmo, pode ser considerado como o verdadeiro instante de abertura glótica.

Os picos máximos, os GCI, são utilizados para determinar a duração de cada período fundamental. A distância entre dois GCI corresponde ao período.

5.7.2 DETERMINAÇÃO DA ORDEM

A ordem do modelo ARX foi determinada de forma interactiva. Testaram-se várias ordens, de 8 a 14, e determinou-se aquela que produzia o menor erro quadrático médio, bem como o melhor ajuste espectral.

5.7.3 PRE-ÊNFASE

Tal como Akande, O. e Murphy, P. [174] referem, a influência da pendente espectral da fonte no sinal de fala pode desviar a correcta identificação do modelo, nomeadamente nas altas frequências, dado o carácter passa-baixo dessa pendente. É aceite que uma forma de reduzir esta influência pode ser conseguida pela aplicação de uma pré-ênfase ao sinal de fala, na forma de um filtro passa-alto. Contudo, tal como Lu, H. e Smith III, J. [175] sugerem, o parâmetro TL do modelo RK, que na prática determina a pendente espectral da fonte gerada por esse modelo, pode ser eficazmente estimado pela adição de um pólo real ao modelo. Assim, a pré-ênfase ao sinal de fala pode ser evitado já que a pendente espectral da fonte é considerado como parte do modelo.

No entanto, o modelo ARX, para ser considerado como tal, tem de apresentar um erro branco, o que não acontece devido ao efeito da pendente espectral da fonte. Quanto o erro não é branco, pode ocorrer um desvio na estimação do modelo ARX. Uma forma de ultrapassar esta dificuldade passa pelo aumento da ordem do sistema de modo a minimizar o erro, especialmente quando a razão sinal-ruído é baixa. Contudo, o aumento da ordem do modelo pode alterar as características dinâmicas do mesmo, podendo inclusive afectar a sua estabilidade. Um outro procedimento simples para compensar o efeito do pendente espectral da fonte no erro é a aplicação de uma pré-ênfase na fonte, de modo a tornar o seu espectro mais próximo do de ruído branco.

5.7.4 DETERMINAÇÃO DOS VALORES DE INICIALIZAÇÃO

Uma boa inicialização dos parâmetros dos modelos a identificar garante uma mais rápida convergência, mas acima de tudo, que o valor óptimo de facto se atinge na quase totalidade das corridas.

Exceptuando os valores do vector de estado inicial do filtro de Kalman, todos os restantes parâmetros iniciais foram determinados de forma interactiva. O vector de estado inicial do filtro de Kalman corresponde aos valores determinados por uma análise LPC, do sinal de fala na sua totalidade, de ordem igual à do modelo ARX.

5.7.5 FASE DE OCLUSÃO

Para efeitos de reconstrução do sinal de fala, em cada iteração é utilizada a média dos coeficientes do tracto vocal estimados ao longo da fase de oclusão, cuja duração é determinada a partir do parâmetro OQ do modelo RK.

Como não há garantia de estabilidade, os pólos instáveis são reflectidos para o interior do círculo unitário.

5.8 SITUAÇÕES EXPERIMENTAIS

5.8.1 SEGUIMENTO DE SINAL E IDENTIFICAÇÃO DE SISTEMA

Numa primeira fase de implementação do algoritmo de identificação conjunta dos parâmetros da fonte e tracto vocais, estudou-se a capacidade do filtro de Kalman, dada uma entrada conhecida e uma saída observada em ruído, identificar os parâmetros do sistema e proceder ao seguimento de um sinal sintético ao longo do tempo.

O sinal sintético foi gerado com recurso a uma cascata de filtros, que actuam sobre um trem de impulsos de entrada, em que cada filtro simula um dos processos da concepção fonte-filtro do processo de produção de fala. O trem de impulsos, com uma frequência de 120 Hz, a uma frequência de amostragem de 10 KHz, é inicialmente filtrado por um filtro IIR de segunda ordem que simula o filtro glotal, sendo o resultado dessa filtragem novamente filtrado por um filtro IIR de ordem 10, cujos parâmetros simulam os coeficientes de um filtro do tracto vocal para a produção da vogal /a/. Finalmente, o sinal resultante é corrompido com ruído aditivo.

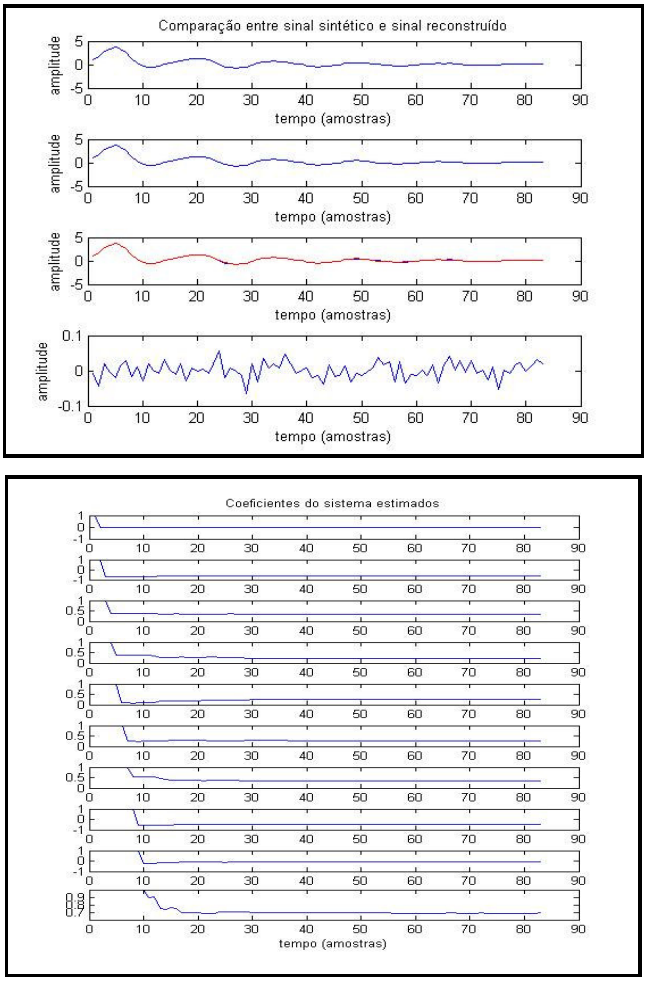

Figura 5.4 – *Em cima: Comparação, para um período, entre o sinal sinteticamente gerado e o sinal reconstruído pelo filtro de Kalman. De cima para baixo: sinal sintético; sinal reconstruído; comparação entre o sinal sintético – azul – e sinal reconstruído – vermelho; erro. Em baixo: Coeficientes do sistema estimados*

Os resultados obtidos, Figura 5.4, mostram o excelente desempenho do filtro de Kalman em seguir o sinal e estimar os parâmetros do sistema, conhecida a entrada e observada a saída em ruído. Note-se a rápida convergência dos parâmetros para um estado estável e também o reduzido erro de estimação.

Estes resultados levam a crer que, num sinal real, em que a entrada é desconhecida, mas exista uma boa estimativa da mesma, ou um bom modelo que a aproxime de forma óptima, o filtro de Kalman irá convergir para uma identificação óptima do sistema, mesmo que o sinal de saída seja observado com ruído.

5.8.2 IMPLEMENTAÇÃO

Como já foi descrito, e na concepção adoptada nesta dissertação, o objectivo de estimar conjuntamente as fonte e tracto vocais resume-se a um problema de identificação do sinal de entrada e do sistema de um modelo ARX em que ambos são à partida desconhecidos, e a única observação disponível é o sinal de fala registado por intermédio do material e métodos citados.

Com o intuito de testar a capacidade do algoritmo a implementar para o efeito, foram realizados, numa primeira fase, ensaios com sinais sintéticos, e posteriormente a sua adequação a sinais reais. Note-se que em ambas as situações experimentais foram obtidos bons resultados.

5.8.2.1 ILUSTRAÇÃO DO MÉTODO EM SINAL SINTÉTICO

5.8.2.1.1 SEM PENDENTE ESPECTRAL

A primeira situação experimental testada consistiu na aplicação do algoritmo de estimação conjunta a um sinal sintético, amostrado a 10 KHz, observado em ruído aditivo, onde o sinal de entrada é gerado por um modelo RK e o tracto vocal é modelado como um filtro IIR de ordem 10, com os coeficientes iguais aos da experiência anterior. Neste primeiro ensaio, o parâmetro TL, responsável pelo pendente espectral da fonte, não foi modelado. Os seus efeitos foram estudados na situação experimental descrita no ponto seguinte.

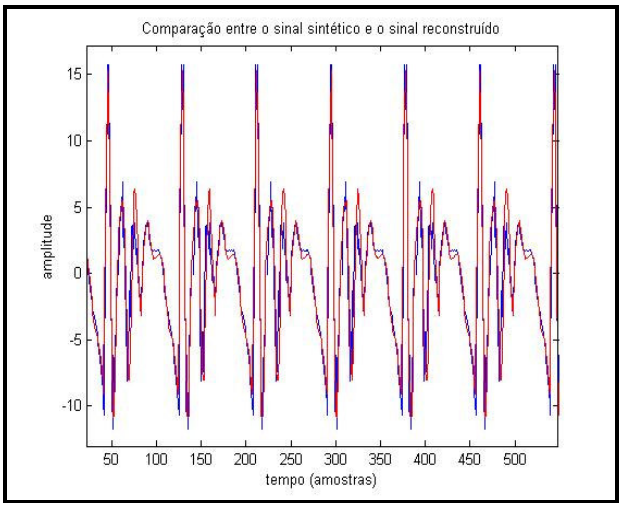

Figura 5.5 - *Comparação, para um período, entre o sinal sinteticamente gerado e o sinal reconstruído pelo método de estimação conjunta da fonte e do tracto vocais – ordem 10.*

Na Figura 5.5 observa-se a capacidade do algoritmo implementado em estimar o sinal pela optimização conjunta dos parâmetros da fonte do tracto vocais, sendo o ajuste da forma de onda excelente.

Em termos espectrais, há, contudo, uma notória dificuldade na estimação correcta dos formantes correspondentes às mais altas frequências. Como descrito atrás, tal deve-se ao facto da pendente espectral da fonte tornar o erro do modelo ARX, cujo modelo o assume branco, em não branco.

Esta dificuldade pode ser ultrapassada pela aplicação de um filtro de pré-ênfase à fonte – ver Figura 5.1. O filtro utilizado para realizar a pré-ênfase é um filtro passa-alto de primeira ordem da forma:

$$PRE(Z) = 1 - \mu z^{-1} \tag{5.18}$$

Onde μ é um valor próximo da unidade. Nesta situação experimental $\mu = 0.9$.

Ao tornar o espectro da fonte mais próximo de um espectro branco, o filtro de pré-ênfase, permite que os pólos do modelo sejam estimados com maior precisão, tal como se pode verificar nas figuras que se seguem.

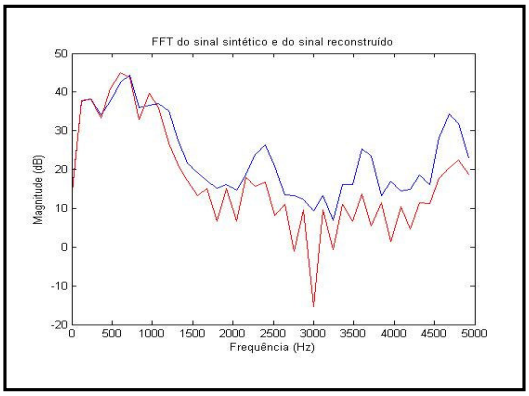

Figura 5.6 - *FFT, de 83 pontos, dos sinais sintético – azul – e reconstruído – vermelho –, sem pré-ênfase na fonte – ordem 10.*

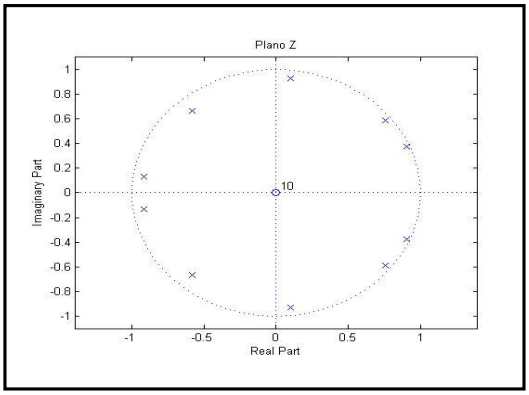

Figura 5.7 - *Plano Z com marcação dos pólos, x, e zeros, o, estimados do sistema – ordem 10 –, sem pré-ênfase na fonte.*

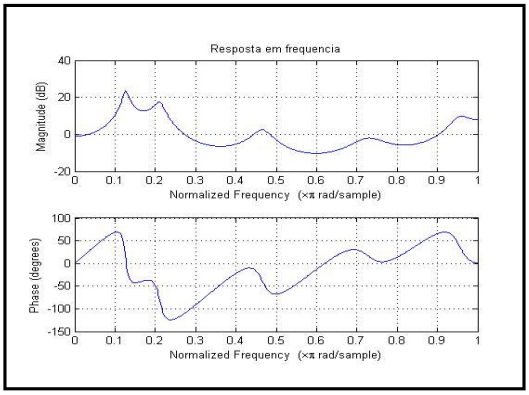

Figura 5.8 - *Resposta em frequência, normalizada, do sistema – ordem 10 –, a partir dos coeficientes estimados, sem pré-ênfase na fonte.*

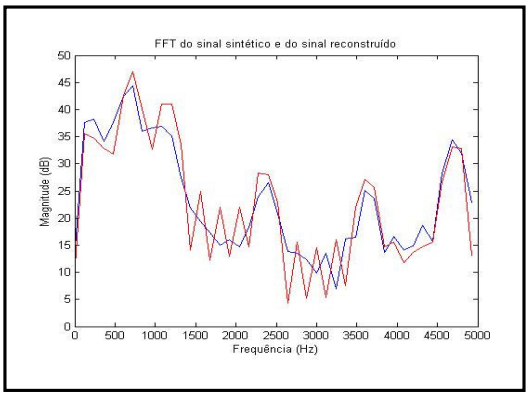

Figura 5.9 - *FFT, de 83 pontos, dos sinais sintético – azul e reconstruído – vermelho –, com pré-ênfase na fonte – ordem 10.*

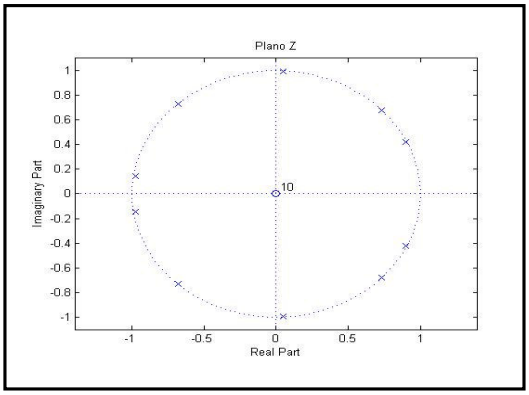

Figura 5.10 - *Plano Z com marcação dos pólos, x, e zeros, o, estimados do sistema – ordem 10 –, com pré-ênfase na fonte.*

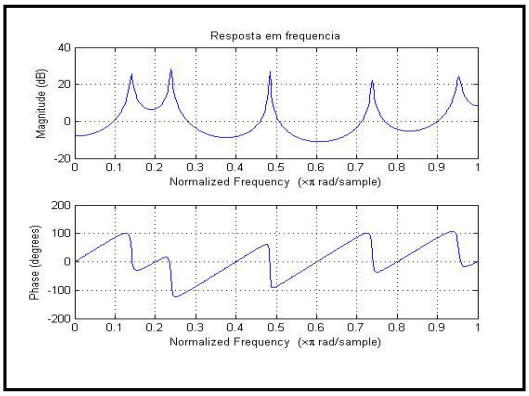

Figura 5.11 - *Resposta em frequência, normalizada, do sistema – ordem 10 –, a partir dos coeficientes estimados, com pré-ênfase na fonte.*

5.8.2.1.2 COM PENDENTE ESPECTRAL

Para estudar os efeitos da pendente espectral da fonte na estimação do sistema, ao procedimento anteriormente utilizado para gerar o sinal sintético foi agora aplicado um filtro IIR de primeira ordem, à onda gerada pelo modelo RK, da forma

$$TL(Z) = \frac{1}{1 - az^{-1}} \tag{5.19}$$

Onde a corresponde ao parâmetro TL do modelo RK. Nesta situação experimental $a = -0.9$

Note-se que em ambas as situações experimentais apresentadas à frente, procedeu-se à pré-ênfase da fonte durante o procedimento de optimização.

Na primeira situação experimental, o modelo ARX é de ordem 10. As figuras que se seguem mostramos resultados obtidos. Observe-se, na representação espectral da Figura 5.12, a falta de precisão na estimativa das frequências formantes mais elevadas, devido à influência da pendente espectral da fonte no sinal de fala.

