



Universidade de Aveiro Departamento de Electrónica, Telecomunicações e
2007 Informática
Departamento de Línguas e Culturas
Secção Autónoma de Ciências da Saúde

**Paula Maria Vaz
Martins**

**Ressonância Magnética no Estudo da Produção do
Português Europeu**



Universidade de Aveiro Departamento de Electrónica, Telecomunicações e
2007 Informática

Departamento de Línguas e Culturas

Secção Autónoma de Ciências da Saúde

**Paula Maria Vaz
Martins**

Ressonância Magnética no Estudo da Produção do Português Europeu

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Ciências da Fala e da Audição, realizada sob a orientação científica dos Doutores António Joaquim Silva Teixeira e Augusto Marques Ferreira da Silva, Professores Auxiliares do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro.

Dedicatória

Dedico este trabalho:

Aos meus filhos Gonçalo e Maria Ana

Ao Rui Miguel

Aos meus pais e familiares

Aos meus amigos

o júri

Presidente

Professor Doutor Francisco António Cardoso Vaz
Professor Catedrático da Universidade de Aveiro

Professor Doutor Luís Filipe Caseiro Alves
Professor Associado com Agregação da Faculdade de Medicina da Universidade de Coimbra

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

Professor Doutor Augusto Marques Ferreira da Silva
Professor Auxiliar da Universidade de Aveiro (Co-Orientador)

agradecimentos

Ao meu orientador Prof. Doutor António Teixeira pela motivação para a realização deste trabalho, pela sua cuidadosa orientação e pelas constantes palavras de incentivo e encorajamento desde o início deste Mestrado.

Ao Prof. Doutor Augusto Silva um agradecimento por ter acreditado na aquisição 3D e pelos seus ensinamentos e orientação na área do processamento de imagem.

Ao Prof. Doutor Filipe Caseiro Alves, Director do Serviço de Imagiologia dos Hospitais da Universidade de Coimbra, à Dr^a Cristina Brito Marques, Responsável pelo Sector de Ressonância Magnética e ao Técnico de Radiologia Emílio Leal, Técnico Coordenador do Sector, a permissão para adquirir as imagens apresentadas neste trabalho e sem as quais este não teria sido viável.

À Escola Superior de Saúde da Universidade de Aveiro pelo apoio concedido.

Ao Dr. João Cunha Pires pela sua disponibilidade e presença na sessão experimental e por todo o seu conhecimento na área das aplicações em Ressonância Magnética.

À minha colega e amiga Alda Pinto por se disponibilizar para efectuar a aquisição das imagens e pelas válidas sugestões para o estabelecimento do protocolo.

À Inês Carbone pelo seu empenho e trabalho na segmentação das imagens obtidas, e por todo o incentivo, companheirismo e amizade durante a realização deste trabalho.

À Marisa, Isabel e Catarina por todos os esclarecimentos, revisões e fundamentalmente pela sua amizade.

Aos informantes deste estudo, pela paciência que tiveram durante a aquisição das imagens.

A todos os meus familiares e amigos pelo incentivo e apoio incondicional, principalmente à minha mãe por me ter concedido as oportunidades que me permitiram chegar até aqui.

Ao Rui e ao Gonçalo, um agradecimento muito especial por aturarem a minha falta de paciência nos momentos de maior tensão e as minhas ausências mesmo quando presente.

Finalmente, para a minha querida Maria Ana um agradecimento do tamanho do mundo pelo seu sorriso e as minhas desculpas por todo o tempo que não lhe dispensei para elaboração deste trabalho.

palavras-chave

Ressonância Magnética, Produção de Fala, Português Europeu, Síntese Articulatória

resumo

A Ressonância Magnética (RM) é um método de imagem extremamente valioso na prática clínica e tem vindo a ser utilizado, nos últimos anos, em estudos de produção de fala. O facto de não utilizar radiações ionizantes, sendo considerado um método de aquisição relativamente inócuo, em conjunto com a sua capacidade multiplanar, boa resolução a nível de tecidos moles e possibilidade de modulação 3D, faz da RM um dos métodos mais promissores para obtenção de dados na área das Ciências da Fala. Existe já um vasto número de estudos com RM, para diversas línguas, mas este tipo de informação é ainda muito escassa para o Português Europeu.

Este estudo com RM teve como principais objectivos a obtenção de uma base de dados relativa aos diversos sons do PE, mediante a aquisição de imagens relativas a produções estáticas (2D e 3D) e produções dinâmicas recorrendo a aquisição de imagem em tempo real. Foi também objectivo deste trabalho a validação de um método de aquisição de imagem (aquisição 3D).

As imagens obtidas no corpus 2D permitiram obter as configurações do tracto vocal no plano sagital para grande parte dos sons do PE, incluindo todos os sons nasais, com uma resolução espacial e relação SNR que permitiu observar com nitidez a maior parte dos articuladores, extrair contornos e parâmetros articulatórios, e observar efeitos de coarticulação em oclusivas e fricativas. Da aquisição 3D, foram obtidas funções de área, parâmetros quantitativos como a abertura nasal, coeficiente de abertura do velo (CAV) e área da cavidade faríngea para vogais nasais, orais e consoantes nasais do PE.

As imagens em tempo real, obtidas com uma velocidade de 5 frames/s, permitiram obter informação preliminar relativa à dinâmica dos articuladores durante a produção de fala.

Os dados obtidos com este trabalho permitiram também o desenvolvimento de ferramentas de segmentação semi-automáticas fundamentais para a extracção de informação das imagens RM.

keywords

MRI, Speech production, European Portuguese, Articulatory synthesis

Abstract

Magnetic Resonance Imaging is a powerful diagnostic tool and has been used successfully to acquire information in speech production studies. Because it does not use ionising radiation, being considered a safe imaging technique, together with its multiplanar capability, good contrast resolution of soft tissues as well as the possibility of 3D modelling, makes MRI one of the most promising imaging methods in the area of speech research. There are several MRI speech production studies, for different languages, but there is not a systematic study for European Portuguese using MRI.

The main goal of this study was to acquire a MRI database relative to European Portuguese sounds. This database included images relative to static productions (2D and 3D) as well as images obtained from dynamic productions during real time acquisition.

2D images allowed to get the configurations of the vocal tract in the medio-sagittal plan for a vast part of the EP sounds, including all nasal sounds, with a SNR and spatial resolution that allowed (1) to observe with clearness most of the articulators, (2) to extract contours and articulatory parameters, and (3) to observe coarticulatory effects in stops and fricatives.

From 3D acquisition, area functions were obtained together with some quantitative parameters such as nasal opening, Velum Port Opening Quotient (VPOQ) and pharyngeal areas, both for nasal and oral vowels and nasal consonants of EP.

Real time images, obtained with a frame rate of five frames per second, allowed to get some (preliminary) information on the dynamics of the articulators (mainly tongue movements) during speech production.

The database, obtained with this work, allowed the development of semi-automatic tools of segmentation for the extraction of information from MR images.

Índice

Lista de Figuras	iii
Lista de Tabelas	xi
Lista de símbolos e abreviaturas	xiii
Capítulo 1	1
Introdução.....	1
1.1 Enquadramento e motivações.....	1
1.2 Objectivos.....	3
1.3 Estrutura da Dissertação.....	4
1.4 Relatórios e publicações efectuadas, no âmbito da dissertação.....	6
Capítulo 2	7
Sons do Português Europeu.....	7
2.1 Conceitos de Fonética e Fonologia.....	7
2.2 Aparelho fonador e produção de fala.....	10
2.3 Sons da fala, em particular do PE.....	22
2.4 Nasalidade.....	31
2.5 Coarticulação.....	32
2.6 Síntese de fala.....	36
2.7 Sumário.....	41
Capítulo 3	43
Técnicas de estudo em Fonética Articulatória.....	43
3.1 Técnicas de Imagem.....	43
3.2 Outras técnicas Laboratoriais.....	49
3.3 Sumário.....	56
Capítulo 4	57
Ressonância Magnética.....	57
4.1 História.....	57
4.2 Equipamento.....	59
4.3 Conceitos fundamentais em RM.....	62
4.4 Codificação da informação e formação de imagem.....	73
4.5. Parâmetros de aquisição e qualidade da imagem.....	82
4.6 Artefactos.....	85
4.7 Sequências de Pulso.....	88
4.8 Sequências e técnicas para aquisição rápida (Fast Imaging).....	94
4.9 Sumário.....	111
Capítulo 5	113
Ressonância Magnética em Estudos de Produção de Fala.....	113

5.1	Potencialidades e Limitações da RM no contexto da produção de fala	113
5.2	Estudos estáticos	115
5.3	Estudos dinâmicos e em tempo real	125
5.4	Sumário	134
Capítulo 6	135
	Metodologias	135
6.1	Corpus	135
6.2	Recolha de dados	139
6.3	Processamento de imagem	151
6.4	Sumário	159
Capítulo 7	161
	Resultados	161
7.1	Corpus 2D	161
7.2	Coarticulação	193
7.3	Corpus 3D	216
7.4	Corpus Tempo real	219
Capítulo 8	227
	Conclusões e Trabalho Futuro	227
8.1	Resumo do trabalho realizado	227
8.2	Principais resultados e conclusões	228
8.3	Limitações do estudo	229
8.4	Desenvolvimentos Futuros	230
Referências Bibliográficas	235
Referências electrónicas	247
Anexo I	249
	a) International Phonetic Alphabet	249
	b) SAMPA - Português Europeu	250
Anexo II	251

Lista de Figuras

Figura 2. 1 - Estruturas anatómicas do aparelho fonador. Fonte: Adaptado de Deller et al. (1993).	11
Figura 2. 2 - Laringe. Vista anterior (a) mediana (b) e posterior(c). Adaptado de: www.gbmc.org/voice.	12
Figura 2. 3 - Posição das pregas vocais durante: (a) respiração normal, em abdução; (b) fonação, em adução e (c) regulação da tensão nas pregas. Fonte: Seeley et al. (2005).	14
Figura 2. 4 - Cavidades supraglotais ou tracto vocal. Fonte: Adaptado de Pruthi (2006).	15
Figura 2. 5 - Tracto nasal: vista lateral (a) e medial (b). Adaptado de: http://webanatomy.net	16
Figura 2. 6 - Seios perinasais. Fonte: https://mywebpages.comcast.net	17
Figura 2. 7 - Articuladores e pontos de articulação. Adaptado de: www.ph.ludwigbsurg.de.htm ...	19
Figura 2. 8 - Divisão da língua, do ponto de vista articulatório. Imagem sagital obtida durante a produção de [6], no falante PAA.	20
Figura 2. 9 - Posição do velo na produção da vogal oral [i] (à esquerda) e da vogal nasal [i~](à direita), para o falante PAA.	22
Figura 2. 10 - Módulos do sintetizador desenvolvido na UA, utilizando síntese articulatória. Fonte: Teixeira et al. (2004a).	39
Figura 2. 11 - Tracto vocal baseado no modelo de Mermelstein. Fonte: Teixeira et al. (2005).	40
Figura 2. 12 - Grelha não regular utilizada para cálculo das áreas. Fonte: Teixeira (2000).	40
Figura 2. 13 - Modelação do tracto nasal. Fonte: Teixeira et al. (2005).	40
Figura 2. 14 - Definição de parâmetros (esquerda) e síntese (direita) de [6~tOniu]. Fonte: Teixeira et al.(2004a).	41
Figura 3. 1 - Exemplo de três <i>frames</i> de um filme cineradiográfico efectuado durante a produção de fala. Fonte: Munhall & Vatikiotis-Bateson &Tohkura (1995) - X-Ray Film Database for Speech Research.	44
Figura 3. 2 - Utilização de bases de dados cineradiográficas para extracção de parâmetros articulatórios. Fonte http://www.idiap.ch/pages/contenuTxt/Demos/demo6/xray.html	45
Figura 3. 3 - Dorso da língua obtido através de ultrasonografia e colocação da sonda sob o queixo do falante (em baixo). Fonte: http://www.speech.umaryland.edu	48
Figura 3. 4 - X-Ray Microbeam (gerador) à esquerda. Fonte: http://www.medsch.wisc.edu/ubeam/gen/generator.html e informação proveniente da base de dados (XRMB) à direita. Fonte: Araújo (2007)	49
Figura 3. 5 - Sistema AG 200 - <i>Helmet</i> onde se encontram as bobines transmissoras. Fonte: www.articulograph.com	51

Figura 3. 6 - Colocação esférica das seis bobines transmissoras utilizadas no sistema 3D AG 500 (a) e <i>Helmet</i> em forma de cubo no mesmo sistema, não sendo necessária a fixação da cabeça. Fonte: Zierdt et al. (2000).....	51
Figura 3. 7 - Imagem referente a uma frame duma palatografia (cima) e respectivo sinal acústico (baixo). Fonte: http://www.speech.umaryland.edu	53
Figura 3. 8 - Palato artificial e disposição dos eléctrodos no sistema Palatometer da Kay Corporation (a) e sistema Reading (b). Adaptado de http://www.linguistics.ucla.edu/faciliti/facilities/physiology/epg.html	54
Figura 3. 9 - EPG <i>frames</i> ou palatogramas: a) sistema Kay Palatometer (disponível em: www.speechmaryland e b) Reading EPG 3 retirado de Nguyen, Gibbon & Hardcastle (1996).....	55
Figura 4. 1 - Estados de energia principais para o átomo de Hidrogénio. O alinhamento paralelo a B_0 corresponde ao estado de menor energia. A diferença de energia entre os dois níveis (ΔE) é proporcional a B_0 . Fonte: Brown & Semelka (2003).....	64
Figura 4. 2 - Sob a acção de um campo magnético, os protões precessam em torno deste. O eixo precessional é paralelo ao campo magnético principal B_0 . A componente z não altera a sua direcção ou magnitude mas as componentes x e y, variam no tempo com uma frequência proporcional a B_0 . Fonte: Brown & Semelka (2003).....	64
Figura 4. 3 - Tempo de recuperação da magnetização longitudinal - Após aplicação de um pulso de excitação de 90° , a magnetização longitudinal (M_z) é recuperada exponencialmente sendo caracterizada pela constante temporal T1. Depois de um período de tempo igual a T1, 63% da magnetização longitudinal é recuperada. Fonte: Bushberg et al. (2002).....	66
Figura 4. 4 - Tempo de relaxamento T2. Fonte: Bushberg et al. (2002).....	67
Figura 4. 5 - Sinal recebido numa bobine, num campo uniforme para uma situação ideal. Fonte: Hashemi & Bradley (1997).....	68
Figura 4. 6 - Decaimento de Indução Livre (FID). Fonte: Hashemi & Bradley (1997).	68
Figura 4. 7 - Decaimento T2 e T2*. Fonte: Hashemi & Bradley (1997).....	69
Figura 4. 8 - TR adequado para obtenção de ponderação T1. Fonte: Hashemi & Bradley (1997).72	
Figura 4. 9 - Coordenadas espaciais para cortes com diferentes orientações. Fonte: Liang & Lauterbur (2000).	74
Figura 4. 10 - Para uma determinada gama de frequências contidas num pulso de RF, a espessura de corte é determinada pela amplitude do gradiente de selecção de corte. Fonte: Silva (2003).	74
Figura 4. 11 - Função <i>sinc</i> com um pulso espectral aproximadamente rectangular. Fonte: Silva (2003).	75
Figura 4. 12 - Codificação em frequência, para um objecto unidimensional com uma densidade de <i>spins</i> $\rho(x)$. Fonte: Silva (2003).....	76

Figura 4. 13 - Codificação de fase, para um objecto unidimensional com uma densidade de spins $\rho(y)$. Fonte: Silva (2003).....	78
Figura 4. 14 - Preenchimento do espaço K, seguindo uma trajectória linear. Fonte: Silva (2003).	79
Figura 4. 15 - Espaço K e espaço da imagem. Fonte: Adaptado de Siemens (2003).....	81
Figura 4. 16 - Relação entre o espaço das medidas, o espaço K e a imagem. Fonte: Silva (2003).....	82
Figura 4. 17 - Artefacto de <i>aliasing</i> (2D FT) na direcção da fase onde se observa o nariz projectado posteriormente.....	86
Figura 4. 18 - Artefacto de “enrolamento” numa aquisição 3D - observa-se sobreposição dos últimos cortes do volume obtido, sobre os primeiros.....	87
Figura 4. 19 - Tempo de Eco. Fonte: Siemens (2003).....	89
Figura 4. 20 - Diagrama de pulso de uma sequência SE. Fonte: Siemens (2003).....	89
Figura 4. 21 - Diagrama de pulso referente a uma sequência IR, neste caso a sequência FLAIR ou <i>dark-fluid</i> . Fonte: Siemens (2003).....	91
Figura 4. 22 - Diagrama de pulso de uma sequência EG. Fonte: Brown & Semelka(1999).	92
Figura 4. 23 - Flip angle parcial (α). Fonte: Siemens (2003).....	93
Figura 4. 24 - Diagrama de pulso de uma sequência TSE. Fonte: Siemens (2003).....	95
Figura 4. 25 - FID, SE e STE . Fonte: Lafuente (2000).....	97
Figura 4. 26 - Sinal composto pela sobreposição do FID, SE’s e ecos estimulados. Fonte: Lafuente (2000).....	98
Figura 4. 27 - Sequência EG <i>Spoiled</i> - FLASH. Fonte: Siemens (2003).....	98
Figura 4. 28 - Diagrama de pulso de uma sequência b-SSFP (TrueFISP). Fonte: Scheffler & Lehnhardt (2003).....	100
Figura 4. 29 - Diagrama de pulso de uma sequência DESS. Fonte: Siemens (2004).....	102
Figura 4. 30 - Diagrama de pulso de uma sequência com preparação da magnetização: Turbo FLASH Fonte: Siemens (2004).....	102
Figura 4. 31 - Esquema de uma sequência EPI: em cima SE-EPI em baixo EG-EPI. Fonte: Bitar et al. (2006).....	104
Figura 4. 32 - Técnica TSE Zoom (Philips). Fonte: Elster & Burdette(2001).....	107
Figura 4. 33 - Efeito da amostragem no FOV. (a) Imagem axial convencional e (b) esquema de amostragem do espaço K para uma aquisição convencional e com iPAT, e (c) artefacto derivado da sub-amostragem. Fonte: Heidemann et al. (2003).	109
Figura 4. 34 - Esquema simplificado de aquisição paralela utilizando a técnica SENSE. Fonte: Pires (2005).	110
Figura 4. 35 - Esquema simplificado de aquisição paralela utilizando a técnica GRAPPA Fonte: Pires (2005).	110

Figura 5. 1 - Marcação das ranges no plano axial, coronal e oblíquo. Fonte: Demolin, Metens & Soquet (1996).	119
Figura 5. 2 - Contorno da língua em: corte axial (em cima à esq), oblíquo (em cima à dta), corte coronal em (baixo à esq.) e range de cortes projectada numa imagem sagital durante a produção de /ipi/ (em baixo à dta). Fonte: Engwall (2000a).	122
Figura 5. 3 - Plano médio sagital para a vogal /i/. À esquerda, grelha inicial para determinação da linha central do tracto vocal (vocal tube centerline) e à direita, grelha final para estimar as áreas transversais do tracto vocal. Fonte: Kröger et al.(2000)	123
Figura 5. 4 - Produção da palavra [tyga] pelo falante 2. Fonte: Demolin et al (1997).	127
Figura 5. 5 - Aquisição em tempo real da repetição da sequência /ieaou/ produzida pelo falante do sexo masculino. As 16 imagens representam um total de 3,2s, com uma imagem a cada 200 ms. Fonte: Demolin, Metens & Soquet(2000).	128
Figura 5. 6 - Procedimento para o método de amostragem sincronizada (SSM). Fonte: Takemoto et al. (2003).	131
Figura 5. 7 - Processo de reorganização dos dados para obter 56 volumes. Fonte: Takemoto et al. (2003).	132
Figura 5. 8 - (a) Sequência de pulso utilizada na aquisição de imagens em tempo real e (b) Esquema de aquisição temporal. Fonte: Narayanan et al.(2004)	133
Figura 6. 1 - Representação da semente numa imagem de RM no plano sagital (falante PAA - corpus 2D).	153
Figura 6. 2 - Noção de vizinhos. Semente a amarelo e vizinhos representados a vermelho.	154
Figura 6. 3 - Contorno criado pela técnica <i>Region Growing</i> .	155
Figura 6. 4 - Obtenção das distâncias médio sagitais no corpus 2D: esquerda - contorno (amarelo), linha central (vermelho) e grelha adaptativa (verde); direita - gráfico das distâncias médio sagitais.	155
Figura 6. 5 – Marcação de pontos referentes a UL, LL, TT, TD, TR e JH, para o falante PAA durante a produção do fone [6].	167
Figura 6. 6 - Exemplo de utilização do algoritmo <i>live wire</i> : a verde - linha já escolhida pelo utilizador e a vermelho - linha proposta pelo algoritmo.	157
Figura 6. 7 - Esquerda: Imagem no plano sagital, respectiva segmentação (amarelo), linha central (azul) e planos perpendiculares (vermelho); Direita: Resultado da segmentação.	158
Figura 7. 1 - Vogais orais anteriores do PE [i] em cima, [e] ao centro e [E] em baixo, produzidas pelo falante PAA (imagem no plano sagital à esquerda, com o respectivo contorno à direita).	163
Figura 7. 2 - Vogais orais médias do PE [@] em cima, [6] ao centro e [a] em baixo, produzidas pelo falante PAA (imagem no plano sagital à esquerda e com o respectivo contorno à direita).	164
Figura 7. 3 - Vogais orais posteriores do PE [u] em cima, [o] ao centro e [O] em baixo, produzidas pelo falante PAA (imagem no plano sagital com o respectivo contorno à direita).	165

Figura 7. 4 - Configurações do tracto para as vogais anteriores [i e E], no plano médio-sagital produzidas pelo falante PAA, tendo como referência as palavras [pipu], pêca [pek6] e leva [IEv6] (em cima). Sobreposição dos contornos referentes às configurações referidas (ao centro) e em baixo os IP relativos aos contornos.	166
Figura 7. 5 - Configurações do tracto para as vogais médias [@ 6 a], no plano médio sagital produzidas pelo falante PAA, tendo como referência as palavras devi [d@vi], cada [k6d6] e pato [patu] (em cima). Sobreposição dos contornos referentes às mesmas configurações (ao centro) e em baixo, os IP para os contornos referidos.	167
Figura 7. 6 - Configurações do tracto para as vogais posteriores ou velares [u o O], no plano médio sagital produzidas pelo falante PAA, tendo como referência as palavras buda [bud6], tôpo [topu] e pote [pOt@] (em cima). Sobreposição dos contornos referentes às mesmas configurações (ao centro) e em baixo os IP para os contornos referidos.....	169
Figura 7. 7 - Sobreposição dos contornos obtidos para todas as vogais orais do PE, para o falante	170
Figura 7. 8 – Parâmetros articulatórios obtidos para todas as vogais do PE. Do canto superior esquerdo: ponto mais alto do dorso da língua (TD), ponto mais recuado da língua, a nível do espaço intervertebral de C3-C4 (TR), ponta da língua (TT), altura da mandíbula (JH) e lábio inferior (LL) e superior (UL).	171
Figura 7. 9 - Ponto mais alto do dorso da língua durante a produção das vogais cardinais do PE: [i] (vermelho), [u] (azul) e [a] (verde).	172
Figura 7. 10 - Sobreposição dos contornos de [i] (verde), [u] (vermelho) e [@] (azul).	173
Figura 7. 11 - Vogais nasais do PE: à esquerda a imagem no plano médio sagital e à direita a imagem com o respectivo contorno, para o falante PAA. De cima para baixo, [i~],[e~],[6~], [o~] e [u~].	175
Figura 7. 12 - Sobreposição dos contornos referentes a todas as vogais nasais do PE, para o falante PAA.....	176
Figura 7. 13 - Sobreposição dos contornos referentes às configurações do tracto para as vogais orais [6] e [a] e da vogal nasal [6~], produzidas pelo falante PAA.	177
Figura 7. 14 - Sobreposição dos contornos referentes às configurações do tracto das vogais orais [@ 6 a] e da vogal nasal [6~], produzidas pelo falante PAA.	178
Figura 7. 15 - Sobreposição dos contornos referentes às configurações do tracto para as vogais orais [e E] e da vogal nasal [e~], produzidas pelo falante PAA.....	178
Figura 7. 16 - Sobreposição dos contornos referentes às configurações do tracto para a vogal oral [i] e da vogal nasal [i~], produzidas pelo falante PAA.....	179
Figura 7. 17 - Sobreposição dos contornos referentes às configurações do tracto para as vogais orais [o] e [O] e para a vogal nasal [o~], produzidas pelo falante PAA.....	180
Figura 7. 18 - Sobreposição dos contornos referentes às configurações do tracto para a vogal oral [u] e para a vogal nasal [u~], produzidas pelo falante PAA.....	180

Figura 7. 19 - Imagens correspondentes à produção das oclusivas surdas [p] em cima, [t] ao meio e [k] em baixo, produzidas pelo falante PAA, no contexto V_1CV_2 em que $V_1= V_2 = [a]$	182
Figura 7. 20- Imagens e respectivos contornos correspondentes à produção das oclusivas sonoras [b] em cima, [d] ao meio e [g] em baixo, produzidas pelo falante PAA, no contexto V_1CV_2 em que $V_1= V_2 = [a]$	183
Figura 7. 21 - Sobreposição dos contornos obtidos para as oclusivas bilabiais [p] e [b] à esquerda, para as dentais [t] e [d] ao centro e para as velares [k] e [g] à direita, produzidas pelo falante PAA, em contexto V_1CV_2 em que $V_1= V_2 = [a]$	184
Figura 7. 22 - Imagens e respectivos contornos correspondentes à produção das consoantes nasais: [m] em cima, [n] ao centro e [ŋ] em baixo, produzidas pelo falante PAA, tendo como referências as palavras <i>cama</i> [k6m6], <i>cana</i> [k6n6] e <i>canha</i> [k6J6].	185
Figura 7. 23 - Comparação das oclusivas orais e nasais pelo ponto de articulação. À esquerda sobreposição dos contornos para as bilabiais [m], [p] e [b], à direita para [n], [t], [d].	186
Figura 7. 24 - Imagens e respectivos contornos, correspondentes à produção das fricativas surdas em início de palavra: [f] em cima, [s] ao centro e [ʃ] em baixo produzidas pelo falante PAA, tendo como referência as palavras <i>fala</i> [fal6], <i>saca</i> [sak6] e <i>chá</i> [Sa] respectivamente.	187
Figura 7. 25 – Sobreposição dos contornos, relativos ao [s] (a azul) e [ʃ] (a vermelho).....	188
Figura 7. 26 - Imagens e respectivos contornos, correspondentes à produção das fricativas sonoras em início de palavra: [v] em cima, [z] ao centro e [ʒ] em baixo produzidas pelo falante PAA, tendo como referência as palavras <i>vaca</i> [vak6], <i>zarpa</i> [zarp6] e <i>jacto</i> [Zatu], respectivamente.....	189
Figura 7. 27 - Comparação dos contornos relativos a pares de fricativas surda /sonora em função do ponto de articulação ([f] e [v] à esquerda, [s] e [z] ao centro e [ʃ] e [ʒ] à direita.	190
Figura 7. 28 - Imagens e respectivos contornos, correspondentes à produção das laterais, [l] em cima, [l~] ao centro e [L] em baixo, produzidas pelo falante PAA, tendo como referência as palavras <i>laço</i> [lasu], <i>mal</i> [mal~] e <i>falha</i> [faL6], respectivamente.	191
Figura 7. 29 - Sobreposição dos contornos, referentes a [l] e [l~] à esquerda e à direita de [l] com [L].	192
Figura 7. 30 - Sobreposição dos contornos obtidos para as oclusivas surdas [p] em cima, [t] ao centro e [k] em baixo produzidas em contexto simétrico com as vogais [a],[i] e [u], para o falante PAA.	194
Figura 7. 31 - Sobreposição dos contornos obtidos para as oclusivas surdas [b] em cima, [d] ao centro e [g] em baixo produzidas em contexto simétrico com as vogais [a],[i] e [u], para o falante PAA.	195
Figura 7. 32 - Sobreposição dos contornos obtidos para as fricativas surdas [f] em cima, [s] ao centro e [ʃ] em baixo, produzidas em contexto simétrico com as vogais [a], [i] e [u], para o falante PAA.	197
Figura 7. 33 - Sobreposição dos contornos obtidos para as fricativas sonoras [v] em cima, [z] ao centro e [ʒ] em baixo, produzidas em contexto simétrico com as vogais [a], [i] e [u].	199

Figura 7. 34 - Distâncias médio-sagitais obtidas, a partir da aplicação de grelha adaptativa sobre as imagens, referentes às vogais [a] em cima, [i] ao centro e [u] em baixo.....	203
Figura 7. 35 - Funções de área obtidas para as vogais [i] em cima, [a] ao centro e [u] em baixo através da aquisição clássica (áreas transversais relativas ao tracto vocal obtidas a partir de cortes efectuados nos planos coronal, oblíquo e transversal).....	205
Figura 7. 36 - Funções de área obtidas para as vogais [i] em cima, [a] ao centro e [u] em baixo, através da aquisição volumétrica (método proposto no nosso estudo).....	206
Figura 7. 37 - Comparação das áreas obtidas através dos dois métodos para as vogais cardinais do PE, [a] à direita, [i] ao centro e [u] à esquerda, por aquisição volumétrica (a vermelho) e clássica (a preto).....	207
Figura 7. 38 - Funções de área do tracto vocal para três contextos do [l] do PE (em cima):à esquerda em início de sílaba e palavra [lasu], ao centro em final de sílaba [mal~], e à direita em início de sílaba, mas intervocálico [pal6]. Em baixo, a sobreposição das três funções de área...	209
Figura 7. 39 – Função de área para a lateral [L].....	210
Figura 7. 40 – Funções de área para as fricativas surdas [f] à esquerda, [s] ao centro e [S] à direita obtidas em contexto com as 3 vogais [a], [i] e [u], no corpus 3D.	210
Figura 7. 41 - Corte perpendicular ao tracto vocal, na zona de acoplamento dos tractos oral e nasal, para as vogais nasais do PE (falante PAA).	213
Figura 7. 42 - Corte perpendicular ao tracto vocal, na zona de acoplamento dos tractos oral e nasal, para a vogais orais do PE adquiridas no corpus 3D (falante PAA).	213
Figura 7. 43 - Corte perpendicular ao tracto vocal, na zona de acoplamento dos tractos oral e nasal, para as consoantes [m] à esquerda, [n] ao centro e [J] à direita.....	214
Figura 7. 44 - Valores médios relativos às áreas nasais e orais e respectivo VPOQ ou CAV (azul) para vogais nasais, orais e consoantes nasais.	215
Figura 7. 45 - Boxplots que representam a mediana (traço a vermelho), desvio-padrão e médias (ponto) relativas ao VPOQ (CAV) para vogais nasais, orais e consoantes nasais.	215
Figura 7. 46 – Visualização 3D, obtida com o ITK Snap, a partir da aquisição em Volume (3D-VIBE) para a vogal oral [6] à esquerda e vogal nasal [6~] à direita, para o falante PAA.....	219
Figura 7. 47 - Vinte e sete frames identificadas, por comparação, referentes à produção da sequencia de vogais [a E i O u] e que correspondem a cerca de 5,4 s de produção.....	221
Figura 7. 48 - Marcação de três pontos na língua: (1) ponta da língua, (2) ponto mais elevado do dorso e (3) posição da raiz a nível do bordo inferior do corpo de C2. <i>Frame</i> referente à produção da vogal [u] durante a produção da palavra [katu].....	222
Figura 7. 49 - <i>Frames</i> referentes à produção de [k a t u], à esquerda [k], centro esquerda [a], centro direita [t] e à direita [u] (frames não sequenciais).	223
Figura 7. 50 - Variação espacial, dos três pontos marcados ao nível da língua durante a produção da palavra [katu].....	223
Figura 7. 51 - Pseudo - variação da posição da língua no tempo durante a produção de [k6~tu] (esquerda) e [katu] (direita).	224

Figura 7. 52 - Sobreposição dos contornos da língua para o [k] (cima à esquerda), [6~] e [a] (cima à direita), [t] (baixo à esquerda) e [u] (baixo à direita) na produção referente a uma frame de [k6~tu] (azul) e [katu] (vermelho).225

Lista de Tabelas

Tabela 2. 1 - Classificação das vogais orais do PE. Fonte: Moutinho (2000).....	24
Tabela 2. 2 - Classificação das vogais nasais do PE. Fonte: Mateus et al. (2005).	25
Tabela 2. 3 - Ponto de articulação, articuladores, movimentos dos articuladores na produção das consoantes do PE padrão. Adaptado de Moutinho (2000) e Mateus et al. (2005).....	28
Tabela 2. 4 - Coarticulação (Adaptado de Farnetani, 1999).	33
Tabela 4. 1 - Valores referentes ao tempo de relaxamento T1 e T2 (em ms) para Intensidade de campo de 0,5 T e 1,5 T. Fonte: Bushberg et al. (2002).	67
Tabela 4. 2 - Gradientes seleccionados para cada um dos planos ortogonais. Fonte: Hashemi & Bradley (1997) e Westbrook & Kaut (1998).....	73
Tabela 6. 1 - Descrição resumida do corpus 2D Sagital, para o falante PAA.....	136
Tabela 6. 2 - Descrição resumida do corpus 3D, para o falante PAA.	136
Tabela 6. 3 - Corpus obtido em tempo real, para a falante MTC.	137
Tabela 6. 4 - Fones adquiridos (estáticos) 2D sagital para identificação das imagens em tempo real obtida na falante MTC.	138
Tabela 6. 5 - Parâmetros de aquisição da sequência real time TrueFISP utilizada para obtenção do corpus tempo real.....	144
Tabela 6. 6 - Parâmetros de aquisição da sequência TSE para o corpus 2D.	145
Tabela 6. 7 - Parâmetros de aquisição da sequência 3DFLASH VIBE (axial) para o corpus 3D.	146
Tabela 6. 8 - Parâmetros de aquisição da sequência DP TSE para obtenção do corpus 3D pelo método clássico.	147
Tabela 6. 9 - Parâmetros de aquisição da sequência 3D MPRAGE utilizada para obtenção do tracto oral e nasal em repouso, do falante PAA.	148
Tabela 6. 10 - Parâmetros de aquisição da sequência 3D TSE, ponderada em T2 , para obtenção das arcadas dentárias do falante PAA.	149
Tabela 7. 1 - Áreas da abertura oral e nasal para vogais nasais, vogais orais e consoantes nasais.	212
Tabela 7. 2 - Áreas (médias) da faringe (orofaringe) para as vogais nasais e orais e consoantes nasais do PE.....	217

Lista de Símbolos e Abreviaturas

α	Flip angle ou ângulo de rotação
α_E	Ernst angle
AFI	Alfabeto Fonético Internacional
AutoSMASH	Auto Simultaneous Acquisition of Spatial Harmonics
AV	Altura do Velo
χ	Susceptibilidade magnética
C6	Sexta vértebra cervical
CASY	Configurable Articulatory Synthesizer
CAV	Coeficiente de Abertura do Velo
CE-FAST	Contrast Enhancement - FAST
CISS	Constructive Interference Steady State
CPRM	Colangiopancreatografia por RM
CSE	Conventional Spin-Echo (ou spin-eco convencional)
DAC	Degree of Articulatory Constrains (grau de limitação ou de impedimento articulatório)
dB/dt	Campo magnético variável no tempo
DE	Direita/Esquerda (coarticulação antecipatória)
DESS	Dual Echo Steady State
DP	Densidade de Protões
EBCT	Electron Beam Computed Tomography (ou Tomografia Computorizada por Feixe de Electrões)
ED	Esquerda/direita (coarticulação <i>carryover</i>)
EG	Eco de Gradiente
EMA	Electromagnetic Articulography (ou Articulografia Electromagnética)
EMMA	Electro Magnetic Mid-Sagittal Articulography
EMR	Experimental Magnetic Resonance
EPG	Electropalatografia
EPI	Echo Planar Imaging (ou imagem eco-planar)
ES	Eco de Spin
ETL	Echo Train Length
F0	Frequência Fundamental
FA	Flip Angle
FFE	Fast Field Echo
FID	Free Induction Decay (ou decaimento de indução livre)
FISP	Fast Imaging with Steady Precession
FLAIR	Fluid Attenuated Inversion Recovery
FLASH	Fast Low Angle SHot
VIBE	Volume Interpolated Breath-hold Exam
FOV	Field of View (ou campo de visão)
fRM	Ressonância Magnética Funcional
FSE	Fast Spin Echo (=TSE)
GRAPPA	Generalized Autocalibrating Partially Parallel Acquisition
GRASE	Gradient and Spin Echo
GRASS	Gradient Recalled Acquisition in Steady State
G_s	Gradiente de selecção de corte
HALTS	Head and Transducer Support System
HASTE	Half-Fourier Single Shot Turbo spin echo
Hz	Hertz
ICP	Institute de la Communication Parléé
IEETA	Instituto de Electrónica e Telemática da Universidade de Aveiro
IP	Índice de Pratt
IPA	International Phonetic Alphabet
IR	Inversão da Recuperação
IRM	Imagem por Ressonância Magnética

IS	Intensidade de Sinal
LoLo	Local Lock
LPC	Linear Predictive Coding
M	Vector de magnetização
MP-GRE	Magnetization Prepared Gradiente Echo
MP-RAGE	Magnetization Prepared - Rapid Gradient Echo
MTC	Informante feminino
M _{xy}	Componente de magnetização transversal
M _z	Magnetização longitudinal
NEX	Number of Excitations (Número de Aquisições)
N _p	Número de passos de codificação de fase
PAA	Informante masculino
PAT	Parallel Acquisition Techniques
PE	Português Europeu
ppm	partes por milhão
PSIF	Reversed FISP
RARE	Rapid Acquisition Relaxation Enhancement
RecFOV	FOV rectangular
RM	Ressonância Magnética
RMN	Ressonância Magnética Nuclear
S/R ou SNR	Relação Sinal/Ruído
SAADS	Sector anterior da arcada dentária superior
SAM	Speech Assessment Methods
SAMPA	SAM Phonetic Alphabet
SAPWindows	Sintetizador Articulatorio do Português para plataforma Windows
SAR	Specific Absorption Rate
SE	Spin Eco
SENSE	Sensitivity Encoding
SFS	Speech Filing System
SMASH	Simultaneous Acquisition of Spacial Harmonics
SPACE RIP	Sensitivity Profiles from an Array Coil for Encoding and Reconstruction in Paralell
SPE	Sound Pattern of English
SSFP	Steady State Free Precession (estado estacionário de precessão livre ou estado de equilíbrio estável)
SSM	Sinchronized Sampling Method
STE	Stimulated Echo (ou eco estimulado)
STIR	Short TI Inversion Recovery
T	Tesla
T1	Tempo de relaxação longitudinal
T2	Tempo de relaxação transversal
TA	Tempo de aquisição
TAC	Tomografia Axial Computorizada
TC	Tomografia Computorizada
TE	Tempo de Eco
TF	Transformada de Fourier
TI	Tempo de Inversão
TM	Transferência de Magnetização
TR	Tempo de Repetição
TR	Tongue Root (no capítulo sete)
T _s	Sampling Time (Tempo de amostragem)
TSE	Turbo Spin Eco (= FSE)
TTS	Text-to-Speech
VC	Vogal - Consoante
VCV	Vogal-Consoante-Vogal
VPOQ	Velum Port Opening Quocient (ou coeficiente de abertura do velo)
Z	Eixo longitudinal

Capítulo 1

Introdução

1.1 Enquadramento e motivações

Uma vez que a fala é o principal meio de comunicação entre os seres humanos, o seu estudo, é objecto de interesse de diferentes áreas do conhecimento (Fant, 2001; Mohammad, 1999; Viana, 2001). Assim, as ciências da fala envolvem, nos dias de hoje, a cooperação entre áreas tão diversas como a Fonética, a Linguística, a Psicologia, a Electrónica, a Imagiologia, a Computação, a Física, e a Medicina entre muitas outras (Fant, 2001; Viana, 2001), sendo frequente a criação de equipas multidisciplinares que se ocupam do estudo da fala, nomeadamente a nível da produção, síntese e reconhecimento.

O recurso a técnicas imagiológicas, a área de formação da autora deste trabalho, está bem presente no campo das ciências da fala. Imagens radiográficas e cineradiográficas estiveram na base de uma das teorias mais importantes nesta área - a teoria acústica de produção da fala - proposta em 1960 por Fant, e em muitos outros estudos de produção de fala (por exemplo, Moll, 1960; Perkell, 1969; Stevens & Öhman, 1963). Mais recentemente, têm sido utilizadas a Tomografia Computorizada, a Ultrasonografia e a Ressonância Magnética (RM).