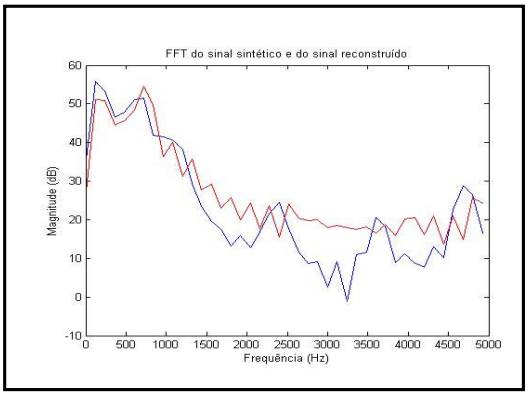

Figura 5.12 - *FFT, de 83 pontos, dos sinais sintético – azul – e reconstruído – vermelho –, com TL modelado e pré-ênfase na fonte – ordem 10.*

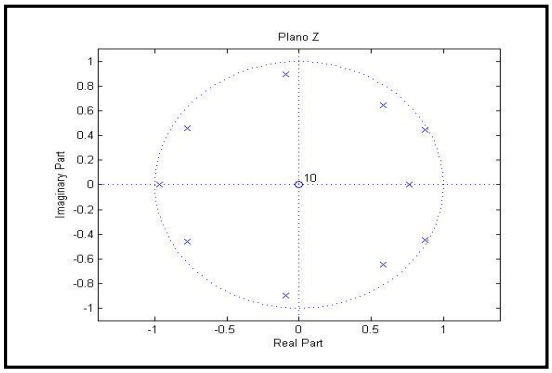

Figura 5.13 - *Plano Z com marcação dos pólos, x, e zeros, o, estimados do sistema – ordem 10 – com TL modelado e pré-ênfase na fonte.*

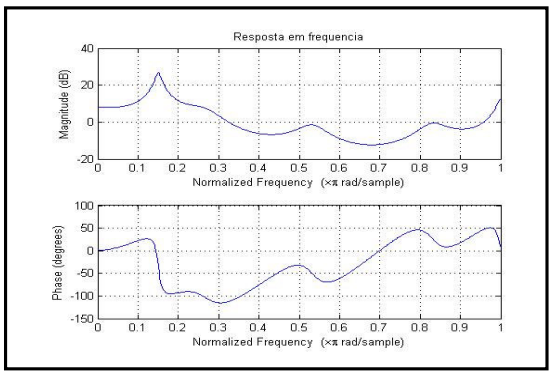

Figura 5.14 - *Resposta em frequência, normalizada, do sistema – ordem 10 –, a partir dos coeficientes estimados, com TL modelado e pré-ênfase na fonte.*

A adição de um pólo real ao sistema ARX, agora de ordem 10+1, permite uma estimação correcta das formantes mais elevadas, tal como se verifica nas figuras seguintes.

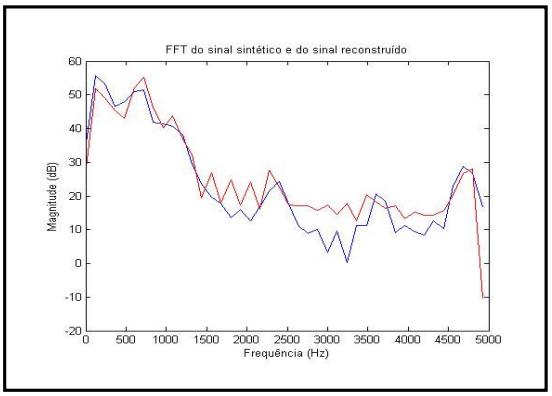

Figura 5.15 - *FFT, de 83 pontos, dos sinais sintético – azul – e reconstruído – vermelho –, com TL modelado, pré-ênfase na fonte e adição de um pólo real ao sistema – ordem 11.*

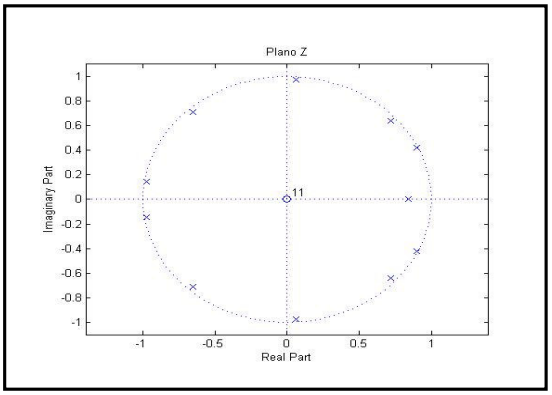

Figura 5.16 - *Plano Z com marcação dos pólos, x, e zeros, o, estimados do sistema – ordem 11 –, com TL modelado, pré-ênfase na fonte e adição de um pólo real.*

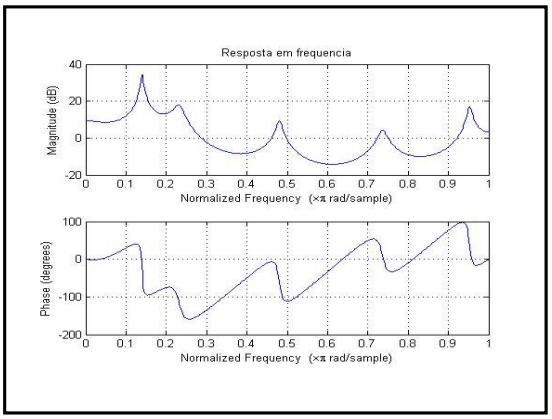

Figura 5.17 - *Resposta em frequência, normalizada, do sistema – ordem 11 –, a partir dos coeficientes estimados, com TL modelado, pré-ênfase na fonte e adição de um pólo real.*

5.8.2.2 APLICAÇÃO EM SINAL REAL

Depois de testada a eficácia do algoritmo quando aplicado a um sinal sintético, testou-se o mesmo para um sinal real – vocalização da vogal /a/ do português europeu – registado como descrito no capítulo 6.1.

A ordem escolhida, de acordo com o procedimento descrito no capítulo 5.7.2, foi 8+1.

Observe-se os resultados obtidos nas figuras que se seguem. Apesar de uma menor resolução nas mais altas frequência, obtêm-se muito boas estimativas do sistema e da fonte.

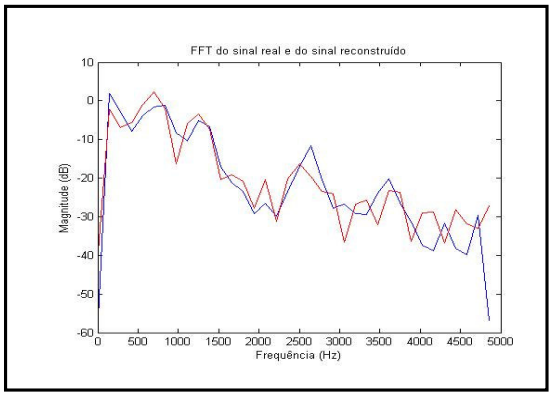

Figura 5.18 - *FFT, de 69 pontos, dos sinais real – azul – e reconstruído – vermelho –, com pré-ênfase na fonte e adição de um pólo real ao sistema – ordem 9.*

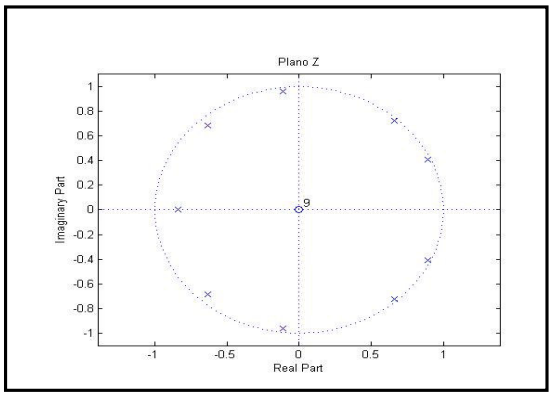

Figura 5.19 - *Plano Z com marcação dos pólos, x, e zeros, o, estimados do sistema – ordem 9 –, com pré-ênfase na fonte e adição de um pólo real.*

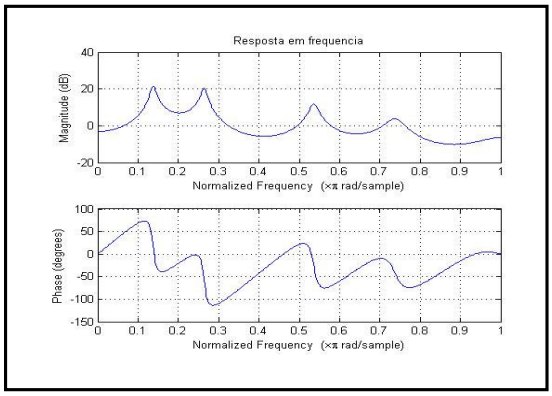

Figura 5.20 - *Resposta em frequência, normalizada, do sistema – ordem 9 –, a partir dos coeficientes estimados, com pré-ênfase na fonte e adição de um pólo real.*

5.9 TESTES PERCEPTUAIS

O sinal real, bem como o reconstruído a partir dos parâmetros da fonte e do tracto vocais, estimados pelo algoritmo implementado, foram reproduzidos para quatro ouvintes em sistema de dupla ocultação, ou seja, desconhecedores do procedimento em causa, bem como dos objectivos da experiência. Quando convidados para classificar o grau de semelhança perceptual entre os dois sinais, onde os graus de classificação possíveis eram: nada parecido, pouco parecido, razoavelmente parecido, muito parecido e igual, dois avaliadores classificaram os sinais como iguais, um classificou-os como muito parecidos, e outro como razoavelmente parecidos.

Estes resultados permitem concluir que, apesar da ligeira redução de resolução nas formantes mais altas, em termos perceptuais, a maioria dos ouvintes classifica o sinal reconstruído como muito próximo ou igual ao sinal real. Tal significa que, em termos perceptuais, apreciados conjuntamente, o modelo RK será uma boa aproximação da fonte real, a par do modelo AR do tracto vocal, e que o algoritmo implementado consegue identificar, de forma perceptualmente óptima, os parâmetros da fonte e do tracto vocais.

6 MODELO FÍSICO TRIDIMENSIONAL DO TRACTO VOCAL

6.1 AQUISIÇÃO DE IMAGENS

As imagens do tracto vocal utilizadas no presente trabalho foram cedidas por Sandra Rua, na sequência da recolha realizada, descrita em Rua, S. [109] – na Figura 6.1 estão representados exemplos de algumas das imagens obtidas. No seu estudo pretendeu caracterizar-se morfologicamente o tracto vocal, através de estudos estáticos e dinâmicos, durante a fonação dos sons da fala, para o Português Europeu. Da análise das imagens foi possível extrair contornos do tracto vocal e a consequente reconstrução tridimensional preliminar, permitindo uma visualização e uma medição parciais da forma do tracto vocal.

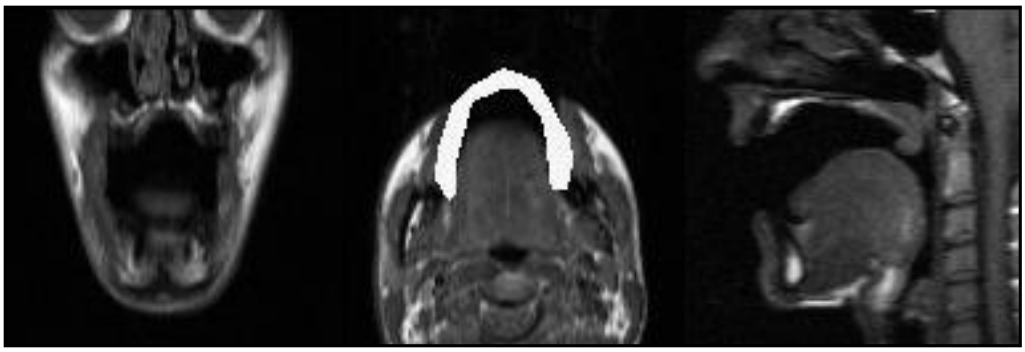

Figura 6.1 - *Exemplos de imagens recolhidas por Rua, S. [109]. Da esquerda para a direita: corte com orientação coronal, corte com orientação axial, corte com orientação sagital.*

As imagens utilizadas na presente dissertação pertencem apenas a um sujeito, do sexo masculino, saudável e com idade situada entre os 20 e 26 anos.

As imagens foram recolhidas por um equipamento de RM *Siemens Magneton Symphony* com um campo magnético de 1,5T, tendo sido utilizada uma antena *Phased Array* de cabeça para recepção do sinal.

Foram obtidos cortes representativos nos três planos anatómicos principais: na orientação sagital; orientação coronal; e orientação axial. Os três conjuntos de imagens

obtidos que representam o volume em estudo – *stacks* – incluem três cortes contíguos com 5 mm de espessura na orientação sagital; quatro cortes espaçados 10 mm e com 6 mm de espessura na orientação coronal; e quatro cortes espaçados 10 mm e com 6 mm de espessura na orientação axial.

6.2 MODELO 3D

6.2.1 SEGMENTAÇÃO DE CORTES

A partir das imagens recolhidas por RM, extraíram-se os contornos referentes ao tracto vocal - fronteira ar-tecidos moles - para cada imagem, por intermédio do *plugin 3D Editing Tool* do programa livre *ImageJ*, utilizando a técnica semi-automática de *threshold* da ferramenta *Segmenting Assistant*. Com o auxílio desta ferramenta, após delimitação da região de interesse, ROI, os contornos são automaticamente definidos através de um limiar ajustável. Ao contorno bidimensional, obtido pela delimitação em cada imagem, dá-se o nome de *outline*, e define um objecto. Cada *outline,* gerada pelo *Segmenting Assistant,* está restringida a um número máximo de 1000 pontos.

A gravação de cada conjunto de *outlines* para cada *stack* em formato *.shapes* permite visualizar as coordenadas dos pontos que formam cada *outline* através do *ImageJ* ou outros programas.

Este trabalho partiu das *outlines* obtidas por Sandra R. [109]. As coordenadas dos pontos dessas *outlines* foram introduzidas manualmente no programa de desenho 3D assistido por computador, CAD 3D, *Solid Edge* de modo a representar todos os pontos que constituem cada *outline*.

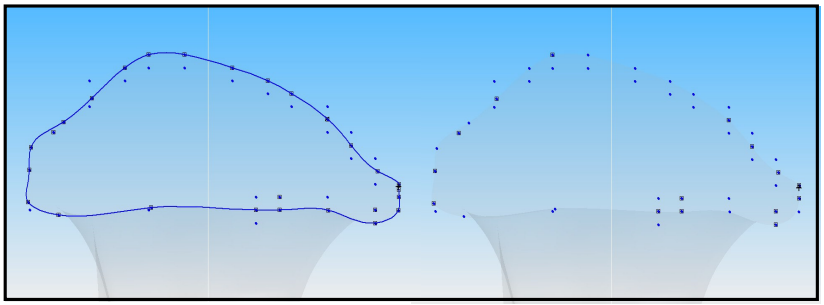

Figura 6.2 - *Pontos de uma outline coronal – esquerda; Contorno obtido a partir dos pontos – direita.*

Após introdução manual dos pontos, procedeu-se à união dos mesmos por intermédio de uma curva. Ao segmento de recta resultante foi aplicado o comando *Simplify*, pertencente ao programa de CAD 3D, que reduz o número de pontos necessários para representar a curva, mantendo uma forma aproximada da originalmente desenhada. Esta operação permite obter formas mais suavizadas – tal como se pode ver na Figura 6.2.

Uma vez obtidos os contornos, eliminou-se a contribuição dos dentes nos cortes de orientação coronal – já que estes surgem na região de interesse, segmentada anteriormente, com densidade idêntica à da cavidade do tracto vocal, ou seja próxima da do ar, visto não possuírem hidrogénio na sua constituição –, de forma interactiva, por observação das imagens RM, de orientação coronal e de orientação sagital, e ajuste manual desses mesmos contornos. A ausência total de informação sobre posição dos dentes impossibilitou a sua compensação automática e tornou a compensação manual uma tarefa árdua, propensa a erros – que se tentaram minimizar através de medições rigorosas nas imagens de RM, e pela assistência de profissionais da área de imagiologia e otorrinolaringologia.

6.2.2 CRIAÇÃO DE SUPERFÍCIE TRIDIMENSIONAL

A partir dos contornos finais obtidos, procedeu-se à união tridimensional dos mesmos. Entre cada dois contornos, foi aplicado um comando de extensão, *Swept Surface*, que cria uma superfície tridimensional, ao longo de um caminho definido pelo utilizador, entre contornos. A Figura 6.3 ilustra este processo.

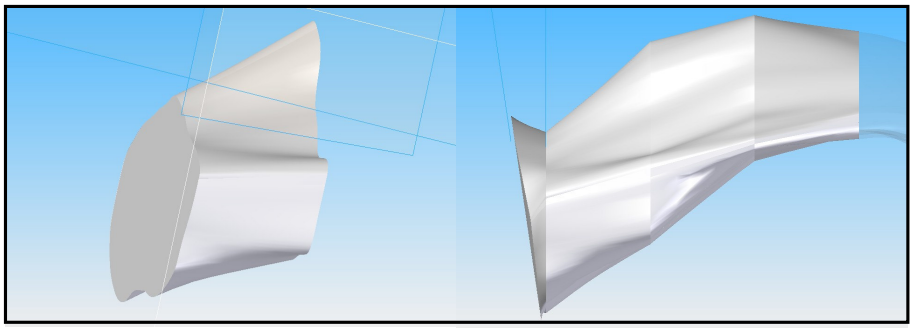

Figura 6.3 - *Swept Surface – esquerda; União tridimensional de todos os contornos coronais, mais contorno dos lábios – direita.*

Da Figura 6.4 observa-se uma correcta união entre os diferentes contornos, e a criação de uma estrutura tridimensional com a forma desejada. Note-se contudo que, pelo facto de se proceder à união faseada entre dois contornos – e esse resultado ao contorno seguinte, e assim em diante –, o sólido apresenta uma forma cuja suavidade é inferior à desejada – e, provavelmente, à real – cujo aperfeiçoamento será realizado após a produção do protótipo sólido – ver capítulo 6.3.2.

6.2.3 MODELO DOS LÁBIOS

Visto que a forma exterior dos lábios não foi amostrada nas imagens de RM, e para se conseguir uma correcta modelação da mesma, foi obtida uma fotografia frontal do sujeito a vocalizar a vogal /a/. Esta fotografia foi capturada numa data posterior à obtenção das imagens de RM, tendo sido o sujeito instruído a replicar, em termos articulatórios e acústicos, a vocalização que efectuou durante a obtenção das imagens de ressonância, o mais fiel e proximamente possível.