A importância da RM nos estudos de produção de fala tem vindo a crescer nas últimas duas décadas, uma vez que torna possível a observação do tracto vocal em toda a sua extensão, sem recurso a radiações ionizantes, o que, em conjunto com a sua capacidade multiplanar, potencial para modelação tridimensional e excelente resolução de contraste a nível dos tecidos moles, representam as principais vantagens deste método de imagem. O ruído produzido pelas bobines de gradientes, o facto de a aquisição ser efectuada com o falante em supinação e a necessidade de, nos estudos estáticos, o falante suste as articulações durante o período de aquisição, têm sido apontadas como as limitações mais evidentes da RM na avaliação do tracto vocal. A resolução temporal, não é, ainda, a desejável para estudar algumas das características dos articuladores, no entanto, permite já a obtenção de alguma informação dinâmica relevante para o estudo da fala. Na realidade, os avanços tecnológicos dos últimos anos (equipamentos com gradientes mais potentes, desenvolvimento de novas sequências, melhores bobines, etc.) têm permitido superar algumas das limitações da RM e aproximá-la da *performance* necessária para o estudo do tracto vocal. Apesar das limitações apontadas, a informação obtida através de RM, em muito tem contribuído para um maior conhecimento do tracto vocal e dos mecanismos envolvidos

na produção de fala, devendo ser considerada informação válida e relevante na área (Engwall, 2003b).

A RM tem sido utilizada em variados estudos, permitindo a obtenção de informação relativa ao tracto vocal para diversas línguas: Inglês Americano (Alwan, Narayanan, & Haker, 1997; Narayanan, Alwan, & Haker, 1995a, 1995b, 1997; Narayanan, Nayak, Lee, & Byrd, 2004; Stone & Lundberg, 1996; Story, 1995; Story, Titze, & Hoffman, 1996), Inglês Britânico (Baer, Gore, Gracco, & Nye, 1991; Greenwood, Goodyear, & Martin, 1992; Mohammad, 1999; Shadle, Mohammad, Carter, & Jackson, 1999), Francês (Badin, Bailly, Raybaudi, & Segebarth, 1998; Badin et al., 2000; Demolin, Metens, & Soquet, 1996, 2000; Soquet, Lecuit, Metens, & Demolin, 2002), Norueguês (Foldvick, Kristiansen, & Kvoerness, 1995), Sueco (Engwall & Badin, 1999; Engwall, 2000 a, 2000 b, 2001, 2003a, 2003b; Ericsson, 2005), Japonês (Dang & Honda, 2002; Dang, Shadle, Kawanishi, Honda, & Suzuki, 1998; Fujita & Honda, 2005; Takemoto, Honda, Masaki, Shimada, & Fujimoto, 2003; Takemoto, Kitamura, Nishimoto, & Honda, 2004), Alemão (Kröger, Winkler, Mooshammer, & Pompino-Marschall, 2000; Hoole et al., 2000; Mathiak et al., 2000), Coreano (Kim, 2004; Yang, 1999), Tamil (Narayanan, Kaun, Byrd, Ladefoged, & Alwan, 1996; Narayanan, Nayak, Lee, & Byrd, 2004), entre outros. Tanto quanto seja do nosso conhecimento não existe nenhum estudo sistemático utilizando RM, para o estudo da produção do Português Europeu (PE), existindo já algum trabalho realizado para as vogais do Português Brasileiro (Gregio, 2006).

O Português Europeu, apesar de apresentar características semelhantes a outras línguas de origem Românica, tem algumas características próprias, como por exemplo, um número elevado de sons nasais (vogais nasais, ditongos e tritongos nasais, consoantes nasais) sobre os quais, existem poucos dados baseados em técnicas laboratoriais.

A escassez ou mesmo inexistência de dados relativos à produção do Português Europeu, obtidos através de RM constitui a principal fonte de motivação para este trabalho, uma vez que é uma área onde trabalhamos nos últimos anos, no Sector de Ressonância Magnética dos Hospitais da Universidade de Coimbra. A oportunidade de adquirir e processar imagens RM aplicando conhecimentos adquiridos durante a componente curricular do Mestrado em Ciências da Fala e Audição, nomeadamente ao nível da produção e processamento de fala, torna-se muito aliciante uma vez que coloca problemas e levanta desafios significativamente diferentes dos que surgem diariamente na prática clínica.

A perspectiva de integrar uma equipa multidisciplinar com profissionais de diferentes áreas do conhecimento (Engenharia Electrónica, Imagem Médica, Linguística, Fonética, Matemática, etc.) decorrente da relação deste estudo com uma das tarefas do projecto FCT HERON (Referência POSI/PLP/57680/2004), de que é investigador principal o orientador desta dissertação, é também extremamente enriquecedora.

Uma área, em que a obtenção de dados de produção é fundamental, está relacionada com os sistemas de síntese da fala. Uma das técnicas de síntese mais promissoras é a síntese articulatória, que se baseia numa descrição pormenorizada da posição e movimento das estruturas anatómicas envolvidas na produção de fala (Teixeira, 2000; Viana, 2001), sendo necessária a obtenção de parâmetros anatómicos (2D e 3D) e fisiológicos que podem ser adquiridos com recurso às técnicas de imagem. A existência de um sintetizador baseado em síntese articulatória, na Universidade de Aveiro, e o interesse nesta área de investigação, justifica a necessidade de obtenção de informação através de RM, relativa à produção do Português Europeu.

Este trabalho, ainda que não directamente relacionado com o processo de síntese, poderá dar um modesto contributo para a continuidade do trabalho já iniciado por Teixeira e colaboradores na síntese do PE, em particular dos sons nasais (Teixeira, 2000; Teixeira & Vaz, 2001; Teixeira et al., 2004a; Teixeira et al., 2005). Este facto, é também motivador, uma vez que a síntese de fala é muito mais que uma aplicação importante da investigação na área das ciências da fala ou um “capricho” da Engenharia. É extremamente importante no conhecimento da Fonética, pode ser utilizada no ensino de línguas estrangeiras, representa uma mais valia na utilização de sistemas de informação por indivíduos com limitações físicas (visuais, auditivas, vocais e outras), tem aplicação em sistemas de segurança e em aplicações militares, entre muitas outras (Simões, 1999; Teixeira, 2003c).

1.2 Objectivos

1.2.1 Objectivo geral

Obtenção de informação anatómica e fisiológica, relativa ao processo de produção do Português, através da utilização de imagens de RM, em falantes do PE, com o intuito de dar suporte ao desenvolvimento de modelos articulatórios fundamentais ao aperfeiçoamento e evolução do sintetizador de voz (SAPWindows - Sintetizador Articulatório do Português para plataforma Windows), desenvolvido na Universidade de Aveiro.

1.2.2 Objectivos específicos

1. Obter imagens RM 2D estáticas do tracto vocal, no plano sagital, durante a produção de vários fones do PE, com o objectivo de obter o traçado dos perfis médio-sagitais, distâncias médio-sagitais e alguns parâmetros articulatórios;
2. Obter informação tridimensional, a partir de cortes efectuados nos diferentes planos (axial, sagital, coronal e oblíquos) ou mediante aquisição volumétrica (3D), de forma a permitir a

determinação das funções de área do tracto vocal e nasal e de parâmetros articulatórios, durante a produção de fones do PE;

3. Adquirir imagens RM, de produções dinâmicas, que permitam observar a variação temporal da abertura e fecho do velo, de forma a melhorar a síntese de vogais nasais do PE;
4. Obter imagens RM em tempo real, durante a produção de sequências e de palavras contendo vogais e consoantes nasais, que permitam estudar a dinâmica do tracto vocal nos sons nasais e estudar fenómenos de coarticulação em tempo real;
5. Adquirir informação 3D das arcadas dentárias e tracto nasal dos falantes intervenientes, de forma a permitir uma modelação mais fidedigna do tracto vocal;
6. Seleccionar e otimizar os protocolos de aquisição de imagem conducentes à obtenção de imagens com uma qualidade que permita atingir cada um dos objectivos atrás definidos. Em particular, é nosso objectivo recorrer a um método de aquisição de informação 3D diferente do que tem sido descrito em outros estudos, nesta área;
7. Contribuir para o desenvolvimento de técnicas de segmentação, em particular, técnicas semi-automáticas e automáticas que venham facilitar a extracção de informação obtida por RM, muito em particular, de imagens obtidas em tempo real.

1.3 Estrutura da Dissertação

A presente dissertação encontra-se dividida em oito capítulos e dois anexos. Neste primeiro capítulo enquadra-se o tema, apresentam-se as motivações subjacentes à realização deste trabalho, definem-se os objectivos a atingir, faz-se uma descrição da estrutura da dissertação. O capítulo termina com apresentação de alguns trabalhos realizados no âmbito do Mestrado e relacionados com a presente dissertação.

No capítulo 2, definem-se alguns conceitos relacionados com a Fonética e Fonologia do Português Europeu e apresenta-se uma breve revisão acerca de conceitos anatómicos e fisiológicos envolvidos na produção da fala. Caracterizam-se, do ponto de vista articulatório, os sons do PE e aborda-se, ainda que sucintamente, tópicos como a coarticulação e nasalização. De forma sintética apresentam-se, no final do capítulo, os principais sistemas de síntese da fala.

Uma vez que, para estudar de forma completa o aparelho produtor de fala e os processos envolvidos na sua produção, é necessária informação proveniente de diversas técnicas de estudo, no capítulo 3, faz-se uma revisão acerca das técnicas mais comuns utilizadas em fonética articulatória. Incluem-se as técnicas de imagem à excepção da Ressonância Magnética que será abordada nos capítulos 4 e 5 da dissertação. Procura-se, para cada uma delas, explicar resumidamente o seu funcionamento e apresentar as vantagens e limitações.

No capítulo 4, faz-se uma revisão acerca de alguns conceitos básicos subjacentes à aquisição de imagem por Ressonância Magnética. Apresentam-se as principais sequências de pulso utilizadas para aquisição de imagem e faz-se uma síntese relativa às técnicas de aquisição rápida e em tempo real. Referem-se os artefactos mais comuns e susceptíveis de interferir com a qualidade de imagem, em particular, nos estudos de produção de fala.

Como o objectivo principal desta dissertação é a obtenção de dados relativos à produção de fala utilizando RM, no capítulo 5, apresenta-se uma revisão dos principais estudos, efectuados para outras línguas, que recorreram à RM para estudar os mecanismos relativos à produção de fala. Dividem-se os estudos em estáticos, dinâmicos ou em tempo real, em função do tipo de aquisição efectuada.

Depois de conhecer o estado da arte relativamente aos estudos RM na área de fala, e da delimitação dos objectivos, elaborou-se o corpus para este trabalho e planeou-se a sua aquisição. A apresentação do corpus, os informantes, o protocolo de aquisição de imagem e de gravação do sinal de voz e, finalmente, uma síntese relativamente às técnicas de segmentação utilizadas no processamento de imagem constituem o capítulo 6, relativo às metodologias.

No capítulo 7, faz-se a apresentação e discussão dos resultados. Apresentam-se, no início, os resultados do Corpus 2D, onde se faz uma descrição articulatória de alguns sons com base em contornos, sobreposição de contornos e parâmetros articulatórios. Mostram-se os resultados relativos ao estudo de fricativas e oclusivas em contexto. Relativamente ao corpus 3D, apresentam-se as funções de área para algumas vogais [i, a, u], e alguns parâmetros obtidos para vogais orais, nasais e consoantes nasais. Finalmente, são apresentados os primeiros resultados do Corpus obtido em tempo real. No final de cada secção, apresenta-se uma pequena discussão.

O oitavo e último capítulo encerra a presente dissertação. Apresentam-se as principais conclusões deste estudo, bem como as suas limitações, apontam-se algumas sugestões com vista a trabalho futuro.

Para complementar a dissertação apresentam-se dois anexos. O anexo I é constituído pelas tabelas relativas aos símbolos dos alfabetos fonéticos para o PE: (1) Alfabeto Fonético Internacional e (2) SAMPA. Do anexo II, fazem parte as tabelas com todos os fones adquiridos relativos ao corpus 2D e 3D (aquisição em volume).

1.4 Relatórios e publicações efectuadas, no âmbito da dissertação

Apresenta-se, de seguida algum, do trabalho mais relevante realizado no âmbito desta dissertação.

Projecto de investigação submetido ao Serviço de Imagiologia dos Hospitais da Universidade de Coimbra (Junho de 2006)

Numa fase inicial deste trabalho, foi elaborado um projecto de investigação, submetido a aprovação ao Exmo Director do Serviço de Imagiologia dos Hospitais da Universidade de Coimbra, para fundamentar e viabilizar a aquisição das imagens no Sector de Ressonância Magnética, do referido hospital:

Martins, P., Carbone, I., Teixeira, A., Silva, A. (2006). *Ressonância Magnética estática e dinâmica aplicada ao estudo da Fonética Articulatória do Português Europeu (PE)*. IEETA. Universidade de Aveiro, Aveiro.

Relatório HERON (Dezembro de 2006)

Foi efectuado um relatório técnico (Laboratório de Processamento de Sinal do IEETA) relativo ao corpus de RM recolhido no âmbito da presente dissertação e enquadrado na Tarefa 1 do projecto HERON, em cuja elaboração a autora participou activamente:

Carbone, I., Martins, P., Teixeira, A., Silva, A. (2006). Base de Dados RM dos Sons do Português Europeu. Projecto HERON. Relatório técnico. IEETA/Universidade de Aveiro, Aveiro.

Artigo aceite para publicação em actas de conferência e apresentação oral

A parte relativa aos resultados obtidos para os sons nasais, foi submetida e aceite para apresentação oral no InterSpeech 2007, conferência principal na área do processamento de voz, a realizar em Agosto de 2007, em Antuérpia:

Martins, P., Carbone, I., Silva, A., Teixeira, A. (2007). *An MRI study of European Portuguese nasals*. InterSpeech 2007. Antuérpia, Bélgica.

Artigos Submetidos

À data da defesa desta dissertação encontram-se também submetidos os seguintes trabalhos:

Carbone, I., P. Martins, et al. (2007). *Vocal Tract Segmentation and Analysis Over a European Portuguese MRI Database*. Revista periódica do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro (Submetido em Abril de 2007).

Martins, P., Carbone, I., Pinto, A., Silva, A., Teixeira, A. (2007). *European Portuguese MRI Based Speech Production Studies*. SPEECH COMMUNICATION – Special issue on Iberian Languages (Submetido em Maio de 2007).

Capítulo 2

Sons do Português Europeu

No presente capítulo, definem-se alguns conceitos básicos relacionados com a fonética e fonologia do PE. O objectivo é proporcionar uma explicação simples destes conceitos a todos os que não estão familiarizados com a área das ciências da fala, mas que estejam interessados na leitura deste trabalho. Assim, faz-se uma descrição das estruturas anatómicas e dos processos envolvidos na produção de fala, segue-se uma caracterização dos sons da fala, em particular dos sons do PE. Explicam-se também alguns conceitos como nasalidade e coarticulação, pois surgem ao longo desta dissertação. No final do capítulo, faz-se uma referência às técnicas mais comuns utilizadas em síntese de fala, em particular à síntese articulatória, uma vez que um dos objectivos deste trabalho é a obtenção de dados anatómicos e funcionais que permitam aperfeiçoar a síntese dos sons do PE.

2.1 Conceitos de Fonética e Fonologia

2.1.1 Fonética

A fonética é a disciplina científica que estuda os sons da fala, em particular o modo como estes são produzidos e percebidos pelos falantes de uma língua (Mateus, Andrade, Viana, & Villalva, 1990). É uma área de intervenção interdisciplinar que integra conhecimentos de diversas áreas científicas como a Anatomia e Fisiologia, Neurologia, Acústica, Linguística, Processamento de Sinal, entre muitas outras (Andrade & Viana, 1996).

Uma vez que, quando há transmissão de uma mensagem entre o locutor e o ouvinte, num acto de comunicação oral, se podem definir três estádios sucessivos: (1) mecanismos de produção de fala, (2) transmissão de ondas sonoras, e (3) mecanismos de percepção, também a fonética é geralmente dividida em três vertentes: a fonética articulatória, acústica e perceptiva (Andrade & Viana, 1996; Mateus, Andrade, Viana, & Villalva, 1990; Mateus, Falé, & Freitas, 2005).

A fonética articulatória estuda o modo como os sons da fala são produzidos, tendo em conta a posição e o movimento dos diversos articuladores. É, por isso, importante conhecer os conceitos anatómicos que se apresentam mais adiante neste capítulo.

A fonética acústica é o ramo da fonética que se ocupa principalmente do estudo e descrição das propriedades físicas dos sons da fala (Mateus, Falé, & Freitas, 2005; Moutinho, 2000).

A fonética perceptiva descreve o modo como os elementos acústicos (objectivos), as ondas sonoras, são percebidos e interpretados pelo sistema auditivo humano de forma subjectiva. De acordo com Moutinho (2000), à fonética perceptiva não se deve chamar fonética auditiva (como acontece por vezes), uma vez que os mecanismos envolvidos na percepção de um som vão muito além da simples audição.

2.1.2 Fonologia

O estudo das características dos sons é do domínio da fonética, mas o modo como os diferentes sons se combinam numa determinada língua, permitindo distinguir significados é já do domínio da Fonologia. Pode dizer-se que a Fonologia determina as distinções fonéticas fundamentais para a comunicação linguística (Moutinho, 2000).

Fonemas

São unidades distintivas, ou seja, são elementos ou segmentos fonológicos que distinguem palavras. A substituição de um fonema por outro provoca alteração do significado (Mateus, Falé, & Freitas, 2005). A função distintiva dos fonemas manifesta-se nas oposições mínimas que a língua apresenta (uma só unidade os distingue). Só distinguimos **bola de mola** ou **bala de mala** porque /b/ e /m/ são unidades distintivas ou fonemas. Estas duas unidades têm em comum o ponto de articulação, o modo de articulação e a presença de vozeamento. O que as distingue é a nasalidade, /m/ é nasal, /b/ não é. A estes pares de palavras (**bola/mola**; **bica/mica**) que permitem a distinção dos fonemas /b/ e /m/ chamam-se **pares mínimos** (Moutinho, 2000) ou **pares distintivos** (Mateus, Andrade, Viana, & Villalva, 1990).

Fones

Às diferentes realizações fonéticas dos fonemas chamam-se fones (Mateus, Andrade, Viana, & Villalva, 1990; Moutinho, 2000). Todos eles deverão ter um grau de semelhança suficiente que permita classificá-los como realizações acústicas pertencentes à mesma classe. O fone, ao contrário do fonema, não é uma entidade abstracta, mas sim uma unidade física real (Simões, 1999).

Alofones

Os fones que correspondem a um mesmo fonema chamam-se alofones ou variantes do mesmo fonema. Por exemplo nas palavras **mal** e **mala**, há duas realizações distintas de /l/. Realiza-se como [l~] (velarizado) na palavra **mal** [mal~] uma vez que se encontra em final de sílaba e como

[l] na palavra mala [mal6], que se encontra em posição de início de sílaba (Moutinho, 2000). Estas variantes dependem do contexto podendo chamar-se variantes contextuais ou combinatórias. Os alofones que não dependem do contexto, mas sim de variedades dialectais designam-se por variantes livres (Mateus, Falé, & Freitas, 2005; Moutinho, 2000).

2.1.3 Alfabetos Fonéticos

Um alfabeto fonético é um conjunto de símbolos criados com o intuito de representar graficamente e de forma simples os sons da fala, uma vez que os sistemas ortográficos são inadequados e causadores de ambiguidades (Andrade & Viana, 1996; Martins, 1988; Mateus, Falé, & Freitas, 2005; Moutinho, 2000). O princípio básico subjacente à criação destes alfabetos é a necessidade de ter uma relação **biunívoca** entre **um** som e o **seu** símbolo. Podemos dizer que **um símbolo** representa apenas **um som** e **um som** é representado por **um único símbolo** (Martins, 1988; Mateus, Falé, & Freitas, 2005; Moutinho, 2000).

Os primeiros alfabetos fonéticos datam do século XVII, mas foi a partir do final do século XIX que começaram a surgir inúmeras propostas de alfabetos fonéticos (Martins, 1988; Moutinho, 2000). Além da função de representar os sons da fala, os alfabetos fonéticos têm ainda aplicação no ensino de línguas estrangeiras e aprendizagem da leitura e escrita (Andrade & Viana, 1996; Martins, 1988; Moutinho, 2000). Um dos alfabetos fonéticos mais conhecidos é o Alfabeto Fonético Internacional (AFI) - ou IPA (*International Phonetic Alphabet*), o acrónimo em língua Inglesa. O objectivo principal subjacente à criação deste alfabeto foi a representação da pronúncia de várias línguas de uma forma racional. Tem sofrido várias actualizações ao longo dos anos, de forma a contemplar sons de várias línguas e corrigir algumas incoerências. Para além dos símbolos, inclui também outros sinais, os diacríticos ou modificadores, como por exemplo o til (~) que serve para indicar nasalização nas vogais do Português. A grande desvantagem deste alfabeto é que utiliza caracteres normalmente não imprimíveis. Assim, neste trabalho utilizaremos um outro alfabeto fonético, o **SAMPA** (SAM Phonetic Alphabet) desenvolvido por um grupo de foneticistas, no âmbito do projecto ESPRIT 1541, **Speech Assessment Methods** (SAM) entre 1987-89, para algumas línguas da Comunidade Europeia. Ao longo dos anos foi contemplando outras línguas, como o Norueguês e Sueco, em 1992, e, em 1993, foi alargado ao Português Europeu, Grego e Espanhol. Foi utilizado para transcrever a versão para o PE do sub-corpus de fala EUROM (Andrade & Viana, 1996; Ribeiro, 2003).

Basicamente, o alfabeto **SAMPA** faz o mapeamento do IPA em códigos ASCII, em particular, os caracteres ASCII de 7 bits imprimíveis. Associado ao código, encontram-se normas para a transcrição das línguas às quais este foi aplicado. Este alfabeto resulta da cooperação de vários especialistas e investigadores de vários países e é uma ferramenta importantíssima no âmbito das ciências da fala. Está disponível *on-line*, em www.phon.ucl.ac.uk, para várias línguas, incluindo

também o PE. Na secção de anexos (Anexo 1) incluem-se os símbolos do alfabeto IPA e SAMPA para o PE.

2.1.4 Transcrição Fonética

De acordo com Moutinho (2000), uma transcrição fonética de um enunciado é um acto individual, podendo variar de um transcritor para outro. O transcritor pode ser influenciado e condicionado pelas suas concepções teóricas e hábitos linguísticos. Existem diferentes tipos de transcrição fonética, usualmente considera-se a **transcrição larga** e a **transcrição estreita**. A escolha de um, ou outro tipo de transcrição, deve variar em função dos objectivos a atingir com a mesma (Andrade & Viana, 1996; Martins, 1988; Moutinho, 2000). Assim, se o objectivo é obter uma representação absolutamente fiel de todos os factos observados, apresentando **maior detalhe fonético**, dever-se-á optar por uma **transcrição estreita**. Se pretendermos apenas **marcar o contraste** que permite a distinção entre dois sons, pode-se optar por uma **transcrição larga**, mas consequentemente menos pormenorizada (Moutinho, 2000). Existem também algumas convenções que deverão ser seguidas quando se efectua uma transcrição. Assim, as transcrições fonéticas deverão aparecer sempre entre parênteses rectos [] e as transcrições fonológicas entre //. Também a indicação da sílaba portadora de acento se deve fazer através do uso de um diacrítico (´), símbolo semelhante ao do acento agudo, colocando-o sobre a vogal da respectiva sílaba ou a anteceder a sílaba na totalidade (Moutinho, 2000).

2.2 Aparelho fonador e produção de fala

Anatomicamente o aparelho fonador é constituído pelas cavidades subglotais ou sublaríngeas, laringe e cavidades supraglotais ou supralaríngeas como se pode observar na Figura 2.1. O tracto ou cavidade oral, o tracto nasal e a cavidade faríngea constituem as cavidades supraglotais (que descreveremos adiante no ponto 2.2.2), responsáveis pela ressonância e pela articulação, essenciais para a produção dos diferentes sons de uma língua. A laringe é constituída pelas pregas ou cordas vocais e pela glote. Finalmente, as cavidades subglotais são constituídas essencialmente pela traqueia, brônquios e pulmões (que geram o fluxo de ar utilizado na fonação) (Andrade & Viana, 1996; Kent, 1997; Mateus, Falé, & Freitas, 2005; Moutinho, 2000; Teixeira, 2000).

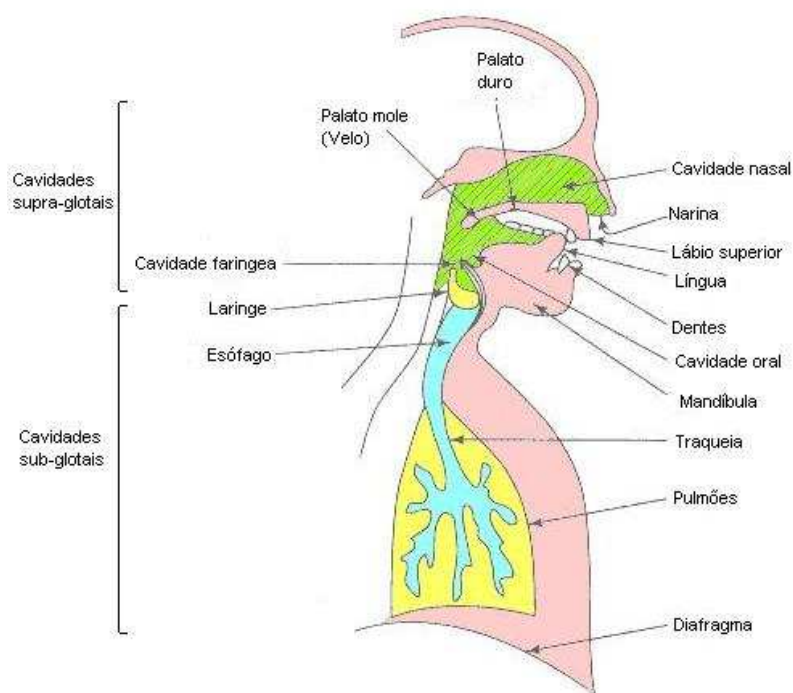


Figura 2. 1 - Estruturas anatómicas do aparelho fonador. Fonte: Adaptado de Deller et al. (1993).

2.2.1 Mecanismos fisiológicos na produção de fala

Do ponto de vista fisiológico, a produção de fala depende de três componentes: a **respiração**, a **fonação** e a **articulação** (Mateus, Falé, & Freitas, 2005; Moutinho, 2000; Zemlin, 2000).

Respiração

A respiração encerra um conjunto de processos vitais (trocas gasosas, controlo do pH), cuja finalidade é o fornecimento de oxigénio ao sangue e, através dele, a todas as células do corpo (Moutinho, 2000; Seeley, Stephens, & Tate, 2005). A par com esta função fisiológica vital, a respiração é também fundamental e ponto de partida para a produção de sons e de fala, se bem que, contribui muito pouco para distinguir os sons entre si (Moutinho, 2000).

A respiração engloba dois momentos ou ciclos distintos: a **inspiração** e a **expiração**. Durante a inspiração, há contracção dos músculos respiratórios (diafragma, músculos intercostais, pequeno peitoral e, em menor escala, os músculos escalenos), fazendo com que haja um aumento das dimensões torácicas nas três dimensões. Em consequência, há expansão da cavidade pulmonar com consequente diminuição da pressão, o que permite a entrada de ar até que as pressões se igualem (Moutinho, 2000; Zemlin, 2000). Na expiração, o ar é expelido para o exterior devido à

redução do volume da caixa torácica e ao aumento da pressão interior. A expiração é o “motor” da fala, pois fornece a pressão necessária para fazer vibrar as pregas vocais (Moutinho, 2000).

Estes dois processos são os responsáveis pela geração dos fluxos de ar (ingressivo e egressivo), que intervêm na produção dos sons da fala. O fluxo de ar **ingressivo** dirige-se para o interior do corpo e o fluxo de ar **egressivo** dirige-se para o exterior (na expiração). O mecanismo do fluxo de ar poderá ser também pulmonar, glótico e velar. O iniciador do fluxo pulmonar é a respiração (fluxo de ar proveniente dos pulmões), no caso do fluxo glotal o responsável é o movimento vertical da laringe, com a glote fechada (sons não vozeados) ou a vibrar (sons vozeados), comprimindo ou rareando a pressão do ar no tracto vocal acima da glote. No fluxo velar, é a língua o iniciador. Os sons produzidos com um **fluxo de ar egressivo** glotal chamam-se **ejectivos**, enquanto que os produzidos com fluxo **ingressivo glotal** são **implosivos**. Aos sons produzidos com um **fluxo velar ingressivo** chamam-se **cliques** (Laver, 1995). A maior parte dos sons, nas diversas línguas, incluindo o Português, são produzidos durante a fase de expiração com modulação do fluxo de ar **egressivo** (Andrade & Viana, 1996; Laver, 1995; Mateus, Falé, & Freitas, 2005; Moutinho, 2000).

Fonação

A fonação resulta da actividade vibratória das pregas vocais, situadas na laringe. A laringe é uma estrutura ímpar, mediana, músculo-cartilaginosa e bastante complexa, situada na região anterior do pescoço (Kent, 1997; Zemlin, 2000). Da sua estrutura fazem parte nove cartilagens: três ímpares (a cartilagem tiróide, a cricóide e a epiglote), três cartilagens pares de menor dimensão (as aritenóides, as corniculadas e as cuneiformes) e um conjunto de músculos e ligamentos intrínsecos e extrínsecos que em muito contribuem para a sua complexidade (Kent, 1997; Seeley, Stephens, & Tate, 2005; Zemlin, 2000).

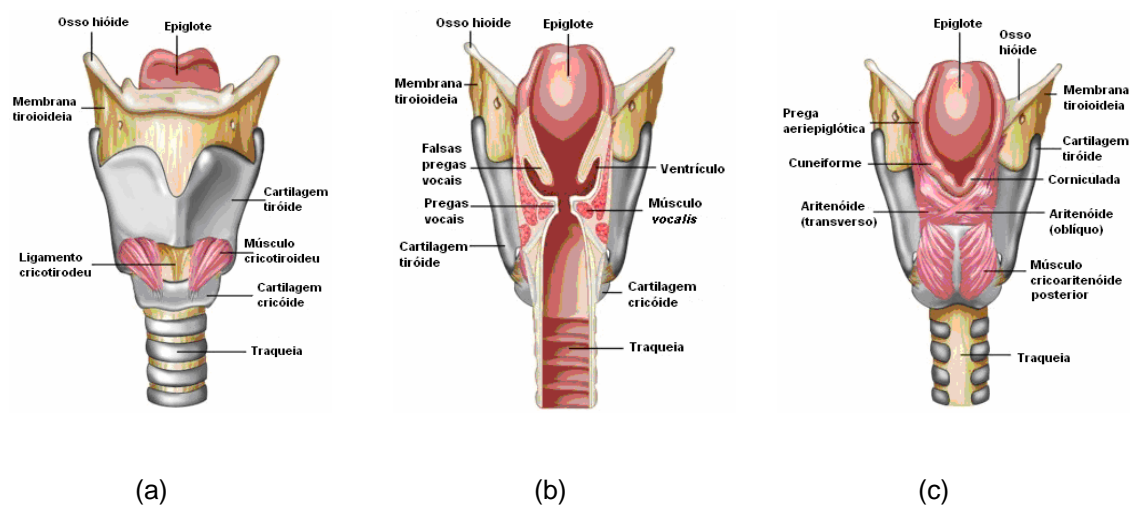


Figura 2. 2 - Laringe. Vista anterior (a), mediana e posterior (c). Adptado de: www.gbmc.org/voice.

Das cartilagens ímpares, a tiróide é a que tem maior dimensão, a cartilagem cricóide é a que tem uma localização mais inferior formando a base da laringe, e na qual todas as outras se apoiam (Figura 2.2a) A epiglote, está ligada à cartilagem tiróide e projecta-se em direcção à raiz da língua. Tem uma função importante, pois como recobre a laringe, previne a entrada de alimentos para as vias aéreas, durante a deglutição. As aritenóides (par) articulam-se com o bordo postero-superior da cartilagem cricóide e o par das corniculadas liga-se às aritenóides. As cartilagens cuneiformes (em forma de cunha) encontram-se numa posição anterior em relação às corniculadas (Seeley, Stephens, & Tate, 2005; Zemlin, 2000).

Os músculos extrínsecos da laringe são aqueles que a suportam, fixam e que permitem a sua movimentação como um todo. Os intrínsecos são responsáveis pelo controlo da produção do som. Os músculos intrínsecos são, em geral, classificados tendo em atenção o seu efeito sobre a forma da glote e na vibração das pregas vocais (abdutores, adutores, tensores e relaxadores), actuando sempre aos pares (Zemlin, 2000). Não havendo espaço para uma descrição detalhada destes músculos, referiremos apenas que: o principal abductor é o músculo cricoaritenóideu posterior (CAP), o maior adutor é o músculo cricoaritenóideu lateral (CAL), sendo o antagonista do CAP. O CAL é responsável, não só, pela adução das pregas, mas também por uma **força de compressão medial** que mantém as pregas próximas. O cricotiroideu faz o ajuste na relação entre a tiróide e a cricóide, o que se reflecte nas aritenóides e, em consequência, no comprimento e tensão das pregas vocais (Kent, 1997; Zemlin, 2000).

O interior da laringe ou cavidade laríngea divide-se no espaço **supraglótico** e **infraglótico** (ou subglótico), tendo como referência a **glote** e as pregas vocais. Ao espaço supraglótico definido entre as pregas vestibulares (falsas pregas vocais) e o ádito chama-se vestibulo (Figura 2.2 b); o espaço entre as pregas vestibulares e as pregas vocais denomina-se ventrículo. O espaço infraglótico desta cavidade é delimitado, em cima, pelas pregas vocais e, inferiormente, pela margem inferior da cartilagem cricóide (Kent, 1997; Zemlin, 2000).

As pregas vocais situam-se logo abaixo das pregas vestibulares, ligam-se anteriormente à cartilagem tiróide (abaixo da incisura tiróideia) e posteriormente ligam-se à face antero-lateral das cartilagens aritenóides. Ao espaço variável (geralmente de configuração triangular) que existe entre as pregas vocais chama-se glote. No entanto, as dimensões e configuração da glote podem variar em função da actividade laríngea e dos ajustes das cartilagens aritenóides (Zemlin, 2000).

As pregas vocais podem assumir posições distintas:

- (1) Durante a **respiração normal** (Figura 2.3a), as cartilagens aritenóides estão afastadas (abduzidas), com **afastamento** das pregas vocais e há passagem de ar;

- (2) Para que haja **produção de sons** as cartilagens aritenóides estão em **adução**, aproximando as pregas vocais, limitando a passagem de ar na glote e criando as condições para que exista vibração (Figura 2.3 b);
- (3) Vibração - a glote entra em vibração (abertura e fechamento extremamente rápido das pregas vocais) por acção da pressão subglotal criada pelo fluxo de ar proveniente dos pulmões.

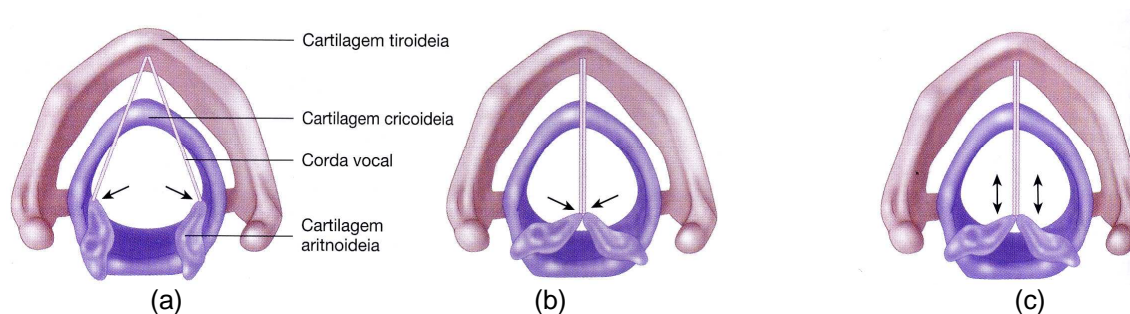


Figura 2. 3 - Posição das pregas vocais durante: (a) respiração normal, em abdução; (b) fonação, em adução e (c) regulação da tensão nas pregas. Fonte: Seeley et al. (2005).

A vibração só é possível se: (1) existir uma diferença de pressão e (2) a posição e estado de tensão das pregas for adequado (Figura 2.3 c). Em termos de pressão, a pressão subglotal deve ser superior à pressão supraglotal para que exista uma força suficiente para afastar as pregas vocais. Com as pregas abduzidas, o ar escapa-se com velocidade elevada, causando uma diminuição temporária da pressão transglotal (Efeito de Bernoulli) que causa uma nova aproximação das pregas (adução). Desta forma, (Andrade & Viana, 1996; Mateus, Andrade, Viana, & Villalva, 1990) a pressão subglotal torna a aumentar e repete-se o ciclo de abertura/fecho das pregas vocais.

Quando não há vibração total das pregas vocais ou quando existe uma adução incompleta durante a vibração produzem-se tipos de voz diferente da voz modal (também denominada normal ou padrão) que se designam, respectivamente, por voz laringalizada e voz murmurada (Andrade & Viana, 1996).

Do que foi exposto, se pode concluir que a laringe é um órgão de extrema relevância na produção de fala, e fundamental também na descrição fonética dos sons, uma vez que permite distinguir entre os sons produzidos com vibração das pregas vocais (sons vozeados ou sonoros) e sons produzidos sem vibração (sons não vozeados ou surdos) (Mateus, Falé, & Freitas, 2005; Moutinho, 2000). Como veremos adiante no ponto 2.3.1.1, todas as vogais do PE são produzidas com vibração das pregas vocais.

Articulação

Como vimos, o fluxo de ar proveniente dos pulmões é modulado ao nível das cavidades supraglotais (cavidade faríngea, tracto nasal, cavidade oral ou bucal). As diferentes configurações destas cavidades são obtidas quando os articuladores (que descreveremos no ponto 2.2.3) se dispõem de forma diferente para produzir os diferentes sons.

De forma resumida, podemos dizer que o fluxo de ar vindo dos pulmões, passa pela traqueia e atinge a laringe sob a forma de uma corrente de ar não modulada. Os movimentos de abertura/fecho rápido das pregas vocais interrompem periodicamente o fluxo de ar subglotal proveniente dos pulmões, de modo a produzir um som glótico ou vocal no interior da cavidade faríngea, oral ou nasal. Ao modificarem a sua configuração, estas cavidades alteram as suas propriedades acústicas, transformando um som indiferenciado num som de fala com significado (Zemlin, 2000).

Descrevemos de seguida a anatomia das cavidades supraglóticas e a forma como se dispõem para produzir diferentes sons.

2.2.2 Cavidades supraglóticas ou tracto vocal¹

São constituídas pelo tracto nasal, cavidade faríngea e tracto oral (Figura 2.4). Começamos por descrever o tracto nasal e só depois o tracto oral, uma vez que finalizamos com a descrição dos articuladores, situados maioritariamente na cavidade oral.

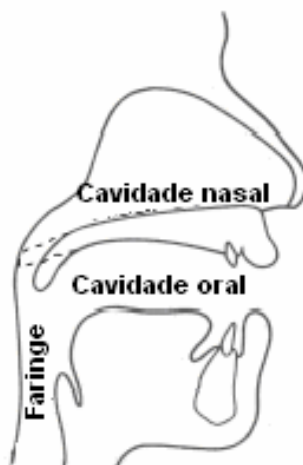


Figura 2. 4 - Cavidades supraglóticas ou tracto vocal. Fonte: Adaptado de Pruthi (2006).

¹ Ao longo desta dissertação chamamos **tracto vocal** ao conjunto das cavidades supralaríngeas. Assim, consideramos que o tracto vocal é constituído pela cavidade oral ou tracto oral, tracto nasal e cavidade faríngea, tal como podemos encontrar em (Pruthi, 2006; Kent, 1997; Zemlin, 2000). Outros autores (Mateus, Andrade, Viana, & Villalva, 1990; Mateus, Falé, & Freitas, 2005; Moutinho, 2000) utilizam o termo tracto vocal quando se referem apenas à cavidade oral ou tracto oral, considerando o tracto nasal separadamente.

2.2.2.1 Tracto nasal ou cavidades nasais

O tracto nasal, ao contrário da cavidade oral, tem uma configuração estática, na sua maior parte. No entanto, alterações ao nível da mucosa nasal (inflamatórias) podem alterar de forma significativa a estrutura das cavidades ao longo do tempo. A única porção do tracto nasal que pode ser responsável por alguma variação dinâmica desta cavidade é a sua porção postero-inferior - o velo. O velo pode encontrar-se abaixado ou elevado, permitindo ou não o acoplamento com o restante tracto vocal (Pruthi, 2006; Teixeira, 2000).

A estrutura do tracto nasal é bastante mais complicada do que a do tracto oral, uma vez que se encontra dividida, pelo septo nasal, em duas passagens paralelas, geralmente assimétricas. As cavidades nasais comunicam com o exterior através das narinas e posteriormente com a nasofaringe através das coanas (Drake & Vogl, 2005; Zemlin, 2000).

As paredes laterais das cavidades nasais, ver Figura 2.5 (a), são compostas pelas conchas nasais: (1) superior, (2) média e (3) inferior e por passagens (os meatos). Esta estrutura faz com que a área de superfície das cavidades aumente e, em conjunto com o revestimento mucoso, interfiram no amortecimento do sinal acústico em todas as frequências (Entenman citado por Teixeira (2000)). Através de orifícios, as paredes laterais comunicam com estruturas arejadas - os seios perinasais. O pavimento das cavidades nasais é côncavo (formado pelas maxilas e osso palatino).