A partir da fotografia obtida e por intermédio do *plugin 3D Editing Tool* do programa *ImageJ*, utilizando a técnica semi-automática de *threshold* da ferramenta *Segmenting Assistant*, obteve-se automaticamente a *outline* dos lábios por ajuste manual de um limiar, *threshold level*. Este procedimento pode ser observado na Figura 6.4.

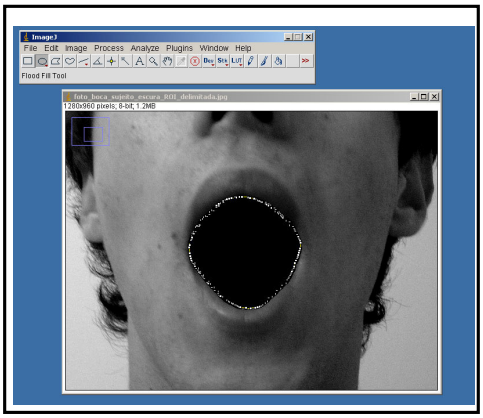

Figura 6.4 - *Outilne dos lábios obtido automaticamente pelo Segmenting Assistant*

As coordenadas dos pontos da *outline* obtida foram introduzidos manualmente no programa de CAD 3D, com posterior união desses pontos. A posição do contorno obtido, relativamente às restantes peças do modelo, foi conseguida, de forma interactiva, pela observação das imagens RM de orientação sagital – bem como da sua relação com as imagens com orientação coronal. Após correcto posicionamento, aplicou-se o comando *Swept Surface* de modo a crescer a superfície que uniu o contorno dos lábios ao primeiro contorno dos cortes com orientação coronal – ver Figura 6.4.

6.2.4 MODELO FINAL

O modelo final foi obtido por ajuste das superfícies obtidas para as orientações de corte coronal e axial. A Figura 6.5 apresenta duas vistas do modelo final.

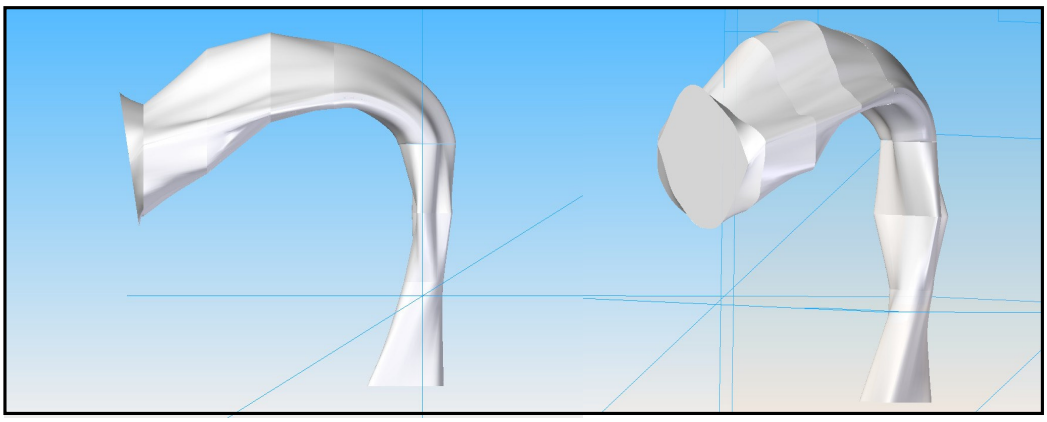

Figura 6.5 - *Modelo 3D do tracto vocal, vista lateral – esquerda; vixsta ortogonal – direita.*

Os cortes sagitais foram utilizados para alinhamento das superfícies coronal e axial, mas não figuram no modelo final por duas razões: os três cortes obtidos são muito centrais pelo que não há qualquer informação sagital lateralizada, nomeadamente na zona da boca; o sujeito parece ter sobre-articulado durante a obtenção destes cortes – notório pela observação dos contornos sagitais referentes à superfície superior da língua – o que leva a erros dimensionais da forma correcta do tracto vocal para a vocalização da vogal em questão. Obstante estas razões, os cortes sagitais foram fundamentais para: o alinhamento das superfícies de orientação coronal e orientação axial – permitindo o ajuste do alinhamento pela observação de concordância entre certos pontos; a assistência na compensação dos dentes e no correcto alinhamento dos lábios nos cortes com orientação coronal – tal como se observa na Figura 6.6.

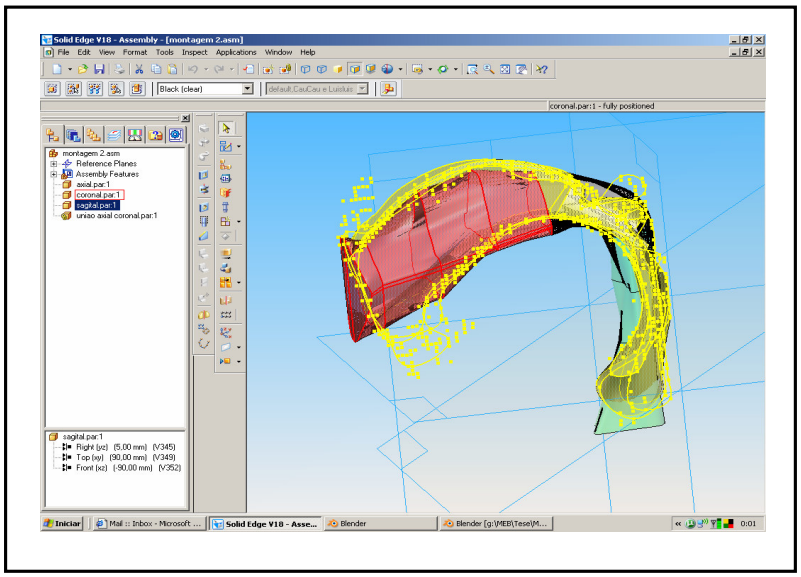

Figura 6.6 - *Fase de alinhamento das superfícies coronal – vermelho – e axial – verde – pelos cortes sagitais – amarelo – em ambiente Solid Edge.*

Da observação das imagens de RM com orientação sagital é, também, possível identificar zonas do tracto vocal, que interessa modelar, que não foram amostradas pelos cortes com orientação coronal e com orientação axial e onde a informação sagital não é fiável ou suficiente. Se a questão da forma exterior dos lábios já foi descrita acima, convém referir outras duas zonas importantes: a zona da úvula e a zona da terminação inferior do tracto vocal junto à laringe – ver Figura 6.8; para indicação das áreasonde foi necessário proceder-se à extrapolação de superfícies.

A zona da curvatura próxima da úvula foi extrapolada pelo comando *Swept Surface* que uniu o último corte coronal com o primeiro axial, de acordo com um caminho – no fundo, correspondente à curvatura – determinada pelos cortes sagitais. Note-se, contudo, que a úvula não foi incluída no modelo, pelo facto do sujeito ter uma úvula de tamanho ligeiramente superior ao normal – o que poderia dificultar a produção do protótipo sólido devido à redução muito elevada da área do tracto vocal dessa zona – bem como pela falta de informação lateral da mesma – considerando-se, aqui, que ela se encontra completamente recolhida.

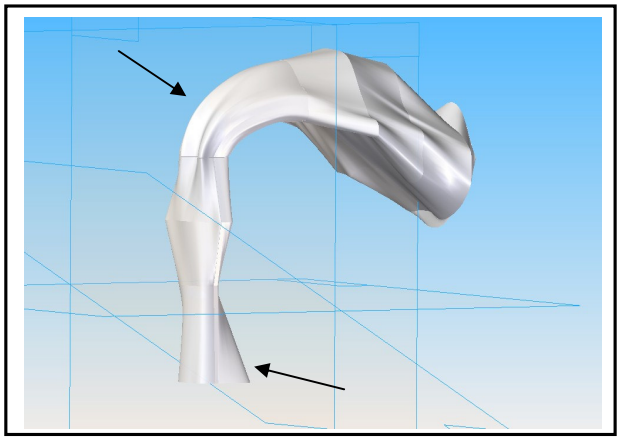

Figura 6.7 - *Modelo 3D do tracto vocal, com destaque das zonas extrapoladas.*

A terminação inferior do tracto vocal, proximamente às pregas vocais, para além de não ter sido amostrada pelos cortes com orientação axial e com orientação coronal, não foi segmentada para os cortes sagitais. Assim, esta região foi criada pela extrapolação do último corte axial, que se aumentou inferiormente. A curvatura e comprimento deste crescimento foram determinados por medições nos cortes com orientação sagital. Note-se, que a pequena formação posterior, consequente do movimento das partes moles da laringe, não foi incluída no modelo devido, mais uma vez, à ausência de informação lateral, mas também ao facto de nenhum dos cortes sagitais ter amostrado esta formação na sua posição mais recolhida – terá sido amostrada, nos três cortes sagitais, em posição que ocupa na fase de oclusão, ou de quase oclusão, do ciclo glotal. Observe-se a Figura 6.8 para uma ilustração exemplificativa desta questão – sendo possível comparar uma imagem de RM, com orientação sagital, do sujeito, onde as partes moles ocupam grande parte do tubo aéreo – para além de elemento de artefacto devido ao movimento dessas mesmas partes moles

durante a aquisição da imagem – tal como ocorre na fase de oclusão e uma outra imagem de RM, de outro sujeito, em situação de repouso, mas onde as partes moles se encontram recolhidas de forma análoga ao que se sucede na fase aberta.

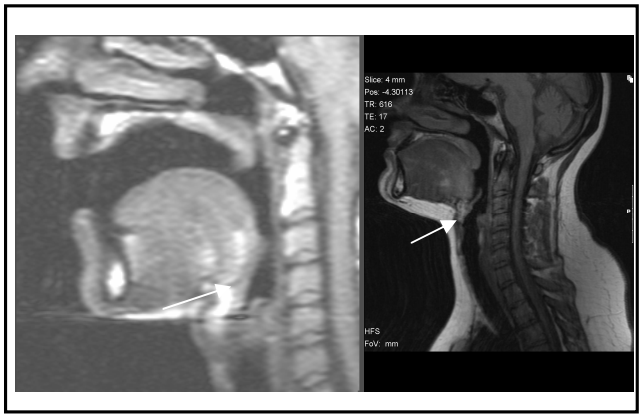

Figura 6.8 – *Ilustração da questão sobre a posição das partes moles próximas da laringe, durante a fonação. Esquerda: imagem de RM, do sujeito X, em fonação, com partes moles em posição análoga à da fase de oclusão. Direita: imagem de RM, de sujeito desconhecido, em repouso, com partes moles em posição análoga à da fase aberta. A imagem da direita foi cedida pelo departamento de Radiologia do Hospital de S. João, E.P.E.*

6.3 PRODUÇÃO DO MODELO SÓLIDO

6.3.1 PROTOTIPAGEM RÁPIDA

Stampfl, J. [176] descreve os diferentes métodos de 3D-shaping correntemente utilizados no fabrico de protótipos físicos, incluindo a prototipagem rápida. Para além dos métodos tradicionais, orientados para o ciclo de produção da indústria, as aplicações da prototipagem têm vindo a crescer e a diversificar-se nos últimos anos, incluindo na engenharia biomédica, onde ocorre frequentemente a necessidade de produzir peças de elevada complexidade e cujo fabrico é impossível por intermédio das técnicas de manufactura convencionais. O facto da prototipagem rápida ser um método adequado contribuirá para o aumento da compreensão funcional e estrutural de tais estruturas e materiais.

A principal característica comum a todas as técnicas de prototipagem rápida é o facto de se basearem numa manufactura por camadas. O modelo da peça é dividido em várias camadas cujo empilhamento automático permite a obtenção da peça final. A grande maioria dos sistemas utilizados baseia-se em técnicas aditivas. Estas podem ser ainda divididas, consoante o material utilizado: material líquido, material colóide, fáscias, etc.

6.3.1.1 ESTEREOLITOGRAFIA

A estereolitografia foi a primeira técnica de prototipagem rápida a estar comercialmente disponível. Foi desenvolvida, de forma independente, por equipas em França, EUA e Japão. Actualmente é a técnica de prototipagem rápida mais utilizada e tornou-se na ferramenta de eleição para a prototipagem industrial e médica.

A estereolitografia baseia-se na solidificação controlada de um líquido monomérico fotossensível através de um laser ou luz ultravioleta, UV, para a criação de uma peça polimérica camada após camada. O princípio de funcionamento mais comum requer uma plataforma imersa num tanque preenchido com a resina líquida fotossensível. Através de um espelho galvânico, o feixe de laser é projectado em regiões seleccionadas da superfície do líquido. Nos locais onde o laser atinge a resina, o monómero solidifica devido a uma reacção fotoquimicamente induzida. Quando o laser já atingiu todas as regiões de uma dada camada que necessitam de solidificar, a peça é coberta por uma nova camada de resina líquida. Tal é conseguido pelo abaixamento da plataforma, e consequentemente da peça cujas camadas já estão solidificadas. A espessura de cada camada pode variar entre 0.02 e 0.2 mm. Assim que se forme uma nova camada, o processo é repetido até a peça estar completa.

De modo a ser possível obter altas velocidades de construção e boa qualidade das superfícies produzidas, as máquinas de estereolitografia utilizam estratégias especiais de exposição. Tipicamente, os contornos exteriores da região a solidificar são expostos num primeiro passo. Depois, as regiões interiores são expostas separadamente.

Quando todas as camadas foram expostas, a peça pode ser removida do líquido , por subida da plataforma, e o excesso da resina líquida é removido , por lavagem com água, ficando esta, após secagem, pronta para qualquer passo de pós-processamento. Geralmente, num primeiro passo, a resina ainda incompletamente curada é completamente curada através da exposição da peça à luz UV. Depois, e para partes onde é exigida uma superfície de alta qualidade, a superfície da peça é tratada para remoção dos efeitos escada típicos da manufactura por processos de camadas. Tal pode ser conseguida pela abrasão com uma lixa suave. Também é possível pintar a peça – a tinta aplicada irá suavizar a superfície.

As principais vantagens da esterolitografia incluem: tratar-se de um método com melhor resolução geométrica e melhor qualidade da superfície produzida; as resinas recentes permitem a produção de partes com propriedades mecânicas comparáveis a muitos polímeros utilizados em engenharia; nos casos onde as resinas não permitem as propriedades mecânicas ou óptica desejadas é possível aplicar variadas técnicas de produção de moldes às peças.

As principiais desvantagens incluem: como a produção da peça é feita em imersão num líquido, é necessário criar estruturas de suporte para peças de reduzida espessura – estas estruturas têm de ser removidas posteriormente por procedimentos manuais; o monómero encolhe durante a polimerização e criam-se stresses internos que podem levar à quebra da peça – a criação de resinas avançadas permitiu minimizar este problema; só podem ser utilizados fotopolímeros o que limita à partida o número de materiais passíveis de ser utilizados para produzir protótipos.

No entanto, deve ser ressaltado que, quando se combina estereolitografia com um processo de moldagem secundário, é possível obter resultados óptimos.

Existe uma larga gama de resinas para estereolitografia comercialmente disponíveis. A maioria desses materiais baseia-se em química de tipo epoxi e depende dos requisitos individuais.

6.3.2 PRODUÇÃO DO PROTÓTIPO

O protótipo foi produzido, via esterolitografia, com resina epóxi SL 7810 da marca Huntsman, que permite a produção de protótipos com bom acabamento de superfície e detalhe. A figura 6.9 apresenta várias vistas do protótipo sólido produzido.

Figura 6.9 - *Protótipo sólido do tracto vocal produzido por estereolitografia – três vistas.*

Após lavagem e secagem, o protótipo foi desbastado mecanicamente por intermédio de uma lixa de grão 100, de modo a suavizar a superfície para o procedimento seguinte.

6.3.3 PRODUÇÃO DO TUBO ACÚSTICO

Tendo em vista a finalidade de simular acusticamente o tracto vocal medido por meio da RM, realizou-se, por moldagem, uma versão do protótipo de resina em negativo.

O protótipo de resina foi então colocado numa caixa de madeira, onde duas das paredes eram coincidentes com as duas extremidades do protótipo, para a qual se vazou silicone líquido. A silicone utilizada foi *Silastic T-4* com o catalizador *T-4º,* a 10%, que é uma silicone translúcida altamente resistente e apropriada para a fabricação de moldes.

Esta silicone foi a escolhida pois apresenta, de entre os materiais disponíveis, as características que mais se aproximavam dos parâmetros definidos por VanIngen-Dunn, C. et al. [177] para as partes moles humana, *flesh*, incluindo densidade de 0.95 g/cm^3 e dureza de 13,8 shore A. A *Silastic T-4* apresenta uma densidade de 1,2 g/cm^3 e uma

dureza de 40 shore A – de modo a reduzir a dureza foi adicionado óleo de silicone a 15%, não tendo sido determinado o valor final de densidade obtida.