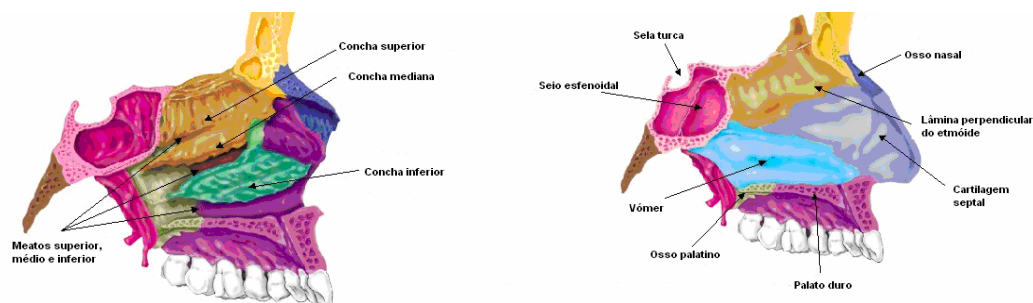


Figura 2. 5 - Tracto nasal: vista lateral (a) e medial (b). Adaptado de: <http://webanatomy.net>.

Alguns autores (Dang, Honda, & Suzuki, 1994; Fant, 1960; Story, 1995) observaram e mediram o comprimento do tracto nasal, utilizando métodos de imagem (Raios X e RM). Story (1995) utilizando RM, e para um falante, verificou que o tracto nasal media cerca de 8 cm desde a bifurcação às narinas e cerca de 2,1 cm entre a zona de acoplamento e a bifurcação. Dang, Honda, & Suzuki (1994) obtiveram também, com RM, um valor de 11,6 cm com desvio padrão de 0,13 cm. Verificaram também, que existe assimetria entre as passagens nasais e que existem diferenças entre indivíduos. Fant (1960) obteve, a partir de imagens radiográficas, um valor de 12,5 cm para o comprimento do tracto nasal.

Seios Perinasais

Os seios perinasais são em número de quatro (Figura 2.6): (1) seios frontais, (2) seios maxilares, (3) seios etmoidais e (4) seio esfenoidal. Todos eles são recobertos por mucosa, segregam muco, contêm ar e abrem-se nas cavidades nasais (Drake & Vogl, 2005; Teixeira, 2000). De acordo com Maeda (1982) citado por Teixeira (2000), estas estruturas desempenham um papel importante no aparecimento das ressonâncias de frequência mais baixa nos sons nasais. No entanto, esta função dos seios perinasais na produção de sons é ainda objecto de alguma controvérsia.

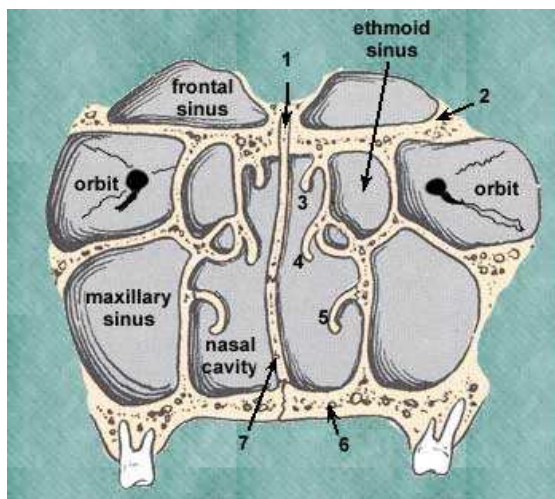


Figura 2. 6 - Seios perinasais. Fonte: <https://mywebpages.comcast.net>.

2.2.2.2. Cavidade faríngea

A faringe é um tubo músculo-membranoso que liga as cavidades nasais e a cavidade oral à laringe e ao esófago, sendo comum aos aparelhos respiratório e digestivo. Estende-se desde a base do crânio até ao nível da sexta vértebra cervical (C6). Tem cerca de 12 cm de comprimento e divide-se em três porções. (1) nasofaringe, (2) orofaringe e (3) laringofaringe (Drake & Vogl, 2005; Kent, 1997; Seeley, Stephens, & Tate, 2005; Zemlin, 2000).

As aberturas posteriores do tracto nasal, as coanas, abrem-se na nasofaringe. O limite inferior desta porção está ao nível do palato mole ou véu do palato. A abertura posterior da cavidade oral abre-se na orofaringe através das *fauces*. Finalmente, a laringofaringe, que se estende desde o topo da epiglote ao esófago, situa-se posteriormente em relação à laringe (Drake & Vogl, 2005; Seeley, Stephens, & Tate, 2005; Zemlin, 2000).

2.2.2.3 Vestíbulo oral

Espaço delimitado externamente pelos lábios e bochechas e internamente pelas gengivas e dentes. Comunica com a cavidade oral ou boca, com as maxilas fechadas, através do espaço entre os dentes (Zemlin, 2000).

2.2.2.4 Tracto oral ou cavidade oral

O tracto oral, cavidade oral ou boca propriamente dita, é limitado anterior e lateralmente pelos dentes e alvéolos; posteriormente pelo arco palatoglossos; superiormente pelo palato duro e palato mole e inferiormente pelo pavimento muscular, constituído essencialmente pela língua. Do ponto de vista biológico, a boca estabelece a comunicação com o exterior dos sistemas digestivo e respiratório. As estruturas da boca ou cavidade oral podem modificar as características de ressonância do tracto vocal e produzir sons de fala (Zemlin, 2000).

Do ponto de vista articulatorio, a cavidade oral é frequentemente dividida em 2 zonas:

- (1) zona anterior - entre os lábios e o palato duro
- (2) zona posterior - desde o palato duro à parede posterior da faringe (Mateus, Falé, & Freitas, 2005)

2.2.3 Articuladores e ponto de articulação

Os articuladores são as estruturas responsáveis pela articulação dos sons e situam-se na cavidade oral (Figura 2.7). Ao assumirem configurações diferentes, os articuladores produzem sons diferentes e regulam a produção de fala. Existem dois tipos de articuladores: (1) passivos e (2) activos. Os activos, e que regra geral apresentam mobilidade, são a língua, os lábios, o palato mole e a mandíbula (que geralmente actua em conjunto com os lábios e a língua). Os articuladores passivos são os alvéolos dentários superiores, a porção anterior da arcada dentária superior e o palato duro (Andrade & Viana, 1996; Mateus, Falé, & Freitas, 2005).

De todos os articuladores, a língua é o que apresenta maior mobilidade e flexibilidade, podendo ser encarada como um conjunto de articuladores relativamente independentes (Andrade & Viana, 1996; Teixeira, 2000). Esta mobilidade resulta da sua função biológica primária, uma vez que durante a mastigação e deglutição, a língua tem que mover os alimentos dentro da boca e faringe (Laver, 1995).

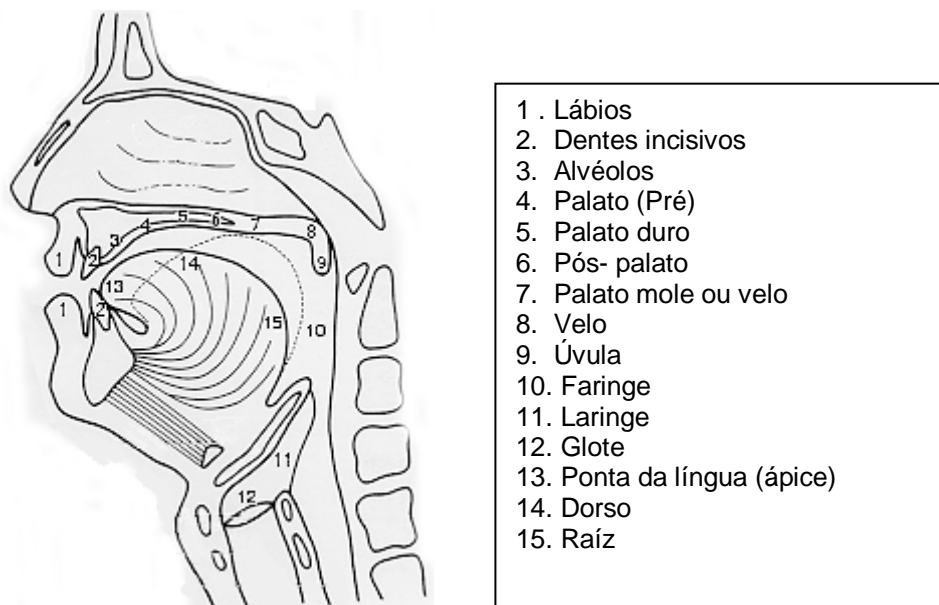


Figura 2. 7 - Articuladores e pontos de articulação. Adaptado de: www.ph.ludwigbsurg.de.htm.

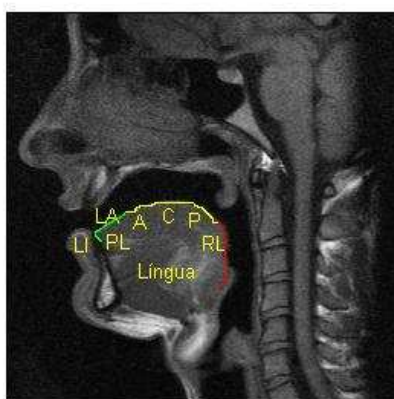
2.2.3.1 Articuladores activos:

Lábios

Os lábios situam-se na parte anterior do tracto vocal podendo obstruir total ou parcialmente a passagem do fluxo de ar para o exterior.

Língua

A língua funciona de modo a alterar a forma da cavidade oral, modificando as características de ressonância desta cavidade e das cavidades associadas. Pode dividir-se a língua (do ponto de vista articulatorio) em: (1) coroa, (2) dorso ou corpo e (3) raiz da língua (ver Figura 2.8). A coroa é a zona frontal da língua e divide-se em ponta ou ápice e lâmina. O dorso situa-se ente a lâmina e a raiz da língua, e é a parte mais extensa deste articulador. O ponto de articulação da maior parte das vogais acontece no dorso da língua. O dorso pode ainda dividir-se em porção anterior ou frontal e porção posterior (Mateus, Falé, & Freitas, 2005). A porção anterior situa-se abaixo do palato duro e a porção posterior abaixo do velo (Ladefoged, 2001). Alguns autores (Ladefoged, 2001; Teixeira, 2000) consideram ainda uma porção média no dorso lingual. A raiz da língua é a porção mais posterior da língua e encontra-se no lado oposto ao da parede da faringe, à frente da epiglote (Ladefoged, 2001; Mateus, Andrade, Viana, & Villalva, 1990; Mateus, Falé, & Freitas, 2005; Teixeira, 2000).



Lábio Inferior (LI)

Coroa:

Lâmina (LA)

Ponta da língua (PL)

Dorso da língua:

Porção anterior (A)

Porção central (C)

Porção posterior (P)

Raiz da língua (RL)

Figura 2. 8 - Divisão da língua, do ponto de vista articulatorio. Imagem sagital obtida durante a produção de [6], no falante PAA.

Palato mole, véu palatino, velo ou *velum*

É considerado como a continuação flexível do palato duro, tendo na sua extremidade uma estrutura móvel, a úvula (Mateus, Falé, & Freitas, 2005). A sua principal função é separar as cavidades nasais da cavidade oral. Quando baixo, permite o acoplamento entre cavidades e passagem do ar pelas cavidades nasais, com produção de sons nasais. Quando está elevado impede a passagem do ar pelas cavidades nasais e são produzidos sons orais (Teixeira, 2000). Entre uma posição aberta e uma posição em que se encontra completamente fechado, existem vários graus de acoplamento com graus de abertura intermédios. Só por si, o facto da passagem velofaríngea se encontrar aberta não significa a percepção da nasalidade (Laver, 1980), sendo também diferente o grau de acoplamento na produção dos vários sons.

Childers e Ding (1991), citados por Teixeira (2000) referem que o grau de abertura do velo varia em função do contexto fonético, encontrando-se aberto na produção dos sons nasais, numa posição intermédia para as vogais baixas, praticamente fechado para as vogais altas e completamente fechado na produção de oclusivas orais.

Num estudo cineradiográfico efectuado por Moll (Moll, 1960) refere-se que na produção das vogais baixas o velo apresenta um menor grau de fechamento enquanto que na produção de vogais altas o grau de fechamento é maior. Verifica-se também que as vogais adjacentes a consoantes nasais como o [n] apresentam fechamento incompleto e que há tendência para um menor fechamento em vogais isoladas do que em vogais inseridas em contexto de consoantes não nasais (p.e. [p] ou [k]).

Num estudo realizado por Rossato, Teixeira & Ferreira (2006), em que se estudaram algumas características de sons nasais e não nasais do PE e do francês, e no que diz respeito à **altura do**

velo (AV) os autores concluíram que para ambas as línguas se verifica que a AV segue a seguinte hierarquia: consoantes orais > vogais orais > consoantes nasais > vogais nasais. Verificaram também que na realização das vogais orais abertas, quando a altura do velo é semelhante à observada na produção de uma consoante nasal, a passagem velofaríngea encontra-se aberta.

Maxilar inferior ou mandíbula

A mandíbula alberga os dentes inferiores que, de forma semelhante aos dentes da arcada superior, estão ligeiramente curvados, vertical e lateralmente. Os dentes inferiores raramente têm papel relevante na produção de fala, excepto passivamente, causando um obstáculo ao fluxo de ar, como na produção de sons como o [s] e para o [t] (Perkell, 1999). A mandíbula pode também ser considerada a “transportadora” da língua, efectuando movimentos visíveis durante a fala, na maior parte dos falantes, quando a língua se eleva e baixa. No entanto, a língua é capaz de efectuar movimentos independentemente da mandíbula, devendo por isso ser considerada semi-independente da mandíbula (Laver, 1995).

Pode dizer-se que os movimentos da mandíbula têm um papel mais relevante na produção de articulações labiais e linguais, quer na produção de vogais quer de consoantes (Perkell, 1999).

2.2.3.2 Articuladores passivos

Sector anterior da arcada dentária superior (SAADS)

Localiza-se na porção supero-anterior da cavidade oral.

Alvéolos dentários superiores

Zona que se situa imediatamente atrás da implantação dos dentes anteriores da arcada dentária superior.

Palato duro

Estrutura óssea situada na parte superior da cavidade oral, imediatamente atrás da zona posterior dos alvéolos dentários superiores. O palato duro curva-se para cima, atingindo o ponto mais alto da cavidade oral. Logo atrás dos alvéolos dentários, o palato apresenta pequenas rugosidades que se pensa terem a função (além da biológica) de causar pequenos obstáculos ao fluxo de ar que se dirige para o exterior, tornando-o turbulento. Esta turbulência é responsável pelo ruído sibilante característico de algumas fricativas (Laver, 1995).

2.3 Sons da fala, em particular do PE

2.3.1 Classificação articulatória dos sons

A classificação articulatória dos sons da fala consiste na categorização dos sons, baseada essencialmente na observação dos movimentos dos articuladores (Mateus, Falé, & Freitas, 2005). Os sons da fala podem ser divididos em vogais e consoantes. As vogais são produzidas sem que existam obstáculos significativos que produzam oclusão ou fricção ao fluxo de ar proveniente dos pulmões. Nas consoantes existe sempre uma constricção significativa que condiciona a passagem do ar na cavidade faríngea ou na cavidade oral (Andrade & Viana, 1996; Cruz-Ferreira, 1999; Mateus, Falé, & Freitas, 2005; Moutinho, 2000; Teixeira, 2000).

2.3.1.1 Vogais

Como já se referiu, as vogais são sons que são produzidos sem constricções à passagem do ar e com vibração das cordas vocais, sendo por isso considerados sons vozeados ou sonoros. As vogais podem ser produzidas com o velo abaixado (aberto) ou elevado (fechado). Quando o velo está aberto (Figura 2.9 (b)) o ar é expelido pela boca e pelo nariz, dando origem à produção das vogais nasais.

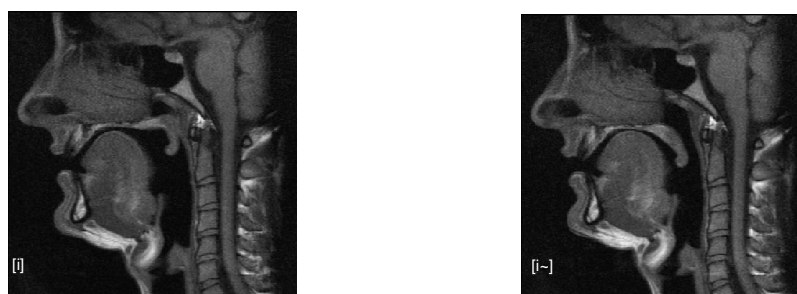


Figura 2. 9 - Posição do velo na produção da vogal oral [i] (à esquerda) e da vogal nasal [i~](à direita), para o falante PAA.

Nas vogais, ao contrário do que se verifica nas consoantes (que veremos a seguir) não é usual a referência a pontos de articulação, uma vez que não existe, neste caso, uma verdadeira obstrução do tracto vocal (Martins, 1988; Moutinho, 2000).

As vogais são habitualmente classificadas em função de:

- Altura (posição do dorso da língua)
- Grau de abertura
- Posição dos lábios ou labialização

A labialização dá conta do papel dos lábios na produção da vogal. Podemos distinguir entre vogais arredondadas ou labializadas e vogais não arredondadas ou não labializadas, produzidas com os lábios abertos. Entre estes dois extremos existem posições intermédias. No PE só as vogais posteriores ou velares [o, O, u] apresentam arredondamento dos lábios, não sendo, no entanto uma característica distintiva para o Português (Andrade & Viana, 1996; Mateus, Falé, & Freitas, 2005; Moutinho, 2000; Teixeira, 2000).

O grau de abertura depende da altura do dorso da língua e da abertura do maxilar inferior no momento da realização da vogal (Andrade & Viana, 1996; Teixeira, 2000). Tendo em atenção o grau de abertura classificam-se as vogais em: fechadas, semi-fechadas, semi-abertas e abertas².

Para um mesmo grau de abertura, o dorso da língua pode mover-se no sentido antero-posterior (avanço-recuo), podendo classificar-se as vogais em:

- (1) anteriores, avançadas ou palatais (dorso da língua avança em relação à posição neutra e em direcção ao palato);
- (2) médias ou centrais (dorso da língua em posição média);
- (3) posteriores, recuadas ou velares (dorso da língua em posição recuada em relação à posição neutra e elevada em direcção ao velo).

Em relação à altura do dorso língua (ponto mais alto) consideram-se 3 posições - alta, média e baixa. Na posição neutra ou de repouso a língua está numa posição central e média dentro da cavidade oral (Mateus, Falé, & Freitas, 2005).

Vogais orais do PE

De acordo com os critérios definidos no ponto anterior definem-se nove vogais orais (para o dialecto padrão) do Português Europeu: [a] (pato), [ɔ] (cada), [ə] (devi), [e] (pêga), [E] (leva), [i] (pipo), [o](tôpo), [O] (pote) e [u] (buda).

² Alguns autores classificam as vogais abertas ou fechadas, de acordo com a sua audibilidade (Mateus, Andrade, Viana, & Villalva, 1990). Assim [a], [E] e [O] são abertas e [i] e [u] são fechadas. Na literatura inglesa (Laver, 1995) a classificação das vogais em abertas ou fechadas tem por base considerar-se uma direcção vertical no espaço articulatório intra-oral e desta forma, uma vogal cuja produção é efectuada mais próxima do “céu da boca” (roof of the mouth, no original) será fechada e outra produzida na parte mais baixa será classificada como aberta. Entre estes extremos definem-se duas outras zonas que permitem classificar as vogais como semi-fechadas e semi-abertas.

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

Tabela 2. 1 - Classificação das vogais orais do PE. Fonte: Moutinho (2000).

Vogal neutra do PE

Existe alguma discussão na bibliografia consultada acerca da vogal neutra do PE e se a existir, esta configuração será idêntica à observada em outras línguas, frequentemente identificada como *schwa* (chevá é o termo em português utilizado por Andrade & Viana, 1996). Em termos de produção, o *schwa* será uma vogal produzida com vibração das pregas vocais, mas em que o restante aparelho fonador (boca, lábios) permanece numa posição de repouso.

A transcrição fonética da vogal que se poderá ou não descrever como a neutra ou *schwa*, o /e/ em palavras como *de*, *devi*, *mote*, *pote* levanta ainda maiores questões. Se considerarmos o alfabeto SAMPA, que optámos por utilizar neste trabalho, transcrevemos foneticamente o /e/ por [@]. O [@] no alfabeto SAMPA corresponde ao símbolo que se utiliza para o *schwa*. No entanto, a quantidade de símbolos utilizados por diferentes autores portugueses utilizando o alfabeto IPA, para representar este fone, é muito variável. Deixamos alguns apontamentos³, em nota de rodapé, que ilustram a complexidade do tema e mostram a variedade de símbolos utilizados.

³ O símbolo [ə] tem sido utilizado pela escola portuguesa tradicional para transcrever o /e/ em palavras como *mote*, *ataque* etc. No IPA este símbolo é reservado para o *schwa* nas diversas línguas (Andrade & Viana, 1996). Para o PE, Andrade e Viana (1996) referem que: “ a utilização do símbolo [i] reflecte uma preocupação mais fonética que tem que ver com o seguinte: a qualidade média dos chevás das diferentes línguas difere bastante, como se pode verificar comparando os do Português, do Francês e do Inglês. No âmbito dos símbolos do IPA para as vogais, o símbolo [i], representa mais adequadamente a qualidade do chevá do Português quando ele é realizado como uma vogal plena, do que o símbolo [ə]”.

Viana (1883), Cunha e Cintra (1986), citados por Veloso (1999) e Barbosa (1994) classificam esta vogal como central, fechada, alta e não arredondada, e que se esta descrição for adequada e correcta, o símbolo que lhe corresponde no IPA é de facto o [i]. Veloso (1999) refere ainda, que na ausência de estudos recentes e seguros, apoiados em técnicas experimentais que contraponham essa utilização, adopta também o símbolo [i] para transcrever a referida vogal. Mateus et al. (2005) utilizam o símbolo [i] para caracterizar uma vogal oral não acentuada.

No entanto, a classificação de Cruz-Ferreira (1995) citada por Veloso (1999) e Cruz-Ferreira (1999) que consta no Handbook of The International Phonetic Association (IPA) para o PE, considera que esta vogal é a contrapartida não arredondada de um [u] ligeiramente avançado e rebaixado, sendo transcrita como [ɯ]. Esta vogal só ocorre em sílabas não acentuadas e segundo a autora não corresponde à qualidade média-central associada ao *schwa*.

Vogais nasais do PE

Além das vogais orais, existem também no PE as vogais nasais, produzidas quando o velo se encontra numa posição baixa, permitindo a passagem de ar pelas cavidades nasais e causando ressonâncias nasais. Assim, vulgarmente consideram-se, para o PE, cinco vogais nasais: o [ɨ̃] (**canto** ou **manto**), [ɛ̃] (**pen**te), [ĩ] (**min**to ou **pin**ta), [õ] **pon**te e [ũ] de **pun**to (Mateus, Falé, & Freitas, 2005; Teixeira, 2000).

	ANTERIOR OU PALATAL	CENTRAL	POSTERIOR OU VELAR
Alta	[ĩ]		[ũ]
Média	[ɛ̃]	[ɨ̃]	[õ]
Baixa			

Tabela 2. 2 - Classificação das vogais nasais do PE. Fonte: Mateus et al. (2005).

As vogais nasais podem ocorrer como alofones ou contrastando fonologicamente com as vogais orais. O primeiro caso ocorre em praticamente todas as línguas, enquanto que no segundo caso, contraste fonológico, só ocorre em algumas línguas como o Polaco, Hindi e Francês. Nem todas as línguas têm vogais nasais, menos de um quarto das línguas apresentam estes sons, destas nenhuma tem mais vogais nasais que orais e só em metade o número de vogais nasais é igual ao das orais (Teixeira, 2000).

Como para o PE, o número de vogais nasais é menor do que as orais, pensa-se que a vogal [ɨ̃] possa ser produzida por configurações correspondentes a [a] e [ɨ], a vogal [ɛ̃] por configurações entre [e] e [E], e o [õ] entre o [o] e [O]. Na realidade, um estudo de Teixeira e colaboradores (Teixeira et al, 2003b), revelou que em termos da altura da vogal e, utilizando F1 como medida da altura, para todas as regiões consideradas no estudo, para os vários informantes e nos diferentes contextos estudados, a altura de [ɨ̃] encontra-se entre a de [a] e [ɨ], a de [ɛ̃] entre [E] e [e] e a de [õ] entre [O] e [o]. Nas vogais nasais [ũ] e [ĩ], a altura da vogal é semelhante à das correspondentes vogais orais. Os autores chegaram também à conclusão que: (1) durante a produção das vogais nasais há variação ao nível da cavidade oral, (2) a altura das vogais [ɨ̃], [ɛ̃] e [õ] apresenta variação regional e (3) a percepção da altura da vogal é influenciada pela nasalidade (Teixeira et al, 2003b). No entanto, muitas dúvidas ainda subsistem relativamente à produção e características dos sons nasais, em particular os do PE.

Semivogais

As semivogais ou glides possuem características articatórias semelhantes às das vogais, mas acusticamente distinguem-se destas na medida em que são produzidas com menor energia e intensidade. Não podem constituir núcleo de sílaba, nunca ocorrendo sozinhas e não podem receber acento. As semivogais estão sempre acompanhadas por uma vogal com a qual formam os

ditongos. No PE existem duas semivogais o [j] (**pai**) e o [w] (**pau**) que correspondem do ponto de vista articulatorio, ao [i] e ao [u] (Mateus, Falé, & Freitas, 2005; Teixeira, 2000).

Ditongos e tritongos orais e nasais do PE

Ditongos Orais

Quando as semivogais ou glides G, se associam às vogais formam um ditongo VG ou GV. Os ditongos podem ser crescentes ou decrescentes. Em Português, considera-se que os verdadeiros ditongos são os ditongos decrescentes. Como exemplos de ditongos decrescentes, ou seja, vogal seguida de glide, poderemos ter [aw] como em **mau**, ou [oj] como em **boi**. Os ditongos crescentes, apesar de, geralmente aceites na Fonética, fonologicamente correspondem a uma sequência de duas vogais VV, que foneticamente se realizam como VV ou GV, em função da taxa de elocução. Como exemplos de ditongos crescentes, poderemos ter, por exemplo [ja] como em **hiato**, ou [jO] como em **viola** (Moutinho, 2000).

Ditongos Nasais

Existem também os ditongos nasais crescentes e decrescentes que resultam também da combinação de uma glide com uma vogal nasal. Exemplos de ditongos nasais crescentes são [w6~] como em **quando** ou [wi~] como em **ruim**. Ditongos nasais decrescentes são, por exemplo, [6~j~] **mãe** ou [6~w~] em **pão** (Moutinho, 2000; Teixeira, 2000).

Tritongos orais e nasais

São encontros vocálicos constituídos por glide-vogal-glide, um exemplo de um tritongo oral é [wow] (**averiguou**) (Teixeira, 2000) e um tritongo nasal é por exemplo [w6~w] como em **saguão** (Laver, 1995).

2.3.1.2 Consoantes

As consoantes ou sons consonânticos são produzidos com uma constrição ou obstrução significativa à passagem do fluxo de ar no tracto vocal. Esta obstrução pode ser total, impedindo a passagem do ar momentaneamente, ou parcial, dando origem à produção de ruído. As consoantes costumam ser classificadas, do ponto de vista articulatorio, por: ponto de articulação, modo de articulação, abertura ou fecho do véu palatino e pela vibração ou não das pregas vocais (Mateus, Falé, & Freitas, 2005; Moutinho, 2000).

Ponto de articulação

Refere-se à localização do ponto de maior constrição imposta pelos articuladores na cavidade oral. Assim, o ponto de articulação⁴ poderá ser bilabial, lábio-dental, dental, alveolar, palatal, velar e uvular (Mateus, Falé, & Freitas, 2005). Moutinho (2000) considera ainda um ponto de articulação pré-palatal⁵ (consoantes fricativas [S] e [Z]) e não considera a úvula como ponto de articulação nos sons do PE. A Tabela 2.3⁶ resume os vários pontos de articulação, os articuladores envolvidos, movimentos dos articuladores e sons do PE assim produzidos.

⁴ Alguns autores portugueses não referem os pontos de articulação faríngeo e glotal, uma vez que eles não existem para os sons do PE, existindo no entanto para outras línguas.

Relativamente à terminologia referente aos pontos de articulação, alguns autores (Teixeira, 2000; Veloso, 1999) fazem sempre referência aos dois articuladores envolvidos na produção da consoante (por exemplo, apico-dental, apico-alveolar etc). Outros autores utilizam apenas, por norma, o nome do articulador superior para indicar o local da constrição e utilizam o nome dos dois articuladores apenas quando possa existir ambiguidade. Segundo Andrade & Viana (1996) algumas constrições não podem simplesmente ocorrer, pelo que não haverá dúvidas nem ambiguidades.

⁵ Na literatura Inglesa, nomeadamente em Ladefoged (2001) e no Handbook of the IPA (1999), utiliza-se o termo pós-alveolar e não pré-palatal, utilizado por alguns autores Portugueses. No entanto, estes dois pontos de articulação não representam exactamente o mesmo local. Em Laver (1995, pp 136) podemos encontrar uma descrição detalhada de todos estes pontos. Assim, para articulações efectuadas entre a zona alveolar e a zona palato-alveolar deve ser utilizado o termo pós-alveolar e para articulações entre a zona alvéolo-palatal e o palato dever-se-á utilizar o termo pré-palatal. A zona alvéolo-palatal é, segundo o autor, a zona articulatória entre a zona palato-alveolar e a zona palatal. Como se pode depreender, para se descrever com rigor o ponto de articulação de uma consoante, será necessário recorrer a técnicas exploratórias como a EPG, que permitem ver o contacto língua-palato.

⁶ Esta tabela foi efectuada com base nas autoras Moutinho (2000) e Mateus et al (2005), existem no entanto outras classificações de autores Portugueses que divergem essencialmente a nível dos pontos de articulação dental, alveolar, pré-palatal e palatal. Assim, para Cruz-Ferreira (1999) no Handbook of the IPA, são classificadas como dentais as consoantes [t, d, n, l], e como alveolares [s, z, r], [S, Z] surgem como palato-alveolares. Nos restantes sons consonânticos, não há diferenças significativas relativamente à classificação aqui apresentada.

Andrade & Viana (1996) classificam à semelhança de Mateus et al (2005) o [R] como uvular e não velar. Estas autoras também designam por sons coronais todas as consoantes que são produzidas na região dento-alveolar e pós alveolar tendo como articulador activo a coroa da língua. Assim, todas as dentais, alveolares e pós alveolares e ainda todas as palatais do PE são, segundo as autoras, consoantes coronais.

Veloso (1999) desdobra as consoantes alveolares em dois grupos: (1) as apico-alveolares ou apico-dentais [t d n l r] e (2) as pré - dorso alveolares [s] e [z].

Esta diferença de classificação encontrada na bibliografia pode dever-se, em nosso entender a vários factores: (1) à ausência de informação proveniente de métodos exploratórios como a EPG, (2) a diferenças entre os próprios falantes nos diversos estudos e (3) a variantes dialectais.

A opção por Moutinho (2000) e Mateus et al (2005) não reflecte a nossa tomada de posição relativa a esta questão, pois não sendo especialistas na área não seria correcto fazê-lo. Fizemos apenas uma escolha, pois seria difícil que um quadro resumo apresentasse toda a diversidade encontrada.

PONTO DE ARTICULAÇÃO	ARTICULADOR ACTIVO	ARTICULADOR PASSIVO	MOVIMENTOS	SONS
Bilabial	Lábio superior e inferior		Aproximação e fecho dos lábios	[p], [b] e [m]
Lábio-dental	Lábio com apoio do maxilar inferior	Arcada dentária superior (porção anterior) - SAADS	Toque do lábio inferior no SAADS, causado por movimento de aproximação do lábio (apoiado pelo maxilar inferior)	[v] e [f]
Dental	Coroa da língua	SAADS	Aproximação e toque da coroa da língua no SAADS	[t], [d], [z] e [s]
Alveolar	Coroa da língua	Alvéolos dentários superiores	Aproximação e/ou toque da coroa da língua nos alvéolos superiores	[n], [l] e [r]
Pré palatal	Coroa ou dorso da língua	Na zona imediatamente a seguir à transição alveolo-palatal mas antes do palato duro		[ʒ] e [s]
Palatal	Dorso da língua	Palato duro	Aproximação e/ou toque do dorso da língua no palato duro	[j] e [l]
Velar	Dorso da língua	Palato mole	Aproximação e/ou toque do dorso da língua no véu do palato	[k] e [g]
Uvular	Dorso da língua	Úvula	Vibração da zona posterior do dorso da língua junto à úvula	[R]

Tabela 2. 3 - Ponto de articulação, articuladores, movimentos dos articuladores na produção das consoantes do PE padrão. Adaptado de Moutinho (2000) e Mateus et al. (2005).

Abertura/fecho do véu palatino

Permite dividir os sons em nasais e não nasais ou orais com passagem do fluxo de ar pelas cavidades nasais.

Vibração ou não das pregas vocais

Permite distinguir os sons em vozeados ou sonoros e em não vozeados ou surdos (Mateus, Falé, & Freitas, 2005; Moutinho, 2000).

Modo de articulação

Descreve o tipo de perturbação que é induzida à passagem do fluxo de ar pelo tracto vocal. Relativamente a esta característica, podem dividir-se os sons em oclusivas, nasais, fricativas⁷, laterais e vibrantes. De acordo com Teixeira (2000), as consoantes nasais como apresentam propriedades acústicas e modo de produção significativamente diferente dos sons oclusivos devem ser apresentados separadamente. Outros autores (Andrade & Viana, 1996; Mateus, Falé, & Freitas, 2005) apresentam as nasais como um caso particular das oclusivas. Tendo em conta o modo de articulação podem considerar-se as seguintes classes de sons:

2.3.1.3. Classificação das consoantes pelo modo de articulação

Oclusivas

Nas oclusivas (*stops* ou *plosives* no inglês) existe uma obstrução total à passagem do ar pela cavidade oral, encontrando-se também impedida a passagem do ar pelas cavidades nasais (no caso das oclusivas orais). Podem ser produzidas com ou sem vibração das pregas vocais, designando-se oclusivas vozeadas e não vozeadas, respectivamente. Nas oclusivas, e em termos acústicos, podem ser consideradas fases distintas durante a sua produção, nomeadamente a fase da oclusão a qual é seguida de libertação da pressão acumulada (a fase de explosão ou *release* em Inglês) e a fase de transição para a vogal.

No PE, existem as oclusivas vozeadas [b] de (buda), [d] de **da**do, e [g] de **ga**to e as não vozeadas [p] de **po**te, [t] de (to)lo, [k] de **ca**ro. Em relação ao ponto de articulação o [b] e o [p] são bi-labiais, o [d] e o [t] são dentais ou alveolares e o [g] e [k] são velares.

Nasais

Quando o velo se encontra abaixado com acoplamento da cavidade oral e nasal, mas existe obstrução na cavidade oral, são produzidas as consoantes nasais ou simplesmente nasais. No PE, existem as consoantes [m] como em **ca**ma, o [n] como em **ca**na e o [ɲ] como em **ca**nha. Ao contrário das oclusivas orais, todas as oclusivas nasais são produzidos com vozeamento (Teixeira, 2000). O [m] é bilabial produzido com os lábios fechados, o [n] é alveolar e, neste caso, é a ponta

⁷ Existem também as africadas, que são sons produzidos com uma obstrução completa do tracto vocal seguida de uma constricção com turbulência característica das fricativas. Correspondem aos sons [tʃ] e [dʒ] (Jesus, 1997). O [tʃ] encontra-se na região Norte e Centro Interior como por exemplo na pronúncia [tʃa]lves de Chaves (Andrade & Viana, 1996)

da língua que toca a região alveolar tocando a porção posterior da língua nos molares superiores, e o [J] é palatal produzido com o dorso da língua a tocar no palato (Jesus, 1997).

Fricativas

As fricativas são sons produzidos com uma constrição do tracto vocal, causando turbulência e consequente produção de ruído. Este ruído propaga-se ao longo do tracto vocal. Usualmente, as fricativas distinguem-se através do ponto de constrição e da presença ou ausência de vozeamento (as fricativas vozeadas apresentam também uma fonte de excitação glotal, devido à vibração das pregas vocais). Em termos acústicos, as fricativas caracterizam-se por uma grande energia contida nas altas frequências, podendo deixar de ser percebidas quando são filtradas por um filtro passa-baixo (Ribeiro, 2003). No PE existem seis fricativas, três vozeadas [v] (**v**aca), [z] (**z**arpa), [ʒ] (**j**acto) e três não vozeadas [f] (**f**ala), [s] (**s**ala) e [ʃ] (**ch**á). Em relação ao ponto de constrição, os sons [v] e [f] são lábio-dentais uma vez que é o lábio inferior que se aproxima dos incisivos superiores. O [s] e o [z] são apico-dentais (ou só dentais) pois a coroa da língua aproxima-se da região dento-alveolar. Finalmente o [ʃ] e o [ʒ] são produzidos com a língua na região palato-alveolar ou pré-palatal (Jesus, 1997; Moutinho, 2000; Ribeiro, 2003; Teixeira, 2000).

Laterais

As consoantes laterais são produzidas com uma obstrução parcial do fluxo de ar provocado pelo contacto da língua com o palato ou com os alvéolos, passando o ar lateralmente (Mateus, Falé, & Freitas, 2005; Ribeiro, 2003; Teixeira, 2000). No PE temos as laterais: [l] como em (**l**aço), e [ʎ] como em **fol**ha e uma variante, o [ʎ~] como em **mal**. O [ʎ~] chama-se velarizado⁸ e “segundo a posição dominante entre os autores portugueses nas décadas de 80 e 90”, ocorre apenas em final de sílaba ou em posição final absoluta (Andrade, 1999; Mateus, Andrade, Viana, & Villalva, 1990). No entanto, Andrade (1997, 1999) refere que num estudo efectuado em três falantes do PE (da zona de Lisboa), houve evidência de que poderia ocorrer velarização também em posição inicial de sílaba, tal como referido em descrições fonéticas mais antigas.

Em termos de ponto de articulação, o [l] é alveolar, produzido com a ponta da língua na zona alveolar, deixando aberturas laterais. A lateral [ʎ], é palatal produzida com uma oclusão central entre o pré-dorso e o palato anterior (Andrade & Viana, 1996; Mateus, Falé, & Freitas, 2005; Teixeira 2000).

Para o estudo das laterais, em particular, é importante obter informação lateral (fora do plano médio sagital), uma vez que a sua produção envolve mecanismos que ocorrem fora da linha média. Informação (3D), informação em tempo real (RM ou Ultrasonografia) e EPG serão as mais indicadas para melhor estudar e compreender a produção destes sons.

⁸ A atribuição do nome velarizado a esta lateral apical, resulta do facto de estar descrita uma segunda constrição posterior formada pelo dorso da língua no velo (Mateus, Andrade, Viana, & Villalva, 1990; Teixeira, 2000).

Vibrantes

As vibrantes são sons que são articulados com uma constrição parcial que provoca vibração de um articulador - a língua (Andrade & Viana, 1996; Mateus, Falé, & Freitas, 2005; Moutinho, 2000; Teixeira, 2000). As vibrantes podem apresentar um só batimento (uma só vibração do articulador) e nesse caso chamam-se vibrantes simples (*taps*) ou dois ou mais batimentos chamando-se vibrantes múltiplas ou *trills* (Laver, 1995; Teixeira, 2000).

No PE, existem três sons deste tipo: o [r], [r] e o [R]. Os dois primeiros são produzidos com vibração da ponta da língua (coroa ou lâmina) na região anterior ou frontal e o último é produzido na região posterior com a vibração do dorso da língua junto à úvula (Andrade & Viana, 1996; Teixeira, 2000). De salientar que no alfabeto SAMPA apenas nos aparecem dois símbolos para as vibrantes, o [r] como em caro e o [R] como em carro.

2.4 Nasalidade

Uma questão importante no estudo da Fonética, nomeadamente do PE, é a nasalidade. A nasalidade é produzida, como vimos atrás, quando há acoplamento entre as cavidades oral e nasal. Esse acoplamento é estabelecido quando há um abaixamento parcial do palato mole ou velo. No entanto, esta apresentação poderá ser demasiado simplificada. Vulgarmente efectuam-se algumas simplificações quando se descreve a nasalidade.

- 1) A primeira simplificação é a de que quando o véu do palato se encontra fechado o som não tem nasalidade e que quando o som é não nasal o véu do palato se encontra fechado;
- 2) A segunda simplificação resulta de se considerar que o fluxo nasal resulta sempre em nasalidade e que a nasalidade requer sempre fluxo nasal. Ou seja, o fluxo de ar pela cavidade nasal não é uma condição suficiente para a produção de nasalidade audível. A nasalidade é essencialmente uma condição de ressonância. Para que exista ressonância da cavidade nasal é necessário ter em atenção a proporção entre as aberturas oral e nasal. A proporção entre a abertura nasal e oral deverá ser de 1 para 11 nos sons não nasais e de 8,8 para 3,1 na voz nasal (Laver, 1980);
- 3) A terceira aproximação é a de que a ressonância da cavidade nasal é a única responsável pela produção da nasalidade, e inversamente que a nasalidade requer sempre ressonância nasal (Laver, 1980).