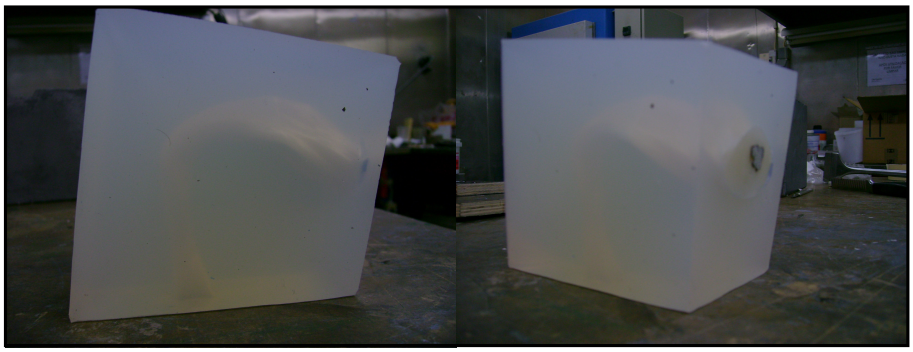

Figura 6.10 - *Protótipo do tracto vocal no interior da silicone, após vazamento e solidificação da mesma.*

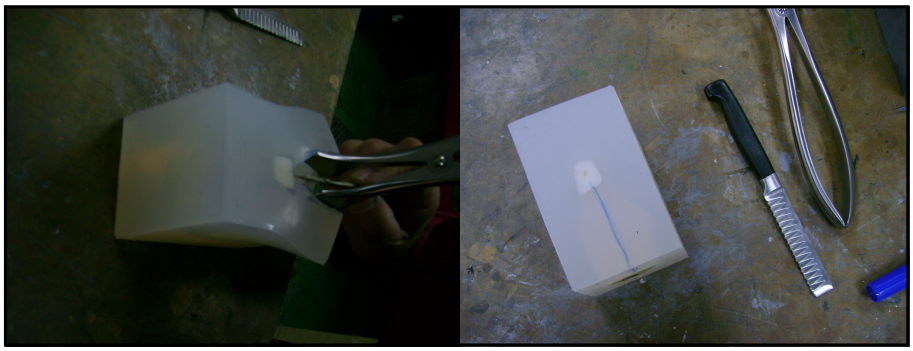

Figura 6.11 - *Remoção do protótipo do interior da silicone.*

Após remoção do protótipo do interior da silicone – tal como se observa na Figura 6.11, obteve-se um tubo, com a conformação do protótipo, aberto nas duas extremidades. A Figura 6.10 mostra o tubo acústico do tracto vocal obtido em silicone.

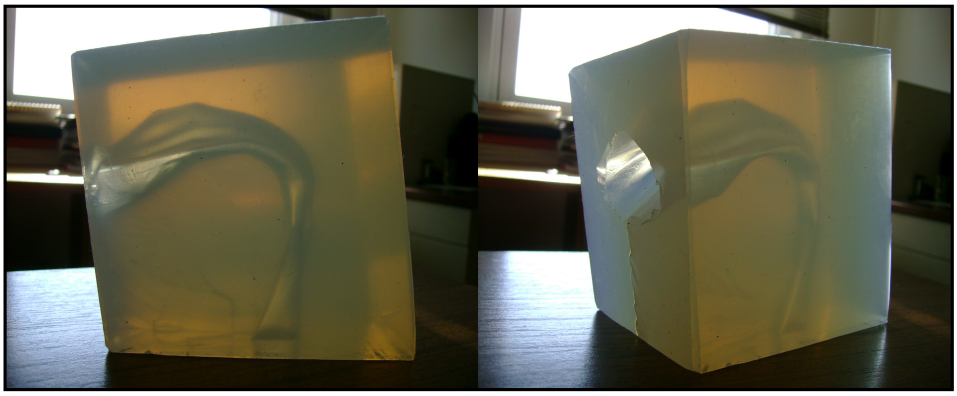

Figura 6.12 - *Tubo acústico do tracto vocal em silicone.*

7 ENSAIOS ACÚSTICOS

De modo a testar a capacidade do modelo sólido do tracto vocal em reproduzir as propriedades acústicas do tracto vocal real, procedeu-se à gravação do sinal de voz, durante a vocalização da vogal /a/ do português europeu, do mesmo sujeito a partir do qual se recolheram as imagens de RM que deram origem ao modelo sólido do tracto vocal – a partir daqui designado por sujeito X. Note-se que, devido à situação de ruído acústico demasiado elevado existente no laboratório de aquisição de imagem de RM, não foi possível a Rua, S. [109], adquirir o sinal de fala correspondente, ficando a pesquisa limitada à utilização de um sinal análogo, produzido, no entanto pelo mesmo sujeito, que, note-se, sendo profissional de saúde na área da terapia da fala, tem a faculdade de reproduzir de forma bastante próxima e dadas as restantes condicionantes, idêntica vocalização.

O algoritmo para identificação da fonte e do tracto vocais desenvolvido foi aplicado a este sinal de modo a obter-se uma estimativa da fonte – modelo RK. O sinal RK estimado foi em seguida utilizado para excitar o tracto vocal em silicone. A recolha do sinal à saída dos "lábios" do modelo em silicone permitirá comparar as propriedades acústicas do modelo e àquelas do sinal real, considerando que a fonte glotal estimada é uma aproximação óptima da fonte real.

7.1 ALINHAMENTO DE SEQUÊNCIAS PARA FILTRAGEM INVERSA

Dada a impossibilidade, por questões logísticas, de registar o sinal EGG simultaneamente com a gravação de voz do sujeito X, a determinação do início da fase de abertura do ciclo glotal, fundamental para o alinhamento de sequências e garantia de convergência do algoritmo de estimação da fonte e tracto vocais, foi realizada, primeiro, pela investigação informada da possível amostra correcta em cada ciclo, e depois, pela observação do erro nas dez amostras em diante e para trás dessa amostra. Foi escolhido o ponto que garantia o menor erro quadrático médio.

Pelas mesmas razões, dada a impossibilidade de determinar o período fundamental, só se correu o algoritmo num único período fundamental. O sinal que irá excitar o tracto vocal em silicone foi construído pela repetição do sinal RK estimado apenas para esse período.

7.2 ISOLAMENTO E EQUILIZAÇÃO DE ALTIFALANTE FONTE

De modo a realizar o ensaio acústico, onde a fonte identificada servirá como função excitatória do tracto vocal moldado em silicone, procedeu-se à construção de um sistema excitador, simulando a fonte glotal. Foram realizadas duas estruturas físicas, a partir de altifalantes de colunas comuns de tipo altifalante satélite de computador, que permitissem, por um lado, a correcta canalização da energia acústica produzida pelo altifalante para a pequena entrada do tracto vocal moldado e, por outro, a realização da gravação do sinal resultante à saída dos "lábios" sem interferência da radiação directa do referido altifalante.

Numa primeira fase foram testados vários altifalantes de pequenas dimensões, $\leq 20mm$, com o intuito de serem facilmente incorporados na pequena abertura inferior do tracto vocal, $\approx 17mm$. Nenhum dos altifalantes mostrou um desempenho satisfatório na gama de frequências a testar, de 20 a 5 KHz, exibindo, de uma foram geral, total ausência, ou muito fraca resposta nas baixas frequências e um comportamento irregular nas frequências mais altas.

Os altifalantes escolhidos para os ensaios acústicos foram-no devido à sua boa resposta em frequência, principalmente nas baixas frequências. No entanto, o seu tamanho torna-os objectos mais complicados de integrar com o tracto vocal de silicone, especialmente na correcta canalização da energia acústica, sem distorção frequencial. Daí a necessidade de se proceder a uma tarefa de igualização dos altifalantes.

7.2.1 ISOLAMENTO DE ALTIFALANTE

Este procedimento pretende garantir que não existe radiação directa da coluna ou do altifalante.

7.2.1.1 ALTIFALANTE CAIXA

O altifalante *Cambridge Soundwork*, pertencente ao sistema *Four Point Surround FPS1600* da marca Creative, foi utilizado como a base da seguinte estrutura. O altifalante foi, primeiro, envolvido numa caixa de cortiça, cujos remates foram selados a silicone. Depois, a estrutura resultante, foi colocada numa caixa de borracha de espessura de parede igual a 1 cm, cuja terminação superior se assemelha a um tronco de cone facetado no topo, realizando uma guia de onda acústica. A Figura 7.1 mostra um diagrama simplificado desta montagem, enquanto que a Figura 7.2 mostra duas imagens obtidas durante o processo de construção do sistema.

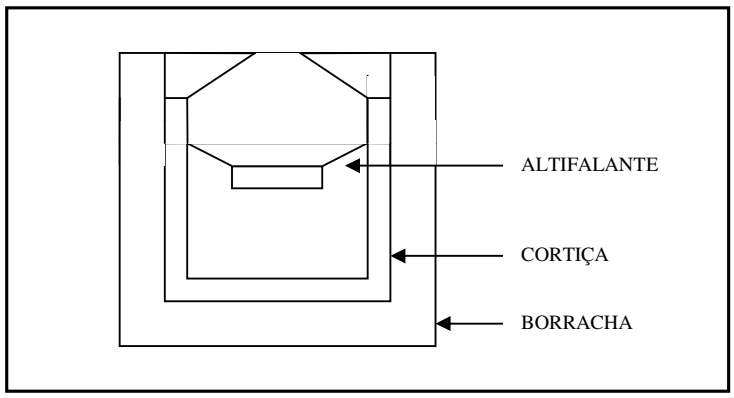

Figura 7.1 - *Diagrama simplificado da montagem do altifalante caixa.*

Figura 7.2 - *Caixa de cortiça e altifalante para coluna doméstica. Esquerda: interior da caixa de borracha com detalhe da caixa em cortiça. Direita: Vista geral da caixa de borracha antes da aplicação da tampa em forma de tronco de cone facetado no topo.*

7.2.1.2 ALTIFALANTE CONE

O altifalante pertencente ao sistema Inspire 2.1, Digital 2800, da marca Creative, foi acoplado a um tubo cónico de silicone, oco no interior, produzido por vazamento de silicone sobre um cone sólido, no interior de uma estrutura metálica, e posterior abertura das duas extremidades. A base do tubo cónico tem o mesmo diâmetro que a coluna. A zona de acoplamento foi isolada por intermédio de plasticina. A Figura 7.3 mostra um diagrama simplificado desta montagem, enquanto que a Figura 7.4 mostra duas imagens do sistema final obtido, durante processo de igualização – ver capítulo 7.2.2.

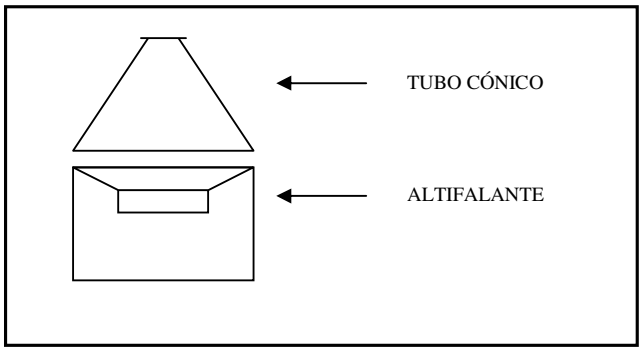

Figura 7.3 - *Diagrama simplificado da montagem do altifalante cone*

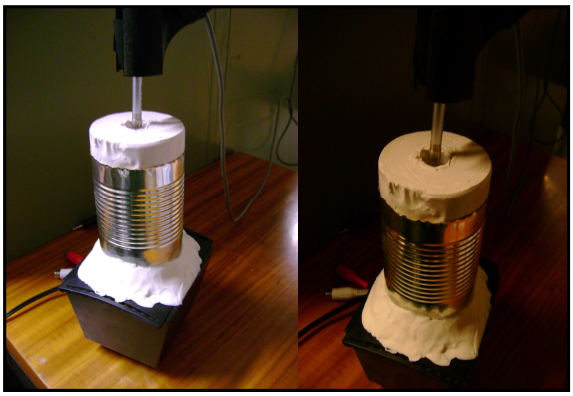

Figura 7.4 - *Vista geral do altifalante durante o procedimento de igualização.*

7.2.2 IGUALIZAÇÃO DE ALTIFALANTE

Visto que a rápida redução do diâmetro da guia de onda acústica desde a saída do altifalante até à pequena terminação por onde a energia será canalizada para o tracto vocal leva à distorção espectral da saída, foi necessário proceder à igualização dos sistemas montados.

De uma forma genérica, e neste caso em particular, a igualização pretende obter um sinal, à saída da montagem, que seja espectralmente igual àquele que sairia do altifalante, se este fosse perfeito e radiasse livremente. Para tal, é necessário determinar a resposta, em frequência inversa, do sistema que se pretende igualizar, e pré-filtrar o sinal de entrada com esse inverso.

Para um sistema de fase mínima, isto é, um sistema linear invariante no tempo, estável e causal, onde se verifica que todos os seus zeros e pólos, bem como aqueles do sistema inverso, estão dentro do círculo unitário, a igualização é um procedimento simples, pela determinação da FFT e cálculo do seu inverso – por intermédio de um simples quociente.

Já a determinação de inversos estáveis e causais de sistemas de fase não mínima, onde parte dos zeros estão fora do círculo unitário, para efeitos de igualização, é uma tarefa difícil. Uma forma de ultrapassar este problema pode ser a determinação de filtros

digitais inversos que realizem a função de transferência inversa do sistema a igualizar, conforme descrito por Marques, A. [178]. Para um sistema de fase não mínima, é possível encontrar um sistema inverso estável e causal a partir da solução estável não causal, introduzindo um atrasado adequado. Este sistema inverso aproximado "atrasado", *delayed inverse filter*, é dado por:

$$H_{inv}(z) = \frac{z^{-\Delta}}{H(z)} \tag{7.1}$$

Onde $H(z)$ é a função de transferência do sinal a igualizar, e Δ é um atraso expresso em amostras.

O atraso necessário, para tornar causal a resposta impulsional infinita não causal do filtro inverso, dependerá da proximidade ao círculo unitário dos zeros de fase não mínima do sistema a igualizar.

O sistema igualizado pode expressar-se por:

$$H(z)H_{inv}(z) = z^{-\Delta} \tag{7.2}$$

o que significa que é um sistema que introduz maior atraso que o sistema original.

De modo a ser aplicada esta concepção de igualização de sistemas de fase não mínima através de filtros digitais inversos à igualização dos sistemas montados para realizar os ensaios acústicos, foi utilizado um algoritmo desenvolvido por Marques, A. [178]. Neste algoritmo, a partir da resposta impulsional do sistema, é obtido um filtro inverso IIR cujos coeficientes são determinados por minimização do erro quadrático até que o sistema igualizado se aproxime do pretendido, ou seja, do impulso unitário, com custo computacional e atraso mínimos.

7.2.2.1 DETERMINAÇÃO DO SISTEMA INVERSO

A reposta impulsional dos sistemas a igualizar foi registada pelo microfone de condensador tipo 4135 da marca Brüel & Kjaer, aplicado na saída de som – tal como se vê na Figura 7.5 para o altifalante caixa..

Figura 7.5 - *Exemplo do registo da resposta impulsional do altifalante caixa*

Foram obtidas as seguintes respostas impulsionais:

Altifalante Caixa

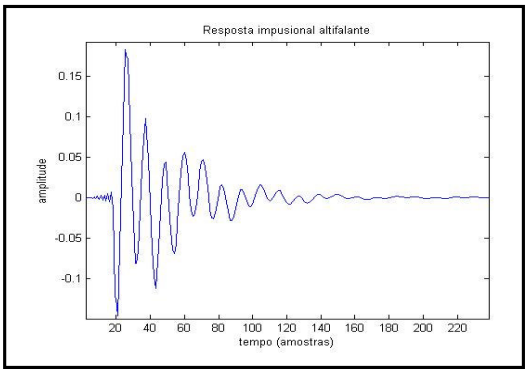

Figura 7.6 - *Resposta impulsional do altifalante caixa*

Altifalante Cone

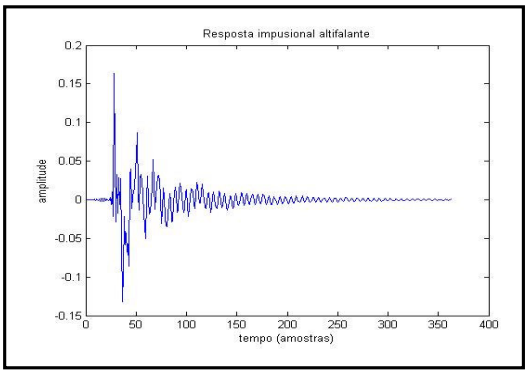

Figura 7.7 - *Resposta impulsional do altifalante cone*

Após obtenção da resposta impulsional dos dois altifalantes, foi aplicado o algoritmo de determinação do filtro inverso para se proceder à igualização dos mesmos. Foram obtidos os resultados que se seguem.

Altifalante Caixa

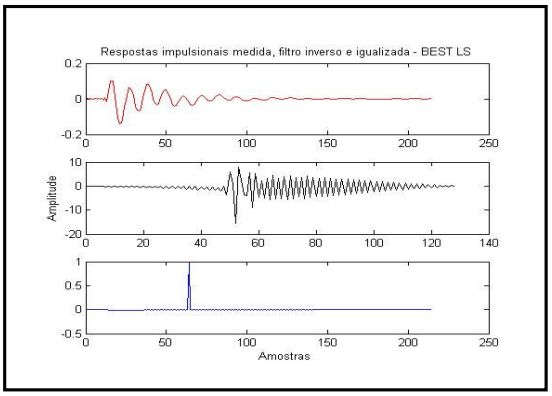

Figura 7.8 - *Algoritmo de determinação do filtro inverso – altifalante caixa. De cima para baixo: resposta impulsional do sistema; filtro inverso no tempo; resposta igualizada*

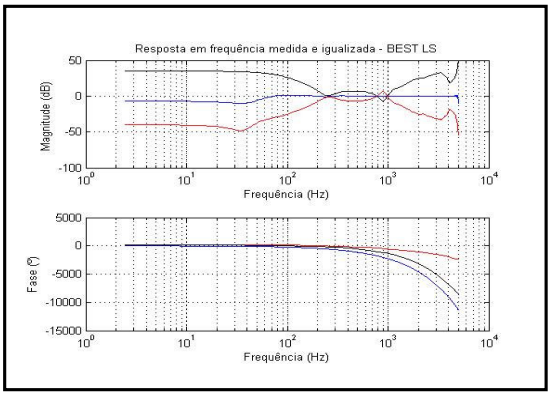

Figura 7.9 - *Respostas em frequência – altifalante caixa: medida – vermelho; inversa – preto; igualizada – azul.*

Altifalante Cone

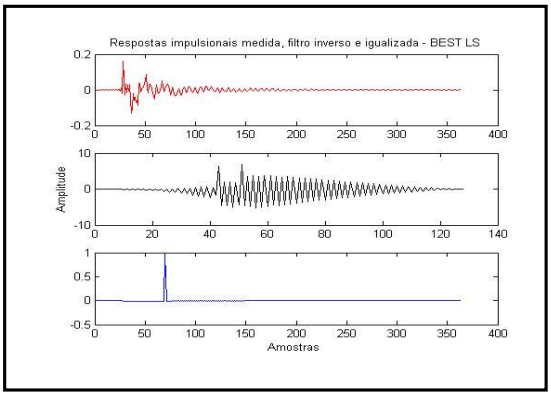

Figura 7.10 - *Algoritmo para determinação do filtro inverso – altifalante cone. De cima para baixo: resposta impulsional do sistema; filtro inverso no tempo; resposta igualizada*

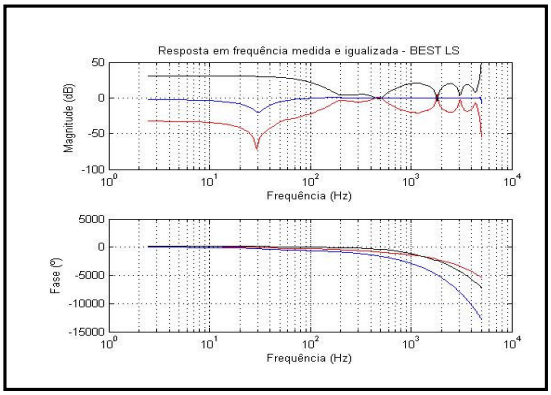

Figura 7.11 - *Respostas em frequência – altifalante cone: medida – vermelho; inversa – preto; igualizada – azul.*

7.2.2.2 APLICAÇÃO AO SINAL RK

O sinal RK foi pré-filtrado com os filtros inversos IIR determinados, e aplicado posteriormente aos altifalantes, tendo sido registado o sinal à sua saída. Os resultados obtidos são apresentados nas figuras que se seguem. Também se apresenta a análise espectral comparativa do sinal RK ideal, do sinal RK, não pré-filtrado, à saída do altifalante e do sinal RK, pré-filtrado, à saída do altifalante para cada um dos altifalantes. A análise destes resultados informou a decisão de qual altifalante escolher para utilização nos ensaios acústicos com o modelo sólido do tracto vocal – ver capítulo 7.3.