Num estudo recente com vogais nasais do Francês (Engwall, Delvaux & Metens, 2006) em que participaram quatro falantes, os autores verificaram que a anatomia das cavidades oral e nasal influencia a articulação, existindo variabilidade entre os sujeitos, quer a nível da forma do tracto vocal, quer de abertura do velo, de forma a estabelecer uma relação adequada entre os ressoadores orais e nasais. Estes autores de alguma forma confirmaram o que já tinha sido observado por Delvaux, Metens & Soquet (2002ab) e Delvaux (2003), ou seja, a oposição oral

/nasal deve-se não só à abertura e posição do velo mas também a um conjunto de alterações articatórias ao nível da configuração lingual e labial que alteram a proporção entre os tractos oral e nasal. A língua portuguesa, como já se referiu é rica em sons nasais (vogais nasais, consoantes nasais, ditongos e tritongos nasais), não existindo muita informação relativamente à sua produção. Há no entanto, algumas evidências, baseadas em estudos fonéticos, de que as vogais nasais do PE têm características próprias e são diferentes, por exemplo das vogais nasais do francês (Delattre, 1968; Rossato, Teixeira, & Ferreira, 2006).

No que diz respeito à síntese dos sons nasais e de acordo com Teixeira, Moutinho, & Coimbra (2001) e Teixeira & Vaz (2001), quando se utiliza uma abertura do velo adequada, as vogais são percebidas como nasais, no entanto a qualidade do som produzido não é natural. Este facto leva a que se perceba que a abertura do velo só por si, com conseqüente acoplamento das cavidades oral e nasal, não é suficiente para produzir um som nasal de qualidade natural. Assim, para melhorar a qualidade dos sons nasais sintetizados, é necessário fazer variar o velo no tempo produzindo uma vogal em três fases, ou seja, é necessário incluir a sua dinâmica (Teixeira, 2000; Teixeira, Moutinho, & Coimbra, 2001; Teixeira & Vaz, 2001). A obtenção de imagens em tempo real que permitam observar essa dinâmica do velo bem como a de outros articuladores é extremamente importante para o processo de síntese.

2.5 Coarticulação

A fala não é uma sequência de sons bem definidos, sendo a transição entre segmentos feita de forma gradual e não abruptamente (Mateus, Falé, & Freitas, 2005). Na realidade, durante a produção de fala os movimentos dos articuladores na produção dos diversos segmentos fonéticos (vogais, consoantes) sobrepõem-se no tempo e no espaço e interagem entre si. Como consequência, a configuração do tracto vocal para cada ponto, ao longo do tempo, é influenciada por mais que um segmento. Este efeito designa-se por coarticulação (Farnetani, 1999). O conceito de coarticulação foi pela primeira vez introduzido em 1933, por Menzerath e de Lacerda, mas só mais recentemente surgiram modelos e teorias que procuram clarificar este conceito. A utilização e acesso mais facilitado a técnicas experimentais (EPG, EMA, análise acústica, etc) em muito têm contribuído para um conhecimento mais completo dos fenómenos de coarticulação.

Na Tabela 2.4, identificam-se os principais articuladores envolvidos na coarticulação, os músculos responsáveis pelos movimentos articulatórios/coarticulatórios, os movimentos que geralmente se sobrepõem nos segmentos contínuos e as consequências acústicas daí resultantes (Farnetani, 1999).

	MIOMOTOR	ARTICULATÓRIO	ACÚSTICO
Lábios	<i>Orbicularis Oris</i> <i>Risorius</i>	Arredondamento dos lábios	Alterações de F1, F2 e F3
Língua	<i>Genioglossus</i> e outros músculos intrínsecos da língua	Movimento Frente /trás e alto/baixo	Alterações em F2, F1 e F3
Velo	Relaxamento do <i>Levator Palatini</i>	Abaixamento do velo	Formantes nasais e alteração formantes orais
Laringe	Cricoaritenoideu posterior/Interaritenoideu, Cricoaritenoideu lateral	Abdução e adução das cordas vocais	Duração acústica e periodicidade do sinal

Tabela 2. 4 - Coarticulação (Adaptado de Farnetani, 1999).

É frequente a coarticulação ser estudada em função dos diferentes articuladores envolvidos, assim fala-se em coarticulação velofaríngea, lingual, labial, mandibular e ainda do efeito conjunto da língua com a mandíbula. A coarticulação tem sido observada em todas as línguas já estudadas, podendo ser considerado um fenómeno universal. No entanto, existem diferenças muito significativas entre as várias línguas, não sendo ainda completamente claro como varia o padrão coarticulatório nas diversas línguas (Farnetani, 1999).

Em geral, durante a produção de fala podem ocorrer dois tipos de coarticulação: (1) coarticulação antecipatória (ou também referida como direita/esquerda (DE)) e (2) coarticulação *carryover* (esquerda/direita (ED)). A primeira só é possível se o falante conseguir olhar para a frente no tempo (*look ahead mode*) e antecipar sons que se seguem. Este tipo de coarticulação envolve um processamento de alto nível (fonológico - fonético) uma vez que um segmento é “visualizado” antes de ser articulado. A coarticulação do tipo ED reflecte a forma como a língua, mandíbula e lábios se movem do som precedente para o som seguinte e ocorre apenas a um nível fisiológico (Dang & Honda, 2004; Dang, Honda, & Perrier, 2004).

Existem algumas teorias e modelos que têm procurado explicar os fenómenos de coarticulação. A coarticulação pode ser vista como um fenómeno de economia do discurso, como “criadora da gramática da língua” e ainda como um fenómeno de coprodução. Uma explicação aprofundada destas teorias pode ser encontrada em Farnetani (1999) e em Farnetani & Recasens (1999). Dada

a complexidade do tema apenas nos referiremos aqui a alguns tópicos decorrentes desses modelos.

1) Economia do discurso

Coarticulação no contexto da teoria da variabilidade adaptada

Do núcleo desta teoria, desenvolvida por Lindblom, fazem parte os conceitos de que (1) o papel fundamental da fala é permitir uma comunicação bem sucedida e (2) que tal como outras funções biológicas o mecanismo envolvido na produção de fala tende a poupar esforços. Neste contexto, variabilidade adaptada significa que os falantes de uma língua são capazes de adaptar a produção às exigências da situação comunicacional, ou seja, podem, em função da situação hiperarticular ou hipoarticular como forma de economia. O ouvinte saberá, nesta última situação, fazer a descodificação da informação e interpretá-la de forma correcta (*top-down information*). Perceptualmente, quando se passa de um discurso hiperarticulado para um discurso hipoarticulado há uma diminuição do contraste fonético e do ponto de vista da articulação há um aumento coarticulatório gradual (Lindblom citado por Farnetani (1999) e Farnetani & Recasens (1999)).

2) Coarticulação como “criadora da gramática da língua” (Creature of language grammar)

Com base na teoria da fonologia generativa clássica (Chomsky & Halle (1968) *The Sound Pattern of English - SPE*) faz-se uma clara distinção dos conceitos de coarticulação e de assimilação. As variações coarticulatórias seriam, à luz desta teoria, originadas a partir das propriedades físicas da fala, sendo governadas por regras universais. Se assim fosse, e de acordo com Hammarberg (1976) citado por Farnetani (1999), não se poderiam explicar as diferenças no padrão de coarticulação encontradas para diferentes línguas. Também, se a coarticulação fosse apenas um processo fisiológico puro, imposto por limitações mecânicas e de inércia do aparelho fonador, estabelecer-se-ia uma marcada dicotomia entre intenção e execução, o que teria como consequência uma incapacidade dos articuladores em executar as ordens de uma forma específica (Daniloff e Hammarberg (1973) e Hammarberg (1976) citados por Farnetani (1999) e Farnetani & Recasens (1999)). A única forma de ultrapassar esta dicotomia seria, de acordo com estes autores, assumir que a coarticulação é, ela própria, parte da componente fonológica. A função da coarticulação seria suavizar/atenuar as diferenças entre segmentos adjacentes. À luz desta explicação, a chamada coarticulação antecipatória (DE) seria sempre um processo fonológico deliberado enquanto que a coarticulação (ED) poderia ser, em parte, explicada pela inércia dos órgãos envolvidos na fala. Uma das limitações deste modelo é a incapacidade de explicar, por exemplo, efeitos extensos de coarticulação (ED) observados em estudos do tipo V-V.

3) Coarticulação como coprodução

Um dos esforços mais significativos no sentido de se estabelecer a ponte entre os aspectos cognitivos e físicos da linguagem resulta de se considerar a coarticulação como um fenómeno de coprodução (Fowler citado por Farnetani (1999) e Farnetani & Recasens (1999)).

A teoria da coprodução está relacionada com os gestos e a sua organização temporal e espacial. A sobreposição entre gestos reflecte a coordenação temporal e é expressa no modelo pela chamada fase intergestual, ou seja, o início de um gesto ocorre numa fase específica do ciclo gestual precedente. Um aumento de sobreposição gestual faz com que a duração segmental diminua e os efeitos coarticulatórios aumentem. Alterações quantitativas na sobreposição intergestual podem contribuir para diferenças a nível coarticulatório em discurso lento e rápido. De acordo com a fonologia dos gestos, as diferenças encontradas entre as diversas línguas podem ser explicadas e resultam do facto de existirem diferenças nos gestos na produção dos sons nas diferentes línguas.

Modelo DAC - Degree of Articulatory Constraints

Enquadrado na teoria da coprodução, surge o modelo DAC, que se baseia no pressuposto de que, gestos articulatórios associados a segmentos consecutivos são coproduzidos e sobrepõem-se em diferentes graus em função das suas propriedades temporo-espaciais, de factores prosódicos e ainda da velocidade do discurso (Recasens, 2002). De acordo com este modelo, a sensibilidade coarticulatória das consoantes à influência das vogais adjacentes em contexto VCV (efeitos V-C) varia inversamente com a força do efeito consonântico (efeitos C-V) e com o grau de impedimento articulatório (DAC) para a consoante intervocálica. O modelo DAC funciona bastante bem na explicação da coarticulação lingual, em particular a nível do dorso da língua, em sequências VCV e CC.

Existe um valor DAC atribuído às consoantes (relativamente ao dorso da língua) em função do seu grau de impedimento. Assim, pode ser atribuído um DAC máximo com o valor três para consoantes alveolopalatais, palatais, velares e fricativas linguais e um valor mínimo (de um) para as bilabiais (Recasens, Pallarés & Fontdevila, 1997; Recasens, 1999; Recasens, 2002). Este modelo apesar de ter sido inicialmente utilizado na previsão coarticulatória ao nível da língua, poderá ser aplicado também ao estudo da coarticulação mandibular, tal como sugerido em Recasens (2004).

Resistência à coarticulação

Destes modelos, surge um conceito, explicado por vezes de forma diferente à luz das diferentes teorias, que está relacionado com a resistência à coarticulação. Esta noção de resistência de alguns sons a efeitos de coarticulação foi proposta por Bladon e Al-Bamerni em 1976 para explicar diferentes graus de coarticulação espacial num estudo envolvendo os alofones velarizado e não

velarizado do /l/ (*clear* e *dark* do /l/ em inglês) (Bladon e Al-Bamerni citados por Farnetani & Recasens, 1999; Recasens, 1999).

A resistência à coarticulação pode ser definida como o grau com que um segmento resiste à influência de um segmento vizinho. Graetzer (2006) refere que a resistência à coarticulação pode ser encontrada em três contextos diferentes; (1) verifica-se quando a resistência possa evitar confusão entre contrastes paradigmáticos by *heightening phonetic clarity*, (2) quando a força articulatória dos segmentos (*strengthening of segments*) é induzida pela prosódia ou pragmática (sendo essa força articulatória definida como: “ an increase in the spacio-temporal magnitude of gestures” (Cho, 2004 citada por Graetzer, 2006) e (3) quando os segmentos são, do ponto de vista articulatório inerentemente fortes.

Este último contexto reflecte a posição de Recasens (1985), segundo o qual a resistência à coarticulação se correlaciona positivamente com (1) elevado grau de contacto entre a porção superior do dorso da língua e o palato na produção dos diferentes sons, (2) a formação de um duplo ponto de articulação como na produção de consoantes palato-alveolares e (3) grau de acoplamento entre ponta e corpo da língua. Desta maneira, a resistência à coarticulação pode variar em função da forma como um ou mais segmentos adjacentes partilham os articuladores.

Como se constatou, a coarticulação é um tema difícil e ainda não completamente explicado, no entanto, no que os autores estão de acordo é que qualquer tentativa para sintetizar fala próxima da fala natural deverá ter em atenção os fenómenos de coarticulação (Dang, Honda, & Perrier, 2004; Jesus, 1997).

2.6 Síntese de fala

A síntese de fala traduz-se na capacidade de uma máquina produzir fala de uma forma artificial. O seu principal objectivo é permitir o estabelecimento de comunicação entre um ser humano e uma máquina. Idealmente, a fala produzida de forma artificial deverá ser tão próxima quanto possível da fala natural. A síntese de fala é, por isso, uma tarefa difícil e que exige conhecimentos de várias áreas científicas (Ribeiro, 2003; Simões, 1999; Teixeira, 2003).

A vontade de criar voz humana de forma artificial não é recente, remontando as primeiras tentativas aos finais do século XVIII (Kratzenstein -1179, von Kempelen -1791 citados por Ribeiro (2003). No entanto, o aparecimento da teoria acústica de produção de fala nos anos 60 e, posteriormente, os grandes desenvolvimentos a nível computacional e tecnológico em muito contribuíram para uma maior investigação nesta área, e com conseqüente melhoria nos sistemas de síntese de fala.

2.6.1 Sistemas de síntese de fala

Existem várias técnicas para se conseguir sintetizar fala:

1. utilização de mensagens digitais previamente gravadas
2. concatenação de palavras e frases guardadas, a partir das quais se geram novas frases, (que não foram previamente pronunciadas)
3. conversão de texto em fala - vulgarmente conhecida, no inglês, por Text-to-Speech (TTS)

O primeiro sistema é o mais simples para produzir um sinal de fala, uma vez que o sistema se limita a reproduzir mensagens pré gravadas. Nestes sistemas, a qualidade do sinal de voz é boa, e o tempo de resposta é curto. É, no entanto, pouco flexível e é necessário um espaço de memória considerável. Geralmente utilizado para aceder a dados bancários via telefone.

2.6.2 Sistemas conversão Texto-Fala (TTS)

Os sistemas TTS são, hoje em dia, os mais divulgados na comunidade científica. Estes sistemas permitem a produção de qualquer mensagem a partir de texto. Neste processo de síntese estão envolvidas duas fases distintas: (1) fase de processamento linguístico e (2) fase de geração do sinal de voz.

A fase de processamento linguístico tem como objectivo obter uma representação fonológica do texto de entrada. Esta fase engloba uma etapa de pré-processamento do texto, que passa pela normalização do texto para que possa ser manipulado nas etapas seguintes. Após esta etapa, segue-se um módulo de transcrição ortográfico-fonética e, posteriormente, uma etapa de processamento prosódico. A naturalidade da fala nestes sistemas depende em muito desta fase do processo de síntese (Simões, 1999; Teixeira, 2003).

A fase de geração de sinal, ou seja, o processo de síntese do sinal pode ser efectuado de formas diferentes:

- (1) síntese de formantes ou síntese baseada em regras
- (2) síntese por concatenação
- (3) síntese articulatória

Cada um dos métodos utiliza de forma diferente a informação gerada nas etapas anteriores do processo de conversão texto-fala. A qualidade do sinal gerado é diferente, existindo vantagens e desvantagens em cada um deles. A síntese concatenativa tem sido a mais utilizada e com bastante sucesso, produzindo sinal de fala de qualidade, mas o sistema que parece ser mais promissor é o de síntese articulatória (Simões, 1999; Teixeira, 2003). Uma vez que o nosso trabalho está de alguma forma relacionado com a síntese articulatória só esta será brevemente abordada.

2.6.2.1 Síntese articulatória

A síntese articulatória procura gerar um sinal de fala o mais realista possível e de melhor qualidade e, para isso, procura que o processo de síntese se aproxime do processo natural da produção de fala. Assim, este tipo de síntese baseia-se na construção de um modelo físico fundamentado na descrição detalhada da dinâmica dos diversos articuladores durante a produção de fala. Os parâmetros típicos, utilizados na síntese, são a posição e movimento dos articuladores nomeadamente da língua, dos lábios, da mandíbula e do velo (Simões, 1999; Teixeira, 2003c). Para implementar um sintetizador articulatório é necessário que se construa um modelo matemático do tracto vocal.

Estes sintetizadores incluem, regra geral, dois subsistemas: um modelo anatomo-fisiológico das estruturas envolvidas na produção de fala e um modelo de produção e propagação do som nessas estruturas. O primeiro modelo permite que se transforme a posição dos diversos articuladores durante a produção de sons em áreas transversais do tracto vocal. O segundo modelo consiste num conjunto de equações que descrevem as propriedades acústicas do tracto vocal. Outros parâmetros têm que ser adicionados para que se consiga modelar correctamente a produção de fala. Uma descrição mais aprofundada destes sistemas poderá ser encontrada em Teixeira, 2000; Teixeira & Vaz, 2001, Teixeira et al, 2004a e Teixeira et al, 2005. Como exemplos deste tipo de sintetizadores podemos referir o sistema CASY (Configurable Articulatory Synthesizer) desenvolvido nos laboratórios Haskins (EUA), o ICP (Institute de la Communication Parléé) Virtual Talking Head, o SAPwindows da Universidade de Aveiro, entre outros (Teixeira et al, 2005).

Comparativamente a outros sistemas de síntese, a síntese articulatória é a mais recente e pouca atenção lhe tem sido atribuída. No entanto, alguns autores (Shadle & Damper citados por Teixeira, et al, 2005) consideram a síntese articulatória como a técnica de síntese mais promissora e com mais futuro. A grande vantagem deste tipo de síntese é a possibilidade de criar fala sintetizada de elevada qualidade. Como desvantagens deste método apontam-se: (1) dificuldades na obtenção de informação relativa ao tracto vocal e às pregas vocais durante a produção de fala, (a maior parte da informação obtida é referente a configurações estáticas, sendo a informação relativa à dinâmica dos articuladores ainda muito incompleta); (2) inexistência de um processo de inversão completo para obter os parâmetros articulatórios a partir de fala natural e (3) complexidade dos cálculos envolvidos (Teixeira et al, 2005).

2.6.2.2 Síntese articulatória para o PE

Sintetizador articulatório da UA

Na Universidade de Aveiro, existe já um sintetizador de fala baseado em síntese articulatória - o seu nome é SAP Windows (Sintetizador Articulatório do Português para o Windows) e tem permitido alguma investigação na síntese de alguns sons do PE (desde 1995), nomeadamente na síntese das vogais orais e nasais, consoantes nasais e mais recentemente de fricativas (Teixeira et al, 2005).

Os módulos e etapas do processo de síntese podem ser observados de forma esquemática na Figura 2.10.

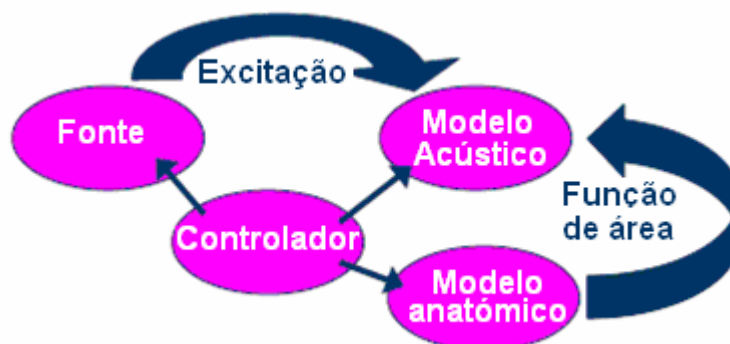


Figura 2. 10 - Módulos do sintetizador desenvolvido na UA, utilizando síntese articulatória. Fonte: Teixeira et al. (2004a).

1) Modelo anatómico

O modelo anatómico utilizado no SAP Windows é uma versão adaptada do sintetizador existente na Universidade da Florida que, por sua vez, é baseado no modelo de Mermelstein (1973) (Figura 2.11). Os articuladores são o corpo da língua, a ponta da língua ou coroa, a mandíbula, os lábios, o velo e o hióide. Neste modelo, assume-se que há simetria médio-sagital para estimar a área transversal. As áreas transversas não são determinadas de forma directa. Utiliza uma grelha não regular para calcular as áreas das diversas secções e o comprimento do tracto vocal (Figura 2.12).

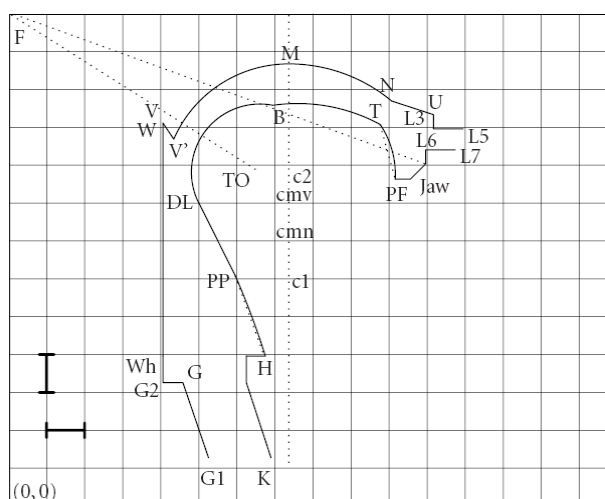


Figura 2.11 - Tracto vocal baseado no modelo de Mermelstein. Fonte: Teixeira et al. (2005).