Altifalante Caixa

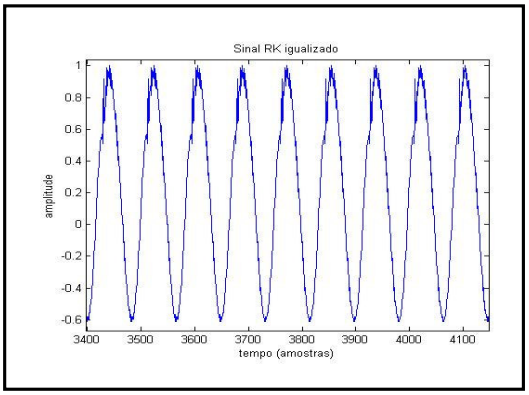

Figura 7.12 - *Sinal RK pré- filtrado pelo filtro inverso IIR, para o altifalante caixa*

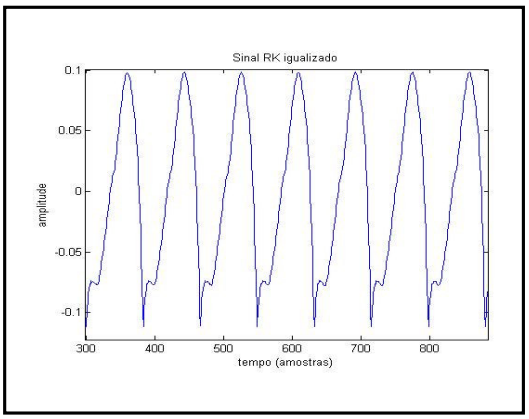

Figura 7.13 - *Sinal pré-filtrado registado à saída do altifalante caixa*

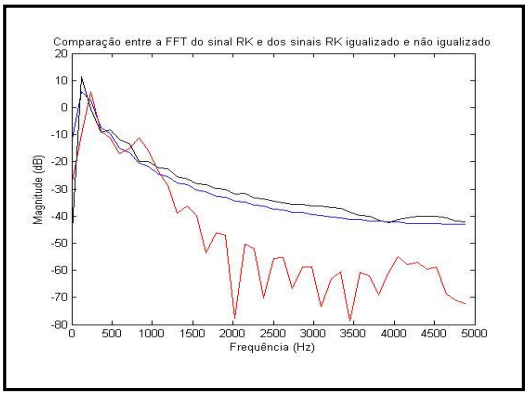

Figura 7.14 - *Espectro do sinal RK – azul, espectro do sinal RK não pré-filtrado, registado à saída do altifalante caixa – vermelho; espectro do sinal RK pré-filtrado, registado à saída do altifalante caixa – preto.*

Altifalante Cone

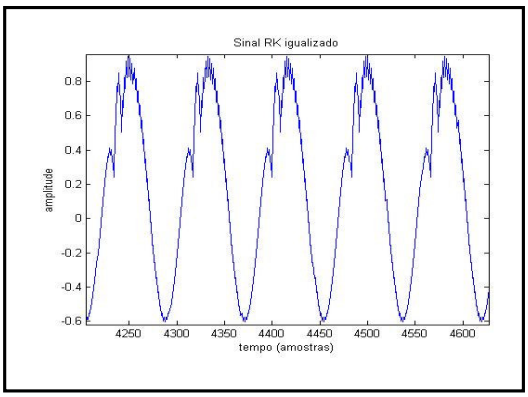

Figura 7.15 - *Sinal RK pré- filtrado pelo filtro inverso IIR, para o altifalante cone*

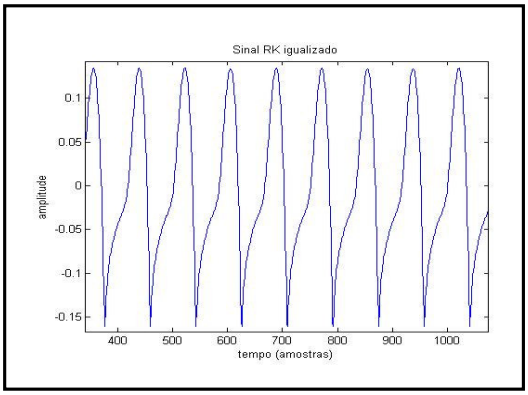

Figura 7.16 - *Sinal pré-filtrado registado à saída do altifalante cone*

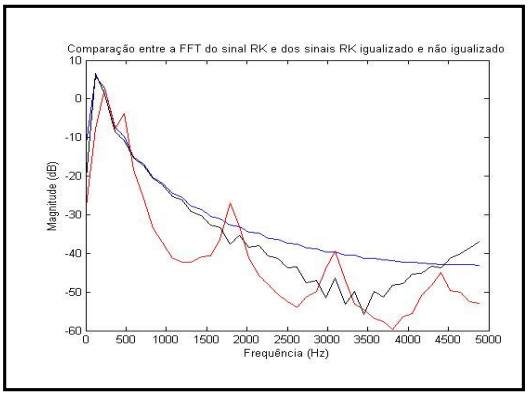

Figura 7.17 - *Espectro do sinal RK – azul, espectro do sinal RK não pré-filtrado, registado à saída do altifalante cone – vermelho; espectro do sinal RK pré-filtrado, registado à saída do altifalante cone – preto.*

Apesar do ajuste da forma de onda obtido para o sinal pré-filtrado registado à saída do altifalante cone ser muito bom, espectralmente, a igualização falha nas altas frequências, com um quebra grande a partir 1.5 KHz. Note-se, contudo, a excelente igualização para as frequências mais baixas, até aproximadamente 1 KHz. Já o ajuste espectral conseguido para o altifalante caixa parece ser mais adequado para o ensaio acústico, apesar de um ganho de cerca de 5 dB para as frequências baixas, 150 a 300 Hz, e um ganho de aproximadamente 2 dB para quase todas as frequências acima de 1 KHz. Assim, no ensaio acústico final será utilizado o altifalante caixa, ao qual será aplicado o sinal RK pré-filtrado pelo filtro inverso IIR determinado anteriormente.

7.3 ENSAIO ACÚSTICO

O sinal à saída dos "lábios" do modelo sólido do tracto vocal foi registado por um microfone supercardióide dinâmico Beta 58A da marca Shure, ligado a um pré-amplificador Pro Mike – Servo Drive Mic Preamp, da marca SPL, a uma distância de 20 cm da "boca", fora do fluxo de ar – tal como se vê na montagem experimental da Figura 7.18..

Figura 7.18 – *Montagem experimental para ensaio acústico*

Os resultados obtidos foram os seguintes:

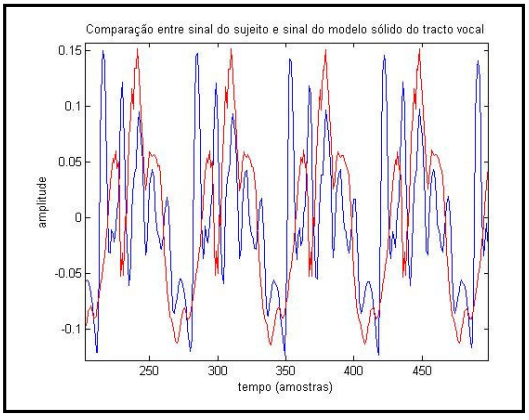

Figura 7.19 – *Comparação entre o sinal real, à saída dos lábios do sujeito – azul – e o sinal registado à saída do tubo acústico em silicone excitado pelo modelo RK pré-filtrado – vermelho –, durante o ensaio acústico*

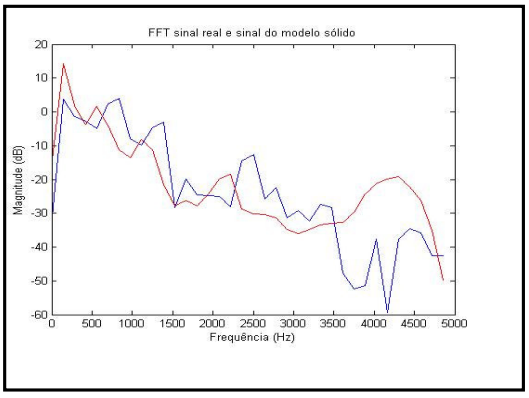

Figura 7.20 - *Sinal Real – azul; sinal registado à saída do tubo acústico em silicone excitado pelo modelo RK pré-filtrado – vermelho.*

Na análise espectral do sinal captado à saída do modelo em silicone é possível distinguir três formantes concordantes com os três primeiros formantes, a menos de um ligeiro desvio em frequência. O quarto formante, correspondente às mais altas frequências, apresenta uma concordância muito fraca, com um ganho de até 20 dB superior ao sinal real e sem um pico espectral bem definido – largura de banda muito grande.

De modo a proceder à análise perceptual subjectiva do sinal obtido à saída do modelo de silicone do tracto vocal realizou-se um pequeno teste perceptual com quatro sujeitos. O sinal real, bem como o sinal registado à saída do modelo sólido do tracto vocal, foram reproduzidos para quatro ouvintes em sistema de dupla ocultação, ou seja, desconhecedores do procedimento em causa, bem como dos objectivos da experiência. Quando convidados para referir que vogal distinguiam durante a audição do sinal registado à saída do modelo sólido do tracto vocal, os quatro avaliadores identificaram a vogal /a/, inequivocamente. Quando convidados para classificar o grau de semelhança perceptual entre os dois sinais, onde os graus de classificação possíveis eram: nada parecido, pouco parecido, razoavelmente parecido, muito parecido e igual, três avaliadores classificaram os sinais como razoavelmente parecidos e um classificou-os como pouco parecidos. Quando convidados para avaliar a naturalidade subjectiva do sinal registado à saída do modelo sólido do tracto vocal, onde os graus de classificação possíveis eram: nada natural, pouco natural, razoavelmente natural, muito natural, três avaliadores classificaram o sinal como razoavelmente natural, enquanto que um outro o classificou como pouco natural.

8 CONCLUSÕES E PERSPECTIVAS FUTURAS

Nesta dissertação propõe-se o estudo da laringe artificial electrónica como uma forma de fala alternativa para o laringectomizado, tendo em vista um projecto, de longo curso, de produção de um novo tipo de electro-laringe que ultrapasse, pelo menos em parte, as limitações dos aparelhos actualmente disponíveis.

Nesta perspectiva, o trabalho focou-se inicialmente na revisão do estado da arte das electro-laringes comercialmente disponíveis, bem como das desenvolvidas experimentalmente, de modo a identificar limitações e possíveis caminhos de investigação. No mesmo sentido, foram estudados os mais recentes trabalhos científicos na área das electro-laringes, seu melhoramento de fala e redução do ruído, bem como as diferentes alternativas ao controlo de activação e de tom.

Aceitando que numa fase inicial deste projecto, a caracterização da fonte e tracto vocais seria um objectivo obrigatório. A maior parte do trabalho, desenvolvido nesta dissertação, focou-se na fonte, nos seus componentes e sinais e no tracto vocal, incluindo a sua anatomia funcional e as suas propriedades acústicas, de forma a poder extrair o sinal de fonte e a poder excitar artificialmente o tracto vocal o mais correctamente possível.

Tendo em vista a implementação futura de uma função geradora do sinal da laringe no aparelho electrónico vicariante, estudaram-se os diversos modelos existentes bem como a obtenção do sinal glotal a partir de sinais de fala, nomeadamente numa forma parametrizada. Com base nestes estudos, e combinando características de trabalhos anteriores, desenvolveu-se, implementou-se, e testou-se um algoritmo semi-automático de estimação conjunta da fonte e tracto vocais, baseado num modelo autorregressivo com entrada exógena, ARX, só com pólos do processo de produção de fala, onde os parâmetros do tracto vocal são identificados adaptativamente por um filtro de Kalman e os parâmetros da fonte identificados por um algoritmo assente no método de optimização por solidificação simulada. O algoritmo foi optimizado para a análise de vogais sustentadas, sendo que a única vogal testada experimentalmente foi o /a/ do português europeu.

O algoritmo implementado produziu muito bons resultados, tanto nos sinais sintéticos, inicialmente usados para os testes, como nos sinais reais subsequentemente utilizados. Os testes em sinais sintéticos permitiram afinar o algoritmo e analisar detalhes de concepção e implementação, nomeadamente: o alinhamento de sequências – com recurso a processamento de sinais de electroglotografia, EGG, previamente recolhidos; a determinação da ordem do sistema – por estratégias de minimização do erro; a influência da pendente espectral da fonte no sinal de fala; a influência da pendente espectral da fonte no ruído do modelo ARX; a determinação dos instantes adequados para o cálculo dos pólos do sistema; a evolução do erro e a convergência dos algoritmos de Kalman e de solidificação simulada. Uma vez encontrada a melhor solução de desempenho para as questões acima citadas, o algoritmo foi aplicado a um sinal real. A convergência manteve-se, bem como a capacidade de aproximar de forma óptima o sinal da fonte e os parâmetros do tracto vocal. Tanto no sinal real utilizado nos testes, como no sinal real do sujeito que deu origem ao modelo sólido do tracto vocal, verificou-se uma maior dificuldade de identificação dos formantes que correspondem às mais altas frequências. Os três primeiros formantes foram, contudo, sempre estimadas de forma óptima ou quase óptima em todos os ensaios. Por outro lado, e relativamente à mais fraca precisão do algoritmo em estimar as frequências formantes mais elevadas, crê-se que, para o sujeito que deu origem ao modelo do tracto vocal – e possivelmente para os restantes sujeitos cujas vocalizações foram utilizadas nos ensaios iniciais –, existem anti-ressonâncias na sua fala, talvez devido a um ligeiro anasalamento da vocalização ou incorrecta vocalização da vogal desejada. As anti-ressonâncias não foram propositadamente modeladas no algoritmo implementado, dada a capacidade de um modelo só com pólos representar a fonação sustentada de vogais para o português europeu, pelo que este não tem capacidade de seguir o sistema optimamente nestas condições. Note-se que o modelo ARX pode ser facilmente expandido para incluir zeros, sem grande custo computacional no algoritmo – esta deverá ser uma extensão a testar em desenvolvimentos futuros.

O modelo utilizado para a fonte, cujos parâmetros foram estimados pelo algoritmo de solidificação simulada, foi o modelo RK. Este foi escolhido devido a: possuir facilidade de implementação, reduzido número de parâmetros a estimar, e capacidade para aproximar o sinal da fonte de uma vocalização normal. De acordo com os resultados obtidos, o modelo RK, parece, de facto, ser uma muito boa aproximação do sinal glotal para vocalizações normais, surgindo como um bom candidato para implementação na laringe electrónica. Note-se, contudo, que para a produção de um sinal de fala perceptualmente próxima do sinal real, ao sinal RK, deve ser adicionado *shimmer*, variação da amplitude, e *jitter,* variação da frequência ciclo-a-ciclo, com valores típicos na fonação humana. A importância destes factores, bem como a do ruído de aspiração glótico, na percepção da naturalidade do sinal de fala, e a sua implementação paramétrica na laringe electrónica, também devem ser alvo de estudos futuros. Outros modelos da fonte glotal, nomeadamente os mais complexos, tal como o LF, também devem ser estudados, e comparados com os resultados obtidos nesta dissertação para o modelo RK. Contudo, a sua integração em procedimentos de estimação conjunta da fonte e tracto vocais pode ser de mais difícil concretização. Será ainda importante estudar os benefícios perceptuais de tais modelos mais complexos, tendo também em conta o aumento do peso computacional final.

Com o objectivo de estudar as características anatomo-funcionais do tracto vocal, procedeu-se ao estudo das suas principais características anatómicas, mas também das acústicas – incluindo os modelos de representação da propagação acústica ao longo do tracto vocal. As características anatómicas foram ainda estudadas com recurso a imagens do tracto vocal em fonação – da vogal /a/ do português europeu – previamente obtidas por Ressonância Magnética, RM, tendo servido estas como base para a criação de um modelo tridimensional, 3D, computorizado do tracto vocal. O objectivo final incluía a produção de um modelo sólido do tracto vocal – a partir do modelo 3D computorizado – que se esperava ser capaz de reproduzir as propriedades acústicas do tracto vocal real e assim ser empregue, de forma sistemática, em ensaios para o desenvolvimento da futura electro-laringe, dada a dificuldade de realizar estudos *in vivo*.

O modelo foi construído no programa de CAD 3D *Solid Edge*, a partir de imagens de RM previamente obtidas e previamente. Esta tarefa revelou-se de extrema dificuldade devido, nomeadamente, à escassez de informação imagiológica – poucos cortes – incluindo a ausência de quaisquer cortes ortogonais – isto é, ortogonais numa perspectiva geral, em relação às orientações principais, como a coronal, a sagital e a axial. Outras dificuldades encontradas, relacionaram-se com: a identificação dos dentes e a compensação do seu desaparecimento nas imagens de RM – os dentes não surgem representados nas imagens de RM, surgindo, inclusive, como espaço vazio que pode interferir na segmentação automática da fronteira tecido-ar e que resulta numa importante limitação desta técnica para estudos do tracto vocal, que tem de ser ultrapassada; a possível sobre-articulação do sujeito, especialmente nos cortes na orientação sagital, o que, para além da falta de informação lateral destes cortes, impediu a sua inclusão no modelo final, tendo sido apenas utilizados, para referenciação das restantes orientações de corte; a ausência de informação precisa na zona do limite inferior do tracto vocal, especialmente da zona da laringe e na sua proximidade, assim como na zona do palato mole e da morfologia dos lábios. Parte das dificuldades foram ultrapassadas de forma interactiva, por observação e medição cuidada das imagens de RM, bem como pela assistência de profissionais da saúde ligados à otorrinolaringologia e à imagiologia. A compensação dos dentes foi conseguida pela observação das imagens na orientação coronal – sendo a posição ocupada pelos dentes identificada, de forma interactiva, e posteriormente segmentada e removida dos contornos do espaço aéreo do tracto vocal. A forma dos lábios foi obtida por intermédio da segmentação de uma fotografia da boca do sujeito a vocalizar a vogal em questão e posterior integração no modelo. A zona do palato mole foi obtida pelo crescimento de uma secção entre o último corte na orientação coronal e o primeiro corte na orientação sagital, cuja curvatura foi definida pelo utilizador, baseada na observação das imagens na orientação sagital. A porção terminal do tracto vocal, nomeadamente na proximidade imediatamente superior à laringe, foi definida pelo crescimento do último corte na orientação axial, informado pelos cortes na orientação sagital. A posição das partes moles em redor da laringe foi determinada pela extrapolação da sua posição com base numa imagem de RM de um segundo sujeito em condição de repouso.

Tendo sido este um primeiro passo no desenvolvimento de um modelo do tracto vocal – não obstante os resultados satisfatórios obtidos – espera-se que modelos futuros permitam ultrapassar as dificuldades encontradas de forma mais precisa. Para o sucesso de tal modelo é obrigatório: a obtenção de mais imagens – também com outras orientações, nomeadamente ortogonais aos planos principais; e o desenvolvimento de métodos automáticos, ou semi-automáticos, de identificação da posição dos dentes e da forma dos lábios. Também é importante estudar as capacidades de modelação 3D de outros programas que permitam obter formas finais mais suavizadas.

Será também importante o estudo da conformação do tracto vocal pós-laringectomia, bem como das suas características acústicas. É necessário perceber se existem variações acústicas perceptualmente importantes e estudar a forma de as compensar, caso existam de facto. À data de conclusão desta dissertação, já tinha sido iniciado um estudo de imagiologia com o objectivo de recolher imagens de RM do tracto vocal de indivíduos laringectomizados. Espera-se que estas imagens venham a contribuir de forma valiosa para o conhecimento das características em questão.

A partir do modelo 3D final, foi produzido um protótipo em resina epoxi pelo método de prototipagem rápida denominado de estereolitografia. A prototipagem rápida apresenta-se como uma técnica valiosa para a produção de modelos biomédicos sólidos dada a sua precisão e capacidade de produção de peças impossíveis de obter pelos métodos convencionais. O protótipo foi utilizado como molde para uma operação de vazamento de silicone com o intuito de se obter um tubo acústico com a forma exacta do modelo tridimensional e aproximado do tubo acústico do tracto vocal real.

Uma vez obtidos o sinal glotal e o modelo sólido do tracto vocal em silicone, pretendeu-se testar a capacidade de reconstruir um sinal de fala, que se esperava próximo do real, a partir desses elementos, por intermédio de um ensaio acústico.

A necessidade de utilização de um altifalante com uma resposta em frequência dentro da gama desejada para o ensaio acima citado levou a que fossem escolhidos dois altifalantes com uma dimensão muito superior à da entrada do tracto vocal em silicone. De modo a realizar o ensaio acústico, procedeu-se à construção de duas estruturas, uma em forma de caixa e outra em forma de cone, que permitem, de forma concomitante, o correcto acoplamento da energia acústica para a pequena entrada do tracto vocal e a realização da gravação do sinal resultante à saída dos "lábios" do modelo sólido, sem interferência da radiação directa do altifalante. Visto que a abrupta redução do diâmetro do espaço acústico, desde a saída do altifalante até à pequena terminação por onde a energia será canalizada para o tracto vocal, levava à distorção espectral da saída, foi necessário proceder à igualização dos sistemas montados. A igualização foi conseguida, com resultados praticamente excelentes, por intermédio de um algoritmo que, a partir da resposta impulsional do sistema, produz um filtro inverso de resposta impulsional infinita, IIR. Os coeficientes respectivos são determinados por minimização do erro quadrático, até que o sistema igualizado se aproxime do pretendido, ou seja, do impulso unitário, com custo computacional e atraso mínimos. Dos dois sistemas de altifalante montados, o sistema em forma de caixa apresentou melhores resultados na igualização e foi, por isso, o escolhido para o ensaio final.

O ensaio final consistiu na pré-filtragem do sinal RK, com os coeficientes estimados para o filtro igualizador inverso IIR, obtido pelo algoritmo de identificação da fonte e tracto vocais para o mesmo sujeito a partir do qual se recolheram as imagens de RM, e sua aplicação à entrada do tracto vocal em silicone. A pressão acústica à saída dos "lábios" do modelo sólido do tracto vocal foi registada por intermédio de um microfone dinâmico. A análise espectral do sinal registado à saída do modelo de silicone, e a sua comparação com o sinal real, mostrou uma forte concordância nos três primeiros formantes, a menos de um ligeiro desvio no valor dos mesmos. Contudo, o quarto formante, de valor mais elevado, não foi reproduzido de forma próxima do sinal real. Pelo contrário, o erro é elevado nessa região do espectro: com a ausência de um pico espectral bem definido, e uma magnitude até 20 dB superior ao sinal real. Neste momento, ainda não se encontram totalmente definidas as razões desse erro, mas qualquer um dos erros referidos para a criação do modelo em CAD 3D, ou uma combinação destes, pode estar a contribuir para tal efeito. Espera-se que a produção de novos modelos sólidos, produzidos com as devidas melhorias atrás referidas, venha a

esclarecer esta questão. É de notar, também, que a comparação realizada neste ensaio é artificial, pois o sinal de fala foi recolhido numa altura posterior à recolha as imagens de RM, devido à impossibilidade de obter gravações áudio durante o procedimento. Ou seja, a forma como foi articulada a fala que deu origem ao sinal real, poderá ser diferente da forma como foi articulada a fala durante a recolha das imagens de RM, isto devido a: existência de sobre-articulação nas imagens RM, notória nos cortes sagitais; obrigatoriedade de sustentação da vocalização por um período longo em RM; influência da gravidade e da posição corporal durante o procedimento de RM; notória dificuldade do sujeito em manter uma vocalização sustentada estável, em termos de amplitude e da frequência fundamental – tal foi observado durante a gravação da voz. Será importante desenvolver um método que permita a gravação áudio de qualidade, com boa razão sinal-ruído, do sinal de fala simultaneamente à obtenção das imagens de RM.

Note-se, contudo, que um ensaio perceptual efectuado com quatro ouvintes, ocultados dos objectivos do ensaio, mostrou que todos eles identificavam a vogal que se pretendia sintetizar, de forma inequívoca, e três deles afirmaram que o sinal à saída do tracto vocal de silicone era razoavelmente próximo do sinal de fala real e que apresentava uma naturalidade razoável.

Não abordados em profundidade nesta dissertação, existem outros objectivos de investigação necessários ao bom progresso deste projecto. A questão do controlo de activação e do tom é um deles. A julgar pelo estado da arte científica actual, os controlo miográfico e neuronal parecem ser os mais promissorers, sendo, contudo, necessários testes experimentais, de modo a avaliar a sua validade e o seu potencial concreto. Outro objectivo está relacionado com a implantabilidade da laringe electrónica. A possibilidade de introdução do som próximo do local onde se encontra a laringe num indivíduo normal, bem como a utilização de um aparelho não visível para os outros interlocutores, apresenta vantagens óbvias. Contudo, este objectivo requer uma investigação profunda no campo da medicina, mas também dos biomateriais e miniaturização de componentes electrónicos. Este ponto levanta outras questões pertinentes como, a alimentação do sistema uma vez implantado, e a capacidade de um sistema miniaturizado em reproduzir a gama de frequências desejada. Note-se que a dificuldade em obter um altifalante de pequenas dimensões com uma resposta frequencial desejada, bem como a necessidade de igualização do sistema – por exemplo,

no sistema a implantar será obrigatório compensar a influência da camada de tecido que irá revestir o dispositivo – foram tratadas de forma introdutória nesta dissertação.

Nesta dissertação foram dados os primeiros passos, conceptuais e experimentais, no desenvolvimento de um novo tipo de electrolaringe. Com base no conhecimento actual, o caminho para a obtenção de um primeiro protótipo, deverá ser longo, constituindo, contudo, um resultado muito promissor.

REFERÊNCIAS

[1] Huang, X., Acero, A., Hon, H.; *Spoken Language Procesing – A Guide to Theory, Algorithm and System Design*; Prentice Hall; 2001

[2] Deller, J., Hansen, J., Proakis, J.; *Discrete-Time Processing of Speech Signals*; IEEE Press / Wiley-Interscience; 2000

[3] Van Den Berg, J.; *Myoelastic-aerodynamic Theory of Voice Production*; The Journal of Speech and Hearing Research; Vol. 2; Nº 3; pp. 227-243; September 1958

[4] Jassar, P., England, R., Stafford, N.; *Restoration of Voice After Laryngectomy*; Journal of the Royal Society of Medicine; Vol. 92; pp. 290-302; June 1999

[5] *Detailed Guide: Laryngeal and Hypopharyngeal Cancer*; American Cancer Society; Atlanta Georgia; 2007

[6] Geertsema, A.; *Tracheostoma Valves and Their Fixation: Towards and Artificial Larynx*; Tese de Doutoramento; University of Groningen; Groningen; Holanda; 2000

[7] *Cancer Facts & Figures 2007*; American Cancer Society; Atlanta, Georgia; 2007

[8] *Cancer Facts & Figures 2006*; American Cancer Society; Atlanta, Georgia; 2006

[9] Shute, B.; *Perception of Artificial Larynx Reliability According to Laryngectomees and Speech-Language Pathologists*; PhD Thesis; Gonzaga University, EUA; 2003

[10] Weinberg, B.; *Speech After Laryngectomy: An Overview of Acoustic and Temporal Characteristics of Esophageal Speech*; in Sekey, A e Hanson, R.

(eds.), Electroacoustic Analysis and Enhancement of Alaryngeal Speech; pp. 5-48; Charles C. Thomas, Springfield; 1982

[11] Weinberg, B. e Westerhouse, J.; *A Study of Bucal Speech*; Readings in Speech Following Total Tracheostomy; University Park Press; pp. 19-25; 1971

[12] Weinberg, B. e Westerhouse, J.; *A Study of Pharyngeal Speech*; Readings in Speech Following Total Tracheostomy; University Park Press; pp. 27-34; 1973

[13] Khaila, H., House, J., Cavalli, L., Nash, E.; *A Phonetic and Phonological Study of So-Called "Bucal" Speech Produced by Two Long-Term Tracheostomised Children*; Proceedings of the 16th International Congress of Phonetic Sciences (ICPHS 2007); 2007

[14] Snidecor, J.; *Speech Rehabiliation of the Laryngectomized*; Charles C. Thomas, Springfield; 1978

[15] Snidecor, J. e Curry, E.; *Temporal and Pitch Aspects of Superior Esophageal Speech*; The Annals of Otology, Rhinology and Laryngology; Nº 68; pp. 623-636; September 1959

[16] Robbins, J., Fisher, H., Blom, E., Singer, M.; *A Comparative Acoustic Study of Normal, Esophageal and Tracheoesophageal Speech Production*; The Journal of Speech and Hearing Research; Vol. 49; pp. 202-210; May1984

[17] Diedrich, W.; *The Mechanism of Esophageal Speech*; The Annals of the New York Academy of Sciences; Vol. 155; pp. 303-317; 1968

[18] Gandour, J. e Weinberg, B.; *Production of Intonation and Constrastive Stress in Esophageal and Tracheoesophageal Speech*; Journal of Phonetics; Vol. 13; Nº 1; pp. 83-95; 1985

[19] Gandour, J. e Weinberg, B.*; Production Syntactic Stress in Alaryngeal Speech*; Language and Speech; Vol. 28; Nº 3; pp. 295-306; July 1985

[20] Gandour, J., Weinberg, B., Petty, S.; *Production Lexical Stress in Alaryngeal Speech*; Folia phoniatrica et logopaedica; Vol. 37; Nº 5-6; pp. 279-286; 1985

[21] Gates, G., Ryan, W., Cooper, J., Lawlis, F., Cantu, E., Hayashi, T., Lauder, E., Welch, R., Heake, E.; *Current Status of Laryngectomees Rehabilitation: I. Results of Therapy*; The American Journal of Otolaryngology; Vol. 3; Nº 1; pp. 1-7; January-February 1982

[22] Angermeier, C. e Weinberg, B.; *Some Aspects of Fundamental Frequency Control by Esophageal Speakers*; The Journal of Speech and Hearing Research; Vol. 24; pp. 85-91; September 1981

[23] Singer, M. e Blom, E.; *An Endoscopic Technique for Restoration of Voice After Laryngectomy*; The Annals of Otolaryngology Rhinology and Laryngology; Vol. 86; Nº 6; Pt. 1; pp. 529-533; December 1980

[24] Van den Berg, J. e Moolenaar-Bijk, A.; *Crico-Pharyngeal Sphincter, Pitch, Intensity and Fluency in Esophageal Speech*; Pratica Oto-Rhino-Laryngologica; Vol. 21; pp. 298-315; 1959

[25] Robbins, J., Fisher, H., Blom, E., Singer, M.; *A Comparative Study of Normal Esophageal and Tracheoesophageal Speech Productions*; The Journal of Speech and Hearing Disorders; Vol. 49; pp. 202-210; May 1984

[26] Blom, E., Singer, M., Hamaker, R.; *A Prospective Study of Tracheoesophageal Speech*; Archives of Otolaryngology – Head and Neck Surgery; Vol. 112; Nº 4; pp. 440-447; 1986

[27] Stalker, J., Hawk, A., Smaldino, J.; *The Intelligibility and Acceptability of Speech Produced by Five Different Electronic Artificial Larynx Devices*; The Journal of Communication Disorders; Vol. 15; Nº 4; pp. 299-307; July 1982

[28] Weiss, M. e Basili, A.; *Electrolaryngeal Speech Produced by Laryngectomized Subjects: Perceptual Characteristics*; The Journal of Speech and Hearing Research; Vol. 28; Nº 2; pp. 294-300; June 1985

[29] Gandour, J. e Weinberg, B.*; Production of Intonation and Contrastive Stress in Electrolaryngeal Speech*; The Journal of Speech and Hearing Research; Vol. 27; Nº 4; pp. 605-612; December 1984

[30] Bennet, S. e Weinberg, B.; *Acceptability Ratings of Normal, Esophageal, and Artificial Larynx Speech*; The Journal of Speech and Hearing Research; Vol. 16; Nº 4; pp. 608-615; December 1973

[31] Hillman, R.; *Improving the Assessment and Treatment of Voice Disorders: Emerging Technologies*; Proceedings of the Conference "From Sound to Sense: 50+ Years of Discoveries in Speech Communication"; B-178; June 2004

[32] Barney, H.; *A Discussion of Some Technical Aspects of Speech Aid for Post-Laryngectomized Patients*; Transactions of the American Laryngology, Rhinology and Otology Society; Vol. 79; pp. 103-115; 1958

[33] Barney, H., Haworth, F., Dunn, H.; *An Experimental Transistorized Artificial Larynx*; Bell System Technical Journal; Vol. 38; Nº 11; pp. 1337-1356; 1959

[34] www.servox.de/servox_eng/index.php

[35] griffinlab.com

[36] www.speechaid.com/nuvois.asp

[37] www.luminaud.com

[38] Choi, H., Park, Y., Lee, S., Kim, K.; *Functional Characteristics of a New Electrolarynx "Evada" Having a Force Sensing Resistor Sensor*; Journal of Voice; Vol. 15; Nº 4; pp.592-599; 2001

[39] Takahashi, H., Nakao, M., Kikuchi, Y., Kaga, K.; *Alaryngeal Speech Aid Using an Intra-Oral Electrolarynx and a Miniature Fingertip Switch*; Auris Nasus Larynx; Vol. 32; Nº 2; pp. 157-162; 2005

[40] Fujisaki, H.; *Modeling in the Study of Tonal Features of Speech With Application to Multilingual Speech Synthesis – Invited Plenary Paper*; Proceedings of the Joint International Conference on SNLP and Oriental COCOSDA 2002; Vol. 2002-5; pp. D1-D10; 2002

[41] Stern, K.; *A Self Contained Intra-Oral Artificial Larynx*; Progress Report; December 1978

[42] Kibuchi, Y. e Kasuya, H.; *Development and Evaluation of Pitch Adjustable Electrolarynx*; Proceedings of the International Conference on Speech Prosody; pp. 761-764; March 2004

[43] Matsushima, J., Uemi, N., Ifukube, T., Takahashi, M.; *Design of a New Electrolarynx Having a Pitch Control Function*; Proceedings of the 3rd. IEEE International Workshop on Robot and Human Communication; pp. 198-203; July 1994

[44] Ifukube, T.; *From Sensory Substitute Technology to Virtual Reality Research*; Artificial Life and Robotics; Vol. 2; Nº 4; December 1998

[45] Sugie, N., Tsunoda, K.; *A Speech Prosthesis Employing a Speech Synthesizer – Vowel Discrimination from Perional Muscle Activities and Vowel Production*; N. Sugie and K. Tsunoda; IEEE Transactions on Biomedical Engineering; Vol. BME-32; Nº 9; pp. 485-490; July 1985

[46] Cole, K., Knopacki, R., Abbs, J.; *A Miniature Electrode for Surface Electromyography During Speech*; Journal of the Acoustical Society of America; Vol. 74; pp. 1362-1366; November 1983

[47] Houston, K., Hillman, R., Kobler, J., Meltzner, G.; *Development of Sound Source Components For a New Electrolarynx Speech Prosthesis*; Proceedings of the 1999 IEEE International Conference on Acoustics, Speech and Signal Processing; Vol. 4; pp. 2347-2350; March 1999

[48] Heaton, J., McMahon, T., Kobler, J., Barry, D., Goldstein, E., Hillman, R.; *Recurrent Laryngeal Nerve Transposition in Guinea Pigs*; The Annals of Otolaryngology Rhinology and Laryngology; Vol. 109; Nº 10; pp. 972-980; October 2000

[49] Heaton, J., Goldstein, E., Kobler, J., Zeitels, S., Randolph, G., Walsh, M., Gooey, J., Hillman, R.; *Surface Electromyographic Activity in Total Laryngectomy Patients Following Laryngeal Nerve Transfer to Neck Strap Muscles*; The Annals of Otolaryngology Rhinology and Laryngology; Vol. 113; Nº 9; pp.754-764; 2004

[50] Goldstein, E., Heaton, J., Kobler, J., Stanley, G., Hillman, R.; *Design and Implementation of a Hands-Free Electrolarynx Device Controlled by Neck Strap Muscle Electromyographic Activity*; IEEE Transactions on Biomedical Engineering; Vol. 51; Nº 2; pp. 325-332; February 2004

[51] Griffiths, M., Fredrickson, J., Bryce, D.; *An Implantable Electromagnetic Sound Source for Speech Productions*; Archives of Otolaryngology – Head and Neck Surgery; Vol. 102; Nº11; November 1976

[52] Painter, C., Fredrickson, J., Kaiser, T., Roanne, K.; *Human Speech Development For an Implantable Artificial Larynx*; The Annals of Otolaryngology Rhinology and Laryngology; Vol. 96; Nº 5; pp.573-577; 1987

[53] Shoureshi, R., Chaghajerdi, A., Aasted, C., Meyes, A.; *Neural-Based Prosthesis for Enhanced Voice Intelligibility in Laryngectomees*; Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering; pp. 173-176; March 2003

[54] Aguilar, G., Meana, H., Miyatake, M., Mendoza, H.; *Speech Enhancement of Voice Produced by an Electronic Larynx*; The 47th Midwest Symposium on Circuits and Systems; Vol. 3; pp. 37-40; July 2004

[55] Aguilar, G., Meana, H., Miyatake, M.; *Enhancement and Restoration of Alaryngeal Speech Signals*; Proceedings of the 16th IEEE International Conference on Electronics, Communications and Computers; pp. 31-26; February 2006

[56] Holly, S., Lernman, C., Randolph, K.; *A Comparison of the Intelligibility of Esophageal, Electrolarynx and Normal Speech in Quiet and in Noise*; The Journal of Communication Disorders; Vol. 16; pp. 143-155; March 1983

[57] Norton, R. e Bernstein, R.; *Improved Laboratory Prototype Electrolarynx (LAPEL): Using Inverse Filtering of Frequency Response Function of the Human Throat*; The Annals of Biomedical Engineering; Vol.21; Nº 2; pp. 163-74; March 1993

[58] Espy-Wilson, C., Chari, V., Huang, C.; *Enhancement of Alaryngeal Speech by Adaptative Filtering*; Proceedings of the Fourth International Conference on Spoken Language (ICSLP 96); Vol. 2; pp. 764-767; October 1996

[59] Niu, H., Wan, M., Wang, S., Liu H.; *Enhancement of Electrolarynx Speech Using Adaptative Noise Cancelling Based on Independent Component Analysis*; The Journal of Medical and Biological Engineering and Computing; Vol. 41; Nº 6; pp. 670-679; November 2003

[60] Cole, D., Sridharan, S., Moody, M., Geva, S.; *Application of Noise Reduction Techniques for Alaryngeal Speech Enhancement*; Proceedings of the 10th Annual Conference on Speech and Image Technologies for Computing and Telecommunications; Vol. 2; pp. 491-494; December 1997

[61] Krubsack, D. e Niederjohn, R.; *Estimation of Noise Corrupting Speech Using Extracted Speech Parameters*; Digital Signal Processing: A Review Journal; Vol. 4; Nº 3; pp. 154-172; July 1994

[62] Pandey, P., Bhandarkar, S., Bachher, G., Lehana, P.; *Enhancement of Alaryngeal Speech Using Spectral Subtraction*; Proceedings of the 14th International Conference on Digital Signal Processing (DSP 2002); Vol. 2; pp. 591-594; 2002

[63] Pratapwar, S., Pandey, C., Lehana, P.; *Reduction of Background Noise in Alaryngeal Speech Using Spectral Subtraction With Quantile Based Noise Estimation*; Proceedings of the 7th World Multiconference on Systematics, Cybernetics and Informatics (SCI 2003); pp. 408-413; July 2003

[64] Liu, H., Zhao, Q., Wan, M.; Wang, S. *Enhancement of Electrolarynx Speech Based on Auditory Masking*; IEEE Transactions on Biomedical Engineering; Vol. 53, Nº 5; pp. 865-874; May 2006

[65] Liu, H., Zhao, Q., Wan, M., Wang, S.; *Application of Spectral Subtraction Method on Enhancement of Electrolarynx Speech*; The Journal of the Acoustical Society of America; Vol. 120; Nº 1; pp. 398-406; July 2006

[66] Strothjohann, M. e Buzug, T.; *Speech Enhancement for an Artificial Larynx Using a Low-Dimensional Model of the Hearing Process*; Proceedings of the 26th Annual International Conference of the IEEE ; Vol. 1; pp. 707-710; September 2004

[67] Strothjohann, M. e Buzug, T.; *Clatter Reduction for Electronic Artificial Larynx*; International Journal of Speech Technology; Vol. 8, Nº 3; pp. 271-281, September 2005

[68] Williams, P., Warwick, R., Dyson, M., Bannister, L.; *Gray's Anatomy: Thirty-Seventh Edition*; Churchill Livingstone; London; 1989

[69] Lucero, J.; *Dynamics of the Vocal Fold Osciallation*; Selected Papers from the XXVII Congresso Nacional de Matemática Aplicada e Computacional; Vol. 6; pp.11-20; 2005

[70] Deliyski, D., Petrushev, P.; *Methods for Objective Assessment of High-Speed Videoendoscopy*; Proceedings of the 6th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL-2003); Nº 28; Vol. 6; pp. 1-16; April 2003

[71] Pulakka, H.; *Analysis of Human Voice Production Using Inverse Filtering, High-Speed Imaging, and Electroglottography*; Masther Thesis; Department of Computer Science and Engineering; Helsinki University of Technology; Helsínquia; Finlândia; February 2005

[72] Henrich, N., Hess, M., Schade, G., Neubauer, J., Mantay, C., Kirchoff, T.; *The Transilumination Technique and its Applications: First Results*; Proceedings of the 6th International Conference on Advances in Quantitive Laryngology, Voice and Speech Research; 2006

[73] Mitra, P.; *Glottography for the Diagnosis of Vocal Disorders*; M. Tech. Credit Seminar Report; Electronic Systems Group; Electronic Engineering Department; IIT Bombay; November 2004

[74] Fabre, P.; *Un Procédé Électrique Percutané d' Inscription de l'Accolement Glottique au Cours de la Phonation : Glottographie de Haute Fréquence*; Bulletin de L' Académie Nationale de Médecine; Vol. 141 ; Nº 3-4 ; pp.66-69 ; January 1957

[75] Marasek, K.; *EGG and Voice Quality - Tutorial*; em http://www.ims.uni-stuttgart.de/phonetik/EGG/pagee1.htm; 1997

[76] Childers, D. e Larar, J.; *Electroglottography for Laryngeal Function Assessment and Speech Analysis*; IEEE Transactions on Biomedical Engineering; Vol. BME-31, Nº 12; pp. 807–817; December 1984

[77] Frokjaer-Jensen, B.; *Construction and Comparative Tests of Two Different Types of Glottographs*; Proceedings of the 17th National Congress on Otolaryngology; Denmark; 1969

[78] Fujisaki, H. e Ljungqvist, M.; *Proposal and Evaluation of Models for the Glottal Source Waveform*; Proceedings of the 1986 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '86); Vol. 11; pp. 1605-1608; April 1986

[79] Childers, D.; *Speech Processing and Synthesis Toolboxes*; Wiley; 1999

[80] Flanagan, J. e Landgraf, L.; *Self-Oscillating Source for Vocal-Tract Synthesizers*; IEEE Transactions on Audio and Electroacoustics; Vol. 16; Nº 1; March 1968

[81] Drioli, C.; *Voice Coding By Means of Physically-Based Models* – PhD Thesis; Università degli Studi di Padova; Padua; 2002

[82] Ishizaka, K. e Flanagan, J.; *Synthesis of Voice Sound from a Two-Mass Model of the Vocal Cords*; Bell System Technical Journal; Vol. 51; Nº 6; pp. 1233-1268; 1972

[83] Koizumi, T., Taniguchi, S., Hiromitsu, S.*; Two-Mass Models of the Vocal Cords for Natural Sounding Voice Synthesis*; The Journal of the Acoustical Society of America; Vol. 82; Nº 4; pp. 1179-1192; October 1987

[84] Hirano, M.; *Morphological Structure of the Vocal Cord as a Vibrator and its Variations*; Folia Phoniatrica; Vol. 26; Nº 2; pp. 89-94; 1974

[85] Titze, I; *The Human Vocal Cords: A Mathematical Model. I.*; Phonetica; Vol. 28; Nº 3; pp. 129-70; 1973

[86] Titze, I. e Strong, W.; *Normal Modes in Vocal Cord Tissues*; The Journal of the Acoustical Society of America; Vol. 57; Nº 3; pp. 736-744; March, 1975

[87] Titze, I.; *The Physics of Small-Amplitude Oscillation of the Vocal Folds*; The Journal of the Acoustical Society of America; Vol. 83; Nº 4; pp. 1536-1552; April 1988

[88] Story, B. e Titze, I.; *Voice Simulation With a Body-Cover Model of the Vocal Folds*; The Journal of the Acoustical Society of America; Vol. 97; Nº 2; pp. 1249-1260; February 1995

[89] Markel, J. e Gray, A.; *Linear Prediction of Speech*; New York: Springer-Verlang; 1976

[90] Rosenberg, A.; *Effect of Glottal Pulse Shape on the Quality of Natural Vowels*; Journal of the Acoustical Society of America; Vol. 49; pp.583-590; 1971

[91] Fant, G.; *Glottal Source and Excitation Analysis*; Speech Music and Hearing – Quarterly Progress and Status Report ; Kungliga Tekniska Högskolan; Vol. 20; Nº 1; pp. 70-85; 1979

[92] Fant, G.; *Voice Source Analysis – A Progress Report*; Speech Music and Hearing – Quarterly Progress and Status Report ; Kungliga Tekniska Högskolan; Vol. 20; Nº 3-4; pp. 31-53; 1979

[93] Ananthapadmanabha, T.; *Acoustic Analysis of Voice Source Dynamics*; Speech Music and Hearing – Quarterly Progress and Status Report ; Kungliga Tekniska Högskolan; Vol. 25; Nº 2-3; pp. 1-24; 1984

[94] Fant, G., Liljencrants, J., Lin, Q.; *A Four Parameter Model of Glottal Flow*; Speech Music and Hearing – Quarterly Progress and Status Report ; Kungliga Tekniska Högskolan; Vol. 26; Nº 4; pp. 1-13; 1985

[95] Klatt, D. e Klatt, L.; Analysis, *Synthesis, and Perception of Voice Quality Variations Among Female and Male Talkers*; The Journal of the Acoustical Society of America; Vol. 87; Nº 2; pp. 820-857; February 1990

[96] Fant, G.; *The LF-Model Revisited: Transformations and Frequency Domaina Analysis*; Speech Music and Hearing – Quarterly Progress and Status Report ; Kungliga Tekniska Högskolan; Vol. 36; Nº 2-3; pp. 119-156; 1995

[97] Rothenberg, M; *An Interactive Model for the Voice Source*; Vocal Fold Physiology: Contemporary Research and Clinical Issues; College Hill Press; San Diego; 1983

[98] Hedelin, P.; *A Glottal LPC-Vocoder*; Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '84); Vol. 9; pp. 21-24; March 1984

[99] Milenkovic, P.; *Voice Source Model For Continuous Control of Pitch Period*; The Journal of the Acoustical Society of America; Vol. 93; Nº2; pp. 1087-1096; 1993

[100] Childers, D.; *Glottal Source Modeling for Voice Conversion*; Speech Communication; Vol. 16; Nº2; pp. 127-138; February 1995

[101] Veldhuis, R.; *A Computationally Efficient Alternative for the LF Model and Its Perceptual Evaluation*; The Journal of the Acoustical Society of America; Vol. 103; Nº 1; pp. 566-571; January 1998

[102] Williams, P., Warwick, R., Dyson, M., Bannister, L.; *Anatomia de Gray – Trigésima Sétima Edição;* Editora Guanabara-Koogan S.A.; Rio de Janeiro; 1995

[103] Roach, P.; A Little Encyclopaedia of Phonetics; 2002; disponível em: http://www.personal.reading.ac.uk/~llsroach/encyc.pdf

[104] Sondhi, M.; *Model for Wave Propagation in a Lossy Vocal Tract*; The Journal of the Acoustical Society of America; Vol. 55; Nº 5; pp. 1070-1075; May 1974

[105] Portnoff, M.; *A Quasi-One-Dimensional Digital Simulation for the Time Varying Vocal Tract – Master Thesis*; Masschussets Institute of Technology; 1973

[106] Sondhi, M.; *Resonances of a Bent Vocal Tract*; Journal of the Acoustical Society of America; Vol. 79; Nº 4; pp. 1113-1116; April 1986

[107] Kim, Y.; *Singing Voice Analysis / Synthesis*; Tese de Doutoramento; Massachussets Institute of Technology; September 2003

[108] Ridouane, R.; *Investigating Articulation in Speech Production: A Review of Some Techniques*; lpp.univ-paris3.fr/equipe/rachid_ridouane/Ridouane_Investigating.pdf; 2006

[109] Rua, S.; *Estudo Morfológico-Dinâmico do Tracto Vocal Humano*; Dissertação de Mestrado; Faculdade de Engenharia da Universidade do Porto; Setembro 2006

[110] Maeda, S., Berger, M., Engwall, O., Laprie, Y., Maragos, P., Potard, B., Schoentgen, J.; 2006; *Acoustic-to-Articulatory Inversion: Methods and Acquisition of Articulatory Data*; Special Targeted Research Program; aspi.loria.fr/Save/survey-1.pdf; 2006

[111] Wrench, A., MacIntosh, A., Hardcastle, W.; *Optopalatograph (OPG): A New Apparatus For Speech Production Analysis*; Proceedings of the Fourth International Conference on Spoken Language (ICSLP 96); Vol. 3: Nº 3-6; pp. 1580-1592; October 1996

[112] Vaseghi, S.; *Advanced Digital Signal Processing and Noise Reduction – Third Edition*; Wiley, London; 2006

[113] Rokkaku, M., Imaizumi, S., Niimi, S., Kiritani, S.; *Measurement of the Three-Dimensional Shape of the Vocal Tract on the Magnetic Resonance Imaging Technique*; Annual Bulletin of the Research Institute for Logopedics and Phoniatrics; Nº20; pp. 47-54; 1986

[114] Baer, T., Gore, J., Boyce, S., Nye P.; *Application of MRI to the Analysis of Speech Production;* Magnetic Resonance Imaging; Vol. 5; Nº 1; pp. 1-7; 1987

[115] Matsumura, M., Sugira, A.; *Modeling of Three-Dimensional Vocal Tract Shapes Obtained by Magnetic Resonance Imaging for Speech Synthesis*; Proceedings of

the 1st International Conference on Spoken Language Processing (ICSLP 90); pp. 425-428; 1990

[116] Baer, T., Gore, J., Gracco L., Nye P.; *Analysis of Vocal Tract Shape and Dimensions Using Magnetic Resonance Imaging: Vowels*; The Journal of the Acoustic Society of America; Vol. 90; Nº 2; pp.799-828; 1991

[117] Lakshiminarayanan, S., Lee, S., McCutcheon, M.; *MR Imaging of the Vocal Tract During Vowel Production*; Journal of Magnetic Resonance Imaging; Vol. 1; Nº1; 1991

[118] Foldvik, A., Husby, O., Kvaerness, J., Norlic, I., Rinck, P.; *MRI Film of Articulatory Movements*; Proceedings of the 1st International Conference on Spoken Language Processing (ICSLP 90); pp. 421-422; 1990

[119] Sulter, A., Miller, M., Wolf, R., Schutte, H., Wit H., Mooyaart, E.; *On the Relation Between the Dimensions and Resonance Characteristics of the Vocal Tract: A Study With MRI*; Magnetic Resonance Imaging; Vol. 10; Nº 3; pp. 365-373; 1992

[120] Badin, P. e Serrurier, A.; *Three-Dimensional Modeling of Speech Organs: Articulatory Data and Models*; IEICE Technical Report; Vol. 106; Nº 177; pp. 29-34; September 2006

[121] Greenwood, A., Goodyear, C., Martin, P.; *Measurements of Vocal Tract Shapes Using Magnetic Resonance Imaging*; IEE Proceedings I, Communications, Speech and Vision; Vol. 139; Nº 6; pp. 553-560; December 1992

[122] Tiede, M., Yeahia, H., Vatikiotis-Bateson, E.; *A Shape Based Approach to Vocal Tract Area Function Estimation*; Proceedings of the 4th Speech Productions Seminal; pp.41-44; 1996

[123] Story, B., Titze, I., Hoffman, E.; *Vocal Tract Area Functions from Magnetic Resonance Imaging*; The Journal of the Acoustical Society of America; Vol. 100; Nº2; pp. 537-554; July 1996

[124] Soquet, A., Lecuit, V., Metens T., Demolin, D.; *From Sagittal Cut to Area Function : an RMI Investigation*; Proceedings of the 4th. International Conference on Spoken Language Processing (ICSLP 96); pp. 1205-1208; 1996

[125] Narayanan, S., Alwan, A., Haker, K.; *Three-Dimensional Tongue Shapes of Sibilant Fricatives*; Proceedings of the 128th Meeting of the Acoustical Society of America; Journal of the Acoustical Society of America; Vol. 96; Nº5; pp. 3342; November 1994

[126] Matsumura, M., Niikawa, T., Shimizu, K., Hashimoto, Y., Morita, T.; *Measurement of 3D Shapes of Vocal Tract, Dental Crown and Nasal Cavity Using MRI: Vowels and Fricatives*; Proceedings of the 3rd. International Conference on Spoken Language Processing (ICSLP 04); pp. 619-622; 1994

[127] Story, B., Titze, I., Hoffman, E.; *Vocal Tract Area Functions from Magnetic Resonance Imaging*; The Journal of the Acoustical Society of America; Vol. 100; Nº2; pp. 537-554; July 1996

[128] Demolin, D., Metens, T., Soquet, A.; *Three-Dimensional Measurement of the Vocal Tract by MRI*; Proceedings of the 4th. International Conference on Spoken Language Processing (ICSLP 96); pp. 272-275; 1996

[129] Soquet, A., Lecuit, V., Metens, T., Nazarian, B., Demolin, D. ; *Segmentation of the Airway From the Surrounding Tissues on Magnetic Resonance Images : A Comparative Study* ; Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP 98); pp. 3083-3086; 1998

[130] Behrends, J. e Wismüller, A.; *A Segmentation and Analysis Method for MRI Data of the Human Vocal Tract*; Forschungsberichte des Institus für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM); Nº 37; pp. 179-189; 2001

[131] Badin, P., Bailly. G., Raybaudi, M., Segebarth, C.; *A Three-Dimensional Linear Articulatory Model Based on MRI Data*; Proceedings of the 3rd ESCA / COCOSDA International Workshop on Speech Synthesis; pp. 249-254; 1998

[132] Badin, P., Borel, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C.; *Towards and Audiovisual Virtual Talking Head: 3D Articulatory Modeling of Tongue, Lips and Face Based on MRI and Video Images*; Proceedings of the 5th Seminar on Speech Production: Models and Data & C REST Workshop on Models of Speech Production; pp. 261-264; May 2000

[133] Kröger, B., Winkler, R., Mooshammer, C., Pompino-Marschall, B.; *Estimation of Vocal Tract Area Function From Magnetic Resonance Imaging: Preliminary Results*; Proceedings of the 5th Seminar on Speech Production: Models and Data & C REST Workshop on Models of Speech Production; pp. 333-336; May 2000

[134] Xiao, B., Bier, P., Thorpe, W.; *A Time-Varying Three-Dimensional Model of the Vocal Tract*; Proceedings of the 11th Australian International Conference on Speech Science & Technology; pp. 159-164; December 2006

[135] Engwall, O.; *Vocal Tract Modeling in 3D*; Speech Music and Hearing – Quarterly Progress and Status Report ; Kungliga Tekniska Högskolan; Vol. 40; Nº 1-2; pp. 31-38; 1999

[136] Chiba, T. e Kajima, M.; *The Vowel, Its Nature and Structure*; Phonetic Society of Japan; Tokyo; 1941

[137] Arai, T. ; *The Replication of Chiba and Kajiyama's Mechanical Models of the Human Vocal Cavity* ; The Journal of the Phonetic Society of Japan ; Vol. 5; Nº 2; pp. 31-38; August 2001

[138] Fujita, S., Kitamura, T., Honda, K.; *Toward Finding Morphological Factors for Deriving Speaker Characteristics Based on Vowel Synthesis from Vocal Tract Solid Models*; Proceedings of the 2003 Autumn Meeting of the Acoustical Society of Japan; Vol. 1; pp. 375-376; 2003

[139] Pavel, R., Jaromir, H., Jan, V.; *Rapid Prototyped Models of Human Vocal Tract Used for Experimental Acoustic Modal Analysis*; Proceedings of the 22nd Danubia-Adria Symposium on Experimental Methods in Solid Mechanics; Paper 66; Session B; October 2005

[140] Fant, G.; *Acoustic Theory of Speech Production*; Mouton, Paris; 1970

[141] Story, B.; *An Overview of the Physiology, Physics and Modeling of the Sound Source for Vowels*; Journal of Acoustical Science and Technology; Vol. 23; Nº 4; pp. 195-206; 2000

[142] Atal, B. e Hanauer, S.; *Speech Analysis and Synthesis by Linear Prediction of the Speech Wave*; The Journal of the Acoustical Society of America; Vol. 50; Nº 2B; pp. 637-655; August 1971

[143] Wong, D., Markel, J., Gray, A; *Least Squares Glottal Inverse Filtering from the Acoustic Speech Waveform*; IEEE Transactions on Acoustics, Speech, and Signal Processing; Vol. ASSP-27; Nº 4; pp. 350-355; August 1979

[144] Alku, P.*; Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering*; Speech Communication; Vol. 11; Nº 2-3; pp.109-118; June 1992

[145] Alku, P. e Laine, U.; *A New Glottal LPC Method for Voice Coding and Inverse Filtering*; IEEE International Symposium on Circuits and Systems ; Vol.3; pp. 1831-1834; May 1989

[146] Backstrom, T., Alku, P., Vilkman, E.; *Time Domain Parameterization of the Closing of Glotal Airflow Waveform from Voices Over a Large Intensity Range*; IEEE Transactions on Speech and Audio Processing; Vol. 10; Nº 3; pp. 186-192; March 2002

[147] El-Jaorudi, A. e Makhoul, J.; *Discrete All-Pole Modeling*; IEEE Transactions on Signal Processing; Vol. 39; Nº 2; pp. 411-423; February 1991

[148] Strube, H.; *Linear Prediction on a Warped Frequency Scale*; Journal of the Acoustical Society of America; Vol. 68; Nº 4; pp. 1071-1076; 1980

[149] Krishnamurthy, A. e Childers, D.; *Two Channel Speech Analysis*; IEEE Transactions on Acoustics, Speech, and Signal Processing; Vol. 34; Nº 4; pp. 730-743; August 1986

[150] Bickley, C. e Stevens, K.; *Effects of the Vocal Tract Constriction on the Glottal Source: Experimental and Modelling Studies*; Journal of Phonetics; Nº 14; pp. 373-382; 1986

[151] Ljunqvist, M., Fujisaki, H.; *A Method For Simultaneous Estimation of Voice Source and Vocal Tract Parameters Based on Linear Predictive Analysis*; Transactions of the Committee on Speech Research, Acoustical Society of Japan; Nº S85-21; pp. 153-160; 1985

[152] Fujisaki, H., Ljunqvist, M.; *Estimation of Voice Source and Vocal Tract Parameters Based on ARMA Analysis and a Model For the Glottal Source Waveform*; Proceedings of the 1987 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '87); Vol. 12; pp. 637-640; April 1987

[153] Ding, W. e Kasuya, H.; *A Novel Approach to the Estimation of Voice Source and Vocal Tract Parameters from Speech Signals*; Proceedings of the Fourth International Conference on Spoken language (ICSLP 96); Vol. 2; pp-1257-1260; October 1996

[154] Ding, W., Campbell, N., Higuchi, N., Kasuya, H.; *Fast and Robust Joint Estimation of Vocal Tract and Voice Source Parameters*; Proceedings of the 1997 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '97); Vol. 2; Nº 21-24; pp. 1291-1294; April 1997

[155] Kasuya, H., Maekawa, K., Kiritani, S.; *Joint Estimation of Voice Source and Vocal Tract Parameters as Applied to the Study of Voice Source Dynamics*;

Proceedings of the International Congress of Phonetic Science (ICPhS99); pp. 2505-2512; 1999

[156] Ding, W., Kasuya, H., Adachi, S. ; *Simultaneous Estimation of Vocal Tract and Voice Source Parameters Base don an ARX Model*; IEICE Transactions in Information and Systems; Vol. E78-D, Nº 6; pp. 738-743; June 1995

[157] Fu, Q. e Murphy, P.; *Adaptative Inverse Filtering for High Accuracy Estimation of the Glottal Source*; 2003 International Speech Communication Association Tutorial and Research Workshop on Non Linear Speech Processing (NOLISP 03); Paper 018; May 2003

[158] Fu, Q. e Murphy, P.; *Robust Glottal Source Estimation Based on Joint Source-Filter Model Optimization*; IEEE Transactions on Audio, Speech, and Language Processing; Vol. 14; Nº 2; pp. 492-501; March 2006

[159] Funaki, K., Miyanaga, Y., Tochinai, K.; *A Time Varying ARMAX Speech Modelling With Phase Compensation Using Glottal Source Model*; Proceedings of the 1997 IEEE International on Acoustics, Speech and Signal Processing (ICASSP '97); Vol. 2; pp. 1299-1302; April 1997

[160] Lu, H. e Smith III, J.; *Joint Estimation of Vocal Tract Filter and Glottal Source Waveform via Convex Optimization*; Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics; pp. 79-82; 1999

[161] Frölich, M., Michaelis, D., Lessing, J., Strube, H., Kruse, E.; *Inverse Filtering and Simultaneous Model Matching*; Proceedings of the 5th International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research; 2001

[162] Fröhlich, M., Michaelis, D., Strube, H.; *SIM – Simultaneous Inverse Filtering and Matching of a Glottal Flow Model for Acoustic Speech Signals*; The Journal of the Acoustical Society of America; Vol. 110; Nº 1; pp. 479-488; July 2001

[163] Shiga, Y. e King, S.; *Estimation of Voice Source and Vocal Tract Characteristics Based on Multi-frame Analysis*; Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003); Vol. 3; pp. 1737-1740; September 2003

[164] Jinachitra, P. e Smith III, J.; *Joint Estimation of Glottal Source and Vocal Tract for Vocal Synthesis Using Kalman Smoothing and EM Algorithm*; Proceedings of the 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics; pp. 327-330; October 2005

[165] Gobl, C.; *The Voice Source in Speech Communication*; Tese de Doutoramento; The Royal Institute of Techonology; Estocolmo; Suécia; 2003

[166] Sondhi, M; *Measurement of the Glottal Waveform*; Journal of the Acoustical Society of America; Vol. 57; Nº 1; pp. 228-232; 1975

[167] Ding, W., Kasuya H., Adachi, S.; *Simultaneous Estimation of Vocal Tract and Voice Source Parameters Based on an ARX Model*; IEICE Transactions on Information and Systems; Vol. E78-D; Nº 6; pp. 738-743; June 1995

[168] *Matla*b; www.mathworks.com

[169] *Praat – Doing Phonetics By Computer*; www.fon.hum.uva.nl/praat/

[170] Kalman, R.; *A New Approach to Linear Filtering and Prediction Problems*; Transactions of the ASME – Journal of Basic Engineering; Nº 82 (Series D); pp. 35-45; 1960

[171] Van Laarhoven, P. e Aarts, E.; *Simulated Annealing: Theory and Applications*; Kluwer Academic Publishers; Dordrecht; 1987

[172] Kirkpatrick, S., Gelatt, Jr., C., Vecchi, M.; *Optimization by Simulated Annealing*; Science; Vol. 220; Nº 4598; pp. 671-680; May 1983

[173] Černy, V.; *A Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm*; Journal of Optimization Theory and Applications; Vol. 45; Nº1; pp. 41-51; January 1985

[174] Akande, O. e Murphy, P.; *Estimation of Vocal Tract Transfer Function With Application to Glottal Wave Analysis*; Speech Communication; Vol. 46; Nº 1; pp. 15-36; May 2005

[175] Shiga, Y. e King, S.; *Estimation of Voice Source and Vocal Tract Characteristics Based on Multi-frame Analysis*; Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003); Vol. 3; pp. 1737-1740; September 2003

[176] Stampfl, J.; *3D-Techniques in Materials Science*; Habilitation in Material Sciences; Technische Universität Wien; Viena; Março 2007

[177] VanIngen-Dunn, C., Hurley, T., Yaniv, G.; *Development of Humanlike Flesh Material for Prosthetic Limbs*; Proceedings of the 15th Annual International Conferene of the IEEE Engineering in Medicine and Biology Society; pp. 1313-1314; 1993

[178] Marques, A.; *Filtros Digitais Inversos para Igualização de Sistemas Electroacústicos de Fase Não Mínima*; Dissertação de Doutoramento; Faculdade de Engenharia da Universidade do Porto; 2005

ERRATA

EM DIRECÇÃO A UMA LARINGE ARTIFICAL ELECTRÓNICA – FUNDAMENTOS TÉCNICO-CIENTÍFICOS E ENSAIOS PRELIMINARES

PÁGINA 9, LINHA 6

Onde se lê: "…consequências negativa de ordem psicológica, social e económicas."
Deve ler-se: "…consequências negativas de ordem psicológica, social e económica."

PÁGINA 23, LINHA 5

Onde se lê: "…e a cavidade nasal nariz."
Deve ler-se: "…e a cavidade nasal."

PÁGINA 28, LINHA 17

Onde se lê: "…estima-se existem…"
Deve ler-se: "…estima-se que existam…"

PÁGINA 32, LINHA 8

Onde se lê: "…posterior à laringectomia, mas num período posterior, acarretando…"
Deve ler-se: "…posterior à laringectomia, acarreta…"

PÁGINA 32, LINHA 11

Onde se lê: "…electrolaringe…"
Deve ler-se: "…electro-laringe…"

PÁGINA 40, LINHA 7

Onde se lê: "…circuito RF…"
Deve ler-se: "…circuito de radiofrequência, RF,…"

PÁGINA 44, LINHA 15

Onde se lê: "…relacionados com assincronia…"
Deve ler-se "…relacionados com a assincronia…"

PÁGINA 66, LINHA 3

Onde se lê: "…feixe ultra sónico manter separado ou numa só palavra contínuo é…"
Deve ler-se "…feixe ultrasónico é…"

PÁGINA 67, LINHA 5

Onde se lê: "...de alta frequência e mas…"
Deve ler-se: "de alta frequências mas…"

PÁGINA 70, LINHA 18

Onde se lê: "…sistema subglótica…"
Deve ler-se "…sistema subglótico…"

PÁGINA 77, LINHA 1

Onde se lê: "…predição linear, ver abaixo."
Deve ler-se: "…predição linear, ver capítulo 3.5.2.2."

PÁGINA 84, LINHA 8

Onde se lê: "…por Rosenberg [90]."
Deve ler-se "…por Rosenberg, A. [90]."

PÁGINA 85, FIGURA 2.15

Onde se lê: "…parâmetro TL."
Deve ler-se: "…parâmetro TL – esquerda."

PÁGINA 85, LINHA 5

Esta linha deveria estar no final da página 83 e não na página 85.

PÁGINA 89, FIGURA 3.1

Onde se lê: "De [102]."
Deve ler-se: "De Williams, P. [102]."

PÁGINA 90, FIGURA 3.2

Onde se lê: "De [102]."
Deve ler-se: "De Williams, P. [102]."

PÁGINA 95, LINHA 19

Onde se lê: "…posição articulatório…"
De ler-se: "…posição articulatória…"

PÁGINA 98, LINHA 23

Onde se lê: "…assumir propagação…"
Deve ler-se: "…assumir a propagação…"

PÁGINA 101, LINHA 1

Onde se lê: "O o tubo…"
Deve ler-se: "O tubo…"

PÁGINA 101, LINHA 3

Onde se lê: "…F_i, ais que…"
Deve ler-se: "…F_i, tais que…"

PÁGINA 111, LINHA 4

Onde se lê: "…várias técnicas do mesmo…"
Deve ler-se: "…várias técnicas de análise do mesmo…"

PÁGINA 133, LINHA16

Onde se lê: "…europeu – no os procedimentos…"
De ler-se: "…europeu – nos procedimentos…"

PÁGINA 135, LINHA 16

Onde se lê: "…estado presente s(n) com…"
Deve ler-se: "…estado presente X(n) com…"

PÁGINA 165, LINHA 17

Onde se lê: "…ver Figura 6.8…"
Deve ler-se: "…ver Figura 6.7…"

PÁGINA 168, LINHA 23

Onde se lê: "…deve ser ressaltado que…"
Deve ler-se: "…deve ser ressalvado que…"

PÁGINA 171, LINHA 5

Onde se lê: "A Figura 6.10 mostra…"
Deve ler-se: "A Figura 6.12 mostra…"