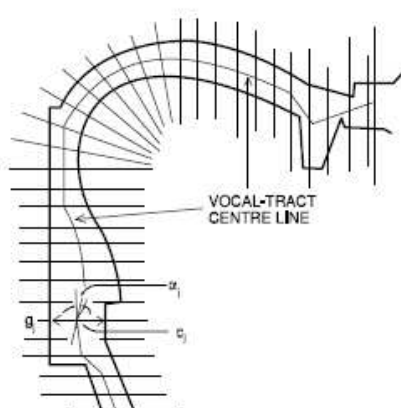


Figura 2.12 - Grelha não regular utilizada para cálculo das áreas. Fonte: Teixeira (2000).

O modelo do tracto nasal (Figura 2.13) é utilizado quando se pretendem sintetizar sons nasais. A cavidade nasal é modelada de forma semelhante à cavidade oral podendo ser considerada um ramo lateral do tracto vocal. Na cavidade nasal, a função de área é fixa para a maior parte do tracto, com excepção da zona do velo, que varia com o grau de acoplamento. O sintetizador permite ainda a inclusão dos seios peri-nasais no modelo.

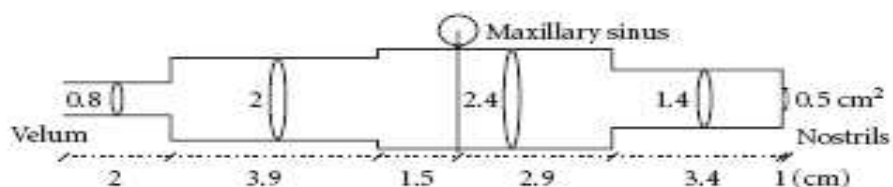


Figura 2.13 - Modelação do tracto nasal. Fonte: Teixeira et al. (2005).

2) Modelo acústico

O modelo acústico utilizado baseia-se na técnica de simulação da propagação do som no domínio da frequência. Foram feitas algumas aproximações: (1) assume-se que a propagação é planar, (2) que o tracto é um tubo e (3) que o tubo é aproximado pela concatenação de tubos acústicos elementares com área constante. Um circuito equivalente, representado por uma matriz de transmissão, modela cada um dos tubos elementares (Teixeira et al, 2004a; Teixeira et al, 2005).

3) Modelo de fonte glotal interactivo

Foi desenvolvido um modelo de excitação glotal que inclui a interacção fonte-tracto, para os sons orais e nasais e que permite o controlo directo dos parâmetros da fonte como, por exemplo, a Frequência Fundamental (F0). Para a modelação de diferentes sons há a necessidade de juntar outros módulos. Por exemplo, para sintetizar fricativas é necessário incluir um modelo acústico que simule o ruído característico na produção desses sons.

Na Figura 2.14 pode observar-se em (a) a definição de parâmetros para obter a palavra António [6~tOniu] e em (b) a síntese da palavra.

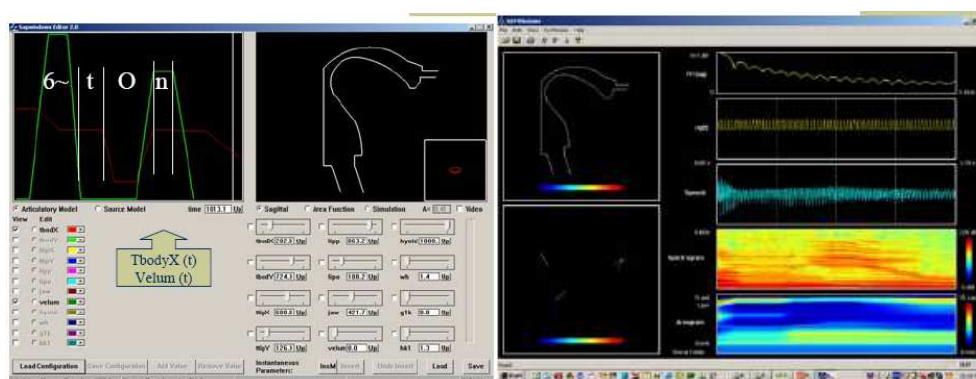


Figura 2.14 - Definição de parâmetros (esquerda) e síntese (direita) de [6~tOniu]. Fonte: Teixeira et al.(2004a).

2.7 Sumário

Procurou fazer-se a clarificação de alguns conceitos importantes na área da Fonética e relacionados com a produção dos sons do PE. Como ficou patente, existem algumas diferenças na terminologia utilizada pelos diferentes autores consultados, quer a nível de descrição anatómica, quer na própria classificação articulatória dos sons. Seria fundamental procurar uma uniformização, uma vez que, quem não tenha uma formação aprofundada na área tem algumas dificuldades quando pretende aumentar os seus conhecimentos na área.

Também se focaram ao longo capítulo algumas áreas em que o nosso conhecimento ao nível da produção dos sons, em particular do PE, é ainda incompleto. Exemplos, são: a existência ou não do *schwa* para o PE e qual a sua representação fonética, os mecanismos de produção das

laterais e vibrantes, a dinâmica do velo na produção dos sons nasais e também fenómenos de coarticulação no PE.

Capítulo 3

Técnicas de estudo em Fonética Articulatória

Como se descreveu no capítulo anterior, ao longo do tracto vocal, existem um conjunto de articuladores, que por terem características diferentes uns dos outros e apresentarem localização interna, tornam difícil o seu estudo. Os articuladores variam não só na sua forma, estrutura e localização mas também na dimensão e velocidade dos seus movimentos. A própria interacção dos articuladores entre si é um mecanismo de grande complexidade e de difícil avaliação. Em consequência, a avaliação de todos estes elementos através de uma só técnica ou instrumento será pouco eficaz e, o mais provável será um método avaliar perfeitamente um determinado articulador, mas ser completamente ineficaz no estudo de outro (Stone, 1999).

Assim, na aquisição de conhecimentos relacionados com os mecanismos de produção da fala e do tracto vocal muitos têm sido os instrumentos e técnicas de observação utilizados: o sistema *X-Ray Microbeam*, a Articulografia Electromagnética, a Electropalatografia, o Optopalatógrafo, o sistema OPTOTRAK, entre muitos outros (Stone, 1999). Dentro destas ferramentas de aquisição de informação estão também incluídos vários métodos de imagem como a radiografia simples, a cineradiografia e, mais recentemente, técnicas de imagem como a Tomografia Computorizada (TC), a Ultrasonografia e a Ressonância Magnética (RM) (Mohammad, 1999; Stone, 1999).

No presente capítulo, abordam-se alguns destes instrumentos de avaliação, referindo para cada um deles, o tipo de informação que podem fornecer relativamente ao tracto vocal, as vantagens e desvantagens de cada método e no caso das técnicas de imagem uma referência aos estudos que as utilizaram na área da fala. Neste capítulo, não será referida a RM, uma vez que será o tema dos capítulos quatro e cinco. Começaremos pelas técnicas de imagem, secção 3.1 e depois apresentaremos, algumas das outras técnicas laboratoriais (secção 3.2).

3.1 Técnicas de Imagem

3.1.1 Radiografia Simples e Cineradiografia

A radiografia simples, obtida em incidência lateral na zona do pescoço, foi o primeiro método de imagem a ser utilizado em estudos do tracto vocal e em investigação na área das ciências da fala. As imagens dessas radiografias permitiram obter uma representação completa do tracto vocal, no plano sagital, desde a glote até aos lábios (Chiba & Kajiyama, 1941; Fant, 1960). No princípio, a

informação proveniente dos raios X era estática mas, a partir dos anos cinquenta, com o aparecimento da cineradiografia foi possível obter informação dinâmica relativa ao tracto vocal. Na área da radiologia, a cineradiografia é utilizada essencialmente em estudos angiográficos e de intervenção, nomeadamente em cateterização cardíaca, onde é necessária, à semelhança do que acontece durante a produção de fala, uma elevada resolução temporal. Na cineradiografia o monitor de televisão é substituído por uma câmara que grava as imagens em película, sendo depois analisada em instrumentos de visualização adequados. É possível a utilização de películas de 16 e 35 mm, e obter cerca de 60 imagens por segundo. No entanto, quanto maior for o número de imagens por segundo, maior será também a dose de radiação a que o indivíduo é sujeito (Bushong, 2001).

A possibilidade de, com este método de imagem, se observarem os movimentos dos articuladores durante a fala abriu novas perspectivas na investigação dos mecanismos de produção da fala e como tal esta técnica foi amplamente utilizada durante os anos subsequentes, sendo produzidas bases de dados de indivíduos falantes para diversas línguas nomeadamente a língua francesa (Straka, 1965 citado por Arnal et al, 2000 e Rochete, 1973) e para a língua inglesa (Moll, 1960; Stevens & Öhman, 1963; Perkell, 1969). Os equipamentos foram modificados e adaptados às necessidades dos estudos de produção de fala, nomeadamente no que diz respeito à insonorização da sala para permitir gravação em simultâneo das produções dos falantes e para posicionamento adequado dos sujeitos. Uma descrição detalhada do equipamento e dos procedimentos seguidos para obtenção da base de dados obtida por Stevens & Öhman em (1963) pode ser encontrada em Perkell (1969), seguida de uma análise minuciosa de algumas produções referentes a palavras sem sentido, construídas com o intuito de avaliar o efeito contextual.

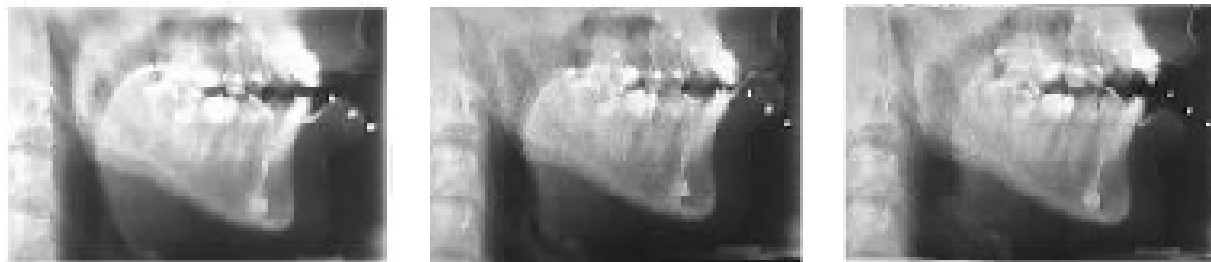


Figura 3. 1 - Exemplo de três *frames* de um filme cineradiográfico efectuado durante a produção de fala. Fonte: Munhall, Vatikiotis-Bateson, & Tohkura (1995) - X-Ray Film Database for Speech Research.

Apesar de existirem instrumentos (EPG e Ultrasonografia) que permitem obter uma resolução temporal superior à da cineradiografia, esta permite um excelente compromisso entre resolução temporal e espacial e a observação completa do tracto em toda a sua extensão, o que não é conseguido com nenhuma das outras duas técnicas (Arnal et al, 2000). No entanto, a grande desvantagem deste método é a utilização de radiações ionizantes, razão pela qual nos dias de hoje será pouco recomendável e mesmo pouco ético a sua utilização em estudos de produção de fala, uma vez que os informantes seriam sujeitos a doses de radiação consideráveis. Ainda assim, recentemente foram divulgados estudos (Ericsson, 2005) onde para além de imagens RM, foram

utilizadas imagens radiográficas dos falantes intervenientes no estudo, para comparação com as imagens de RM.

Uma vez que a informação proveniente das imagens radiográficas, em particular dos filmes cineradiográficos, é relevante, tem-se verificado uma conjugação de meios, de alguns grupos de investigação, na compilação e digitalização das bases de dados recolhidas no passado, com o intuito de as disponibilizar à comunidade científica. Munhall, Vatikiotis-Bateson & Tohkura, (1995) juntaram em videodisk cerca de 25 filmes que totalizam cerca de 55 minutos de produção, adquiridos por Rochete em 1973, no Canadá (Québec). Estes filmes são conhecidos como Université Laval Films ou simplesmente Cine-Laval. O restante material está identificado como M.I.T. film e corresponde à base de dados obtida por Stevens & Öhman em (1963) e analisada por Perkell (1969). Também o Instituto de Fonética de Estrasburgo, em conjunto com o Institute de la Communication Parlée (Grenoble) está a reunir informação referente a cerca de 40 filmes cineradiográficos relativos a estudos de produção de várias línguas, em conjunto com traçados sagitais do tracto vocal efectuados manualmente por conceituados foneticistas. A disponibilização desta informação, em formato digital, tem permitido desenvolver métodos de segmentação automáticos, para obtenção mais facilitada da informação, tal como efectuado por Thimm & Luetlin (1999).

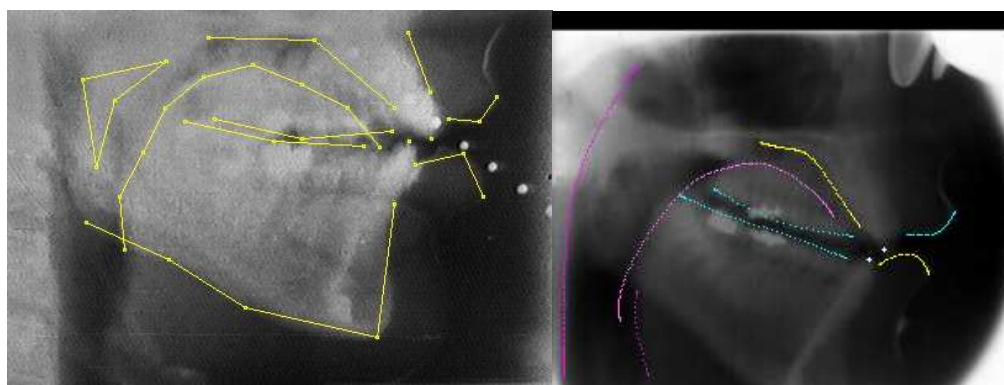


Figura 3. 2 - Utilização de bases de dados cineradiográficas para extracção de parâmetros articulatórios. Fonte <http://www.idiap.ch/pages/contenuTxt/Demos/demo6/xray.html>

3.1.2 Tomografia Computorizada

A tomografia axial computadorizada (TAC) ou simplesmente (TC) é uma técnica de imagem que recorre também à utilização de radiação X, mas utilizando um processo de aquisição diferente das técnicas projectivas. Esta técnica de imagem foi viabilizada por Hounsfield e Cormack em 1972 (Lima, 1995). O processo de aquisição consiste na obtenção de informação relativa a uma grandeza física, mediante a passagem de um feixe extremamente colimado, no plano axial através do corpo ou objecto a estudar. A finalidade da TC é obter os coeficientes de atenuação linear (μ),

em cada secção, a partir de um conjunto de projecções obtidas para diferentes ângulos em torno do paciente (Lima, 1995). Dos primeiros equipamentos de TC (sistemas de translação-rotação) em que o tempo de aquisição era da ordem dos minutos, até aos recentes equipamentos de TC multicorte (rotação contínua da ampola, com avanço contínuo da mesa) que permitem a aquisição de múltiplos cortes (até 64 cortes/s), com tempos de rotação da ordem dos 0,4s, distam 7 gerações de equipamentos e uma evolução tecnológica que permite, nos dias de hoje, utilização de matrizes superiores a 512x512 elementos, com voxel isotrópico, permitindo reformatações perfeitas em qualquer plano (Teixeira, 2003).

Não fosse a utilização de radiação ionizante, que representa claramente uma grande desvantagem, seria um método com muito potencial na obtenção de informação relativa ao tracto vocal. Quando comparada com a RM, permite obter uma resolução espacial superior, a nível de estruturas calcificadas como os dentes e palato duro. O facto de, o processo de aquisição, não ser ruidoso, ao contrário do que se passa na RM, permitiria a gravação simultânea do sinal áudio sem necessidade de recorrer a tecnologia específica e dispendiosa, como os microfones ópticos. O contraste ar/tecido é elevado, permitindo uma boa visualização dos articuladores, apesar da resolução de contraste, a nível dos tecidos moles, ser inferior à RM. Nos dias de hoje, com os mais recentes avanços tecnológicos é também possível a obtenção de imagens em tempo real (Fluoro TC) a uma razão de 5 frames/s, ou seja, equivalente ao conseguido com RM. Permite a obtenção de cortes ou volumes no plano axial, de forma directa, e também reconstruções perfeitas nos outros planos de forma relativamente rápida. O facto de, não permitir a obtenção de cortes sagitais directos, é uma limitação em relação à RM. Contudo, mais uma vez, reforçamos a ideia de que, ainda que com doses muito baixas, não será recomendável a sua utilização no âmbito da investigação em Ciências da Fala, uma vez que, o risco inerente à exposição às radiações ionizantes, não trará qualquer benefício ao individuo exposto.

Em termos de estudos de produção de fala a TC, foi utilizada pela primeira vez por Kiritani (1978) e Johansson & Sunberg (1983), não muito depois do início da sua aplicação na área da clínica, com o objectivo de obter as áreas transversas (3 a 4 cortes axiais) do tracto vocal na zona da faringe, durante a produção de algumas vogais da língua Sueca. Não tem sido tão amplamente utilizada como a RM, devido à exposição dos falantes às radiações ionizantes, mas alguns grupos de investigação têm recorrido à TC, a maior parte das vezes, para complemento das imagens obtidas por RM (Serrurier & Badin, 2005a). A Tomografia Computorizada por Feixe de Electrões (EBCT), que antes da divulgação das últimas gerações de TC multi-corte permitiam a obtenção de tempos de aquisição muito mais baixos e com menor dose no paciente, foi também já utilizada no âmbito destes estudos. Um dos estudos mais importantes recorrendo a EBCT, foi efectuado por Tom e al (1999). Story (1995) e Story, Titze & Hoffman (1996) referem a utilização de EBCT para adquirir as arcadas dentárias do falante envolvido no estudo.

3.1.3 Ultrasonografia

Os ultrasons foram utilizados pela primeira vez na área da medicina com finalidade terapêutica (produção de calor durante reabilitação - em 1920 - cf. Stone, 2005). Em termos de diagnóstico médico, foi pela primeira vez utilizada nos finais dos anos 30, por um neurologista, Theodore Dussik. Contudo, a ultrasonografia ou ecografia só se tornou popular a partir dos anos 60 com o desenvolvimento de equipamentos adequados. É uma técnica de imagem considerada inócua (se respeitadas os limites de segurança estabelecidos), não invasiva, fácil de executar, bem aceite e com custos relativamente baixos o que conduziu a uma ampla utilização nas mais diversas áreas: abdominal (avaliação de órgãos sólidos), ginecológica, obstétrica, senológica, cardíaca, vascular (eco-doppler), oftalmológica, entre outras. A principal desvantagem é a sua incapacidade em discriminar as estruturas ósseas e os artefactos derivados da interposição de estruturas com gás (Martins, 2003).

Na área das ciências da fala, a ultrasonografia tem sido utilizada nas últimas décadas, em particular, no estudo dos movimentos da língua, com sucesso considerável (Stone, 1999). A US permite velocidades de amostragem elevadas (30 a 60 Hz) para obtenção de informação 2D em tempo real ou, em alternativa, obter informação 3D estática. Por exemplo, Stone & Lundberg, (1996), utilizaram US para efectuar reconstrução 3D da superfície da língua, a partir de 6 cortes coronais.

A US permite velocidades de amostragem elevadas mas ainda insuficientes, por exemplo para estudar movimentos da ponta da língua em sons como oclusivas, *flaps* e cliques. Outras limitações da US neste contexto, são: a não visualização do palato mole e o facto de não permitir a observação do tracto vocal em toda a sua extensão.

Tem sido necessário, no entanto, adaptar esta técnica aos requisitos dos estudos nesta área. Assim, para obter imagens da língua, a sonda é colocada sob o queixo do informante sendo necessária uma correcta estabilização da sonda e da cabeça do informante, em relação ao feixe de US. Foram desenvolvidos e validados alguns métodos para estabilização, uns rígidos outros flexíveis. Em Stone (2005) pode ser encontrada uma revisão exaustiva destes métodos e os métodos de validação utilizados. No âmbito deste trabalho, e segundo a autora referenciada, referimos apenas um dos sistemas, que é um sistema de imobilização rígido - sistema HALTS (Head and Transducer Support System). Este sistema imobiliza completamente a cabeça e a sonda, sob o queixo, de tal forma que os movimentos obtidos para a língua representam o movimento combinado do sistema língua-mandíbula.

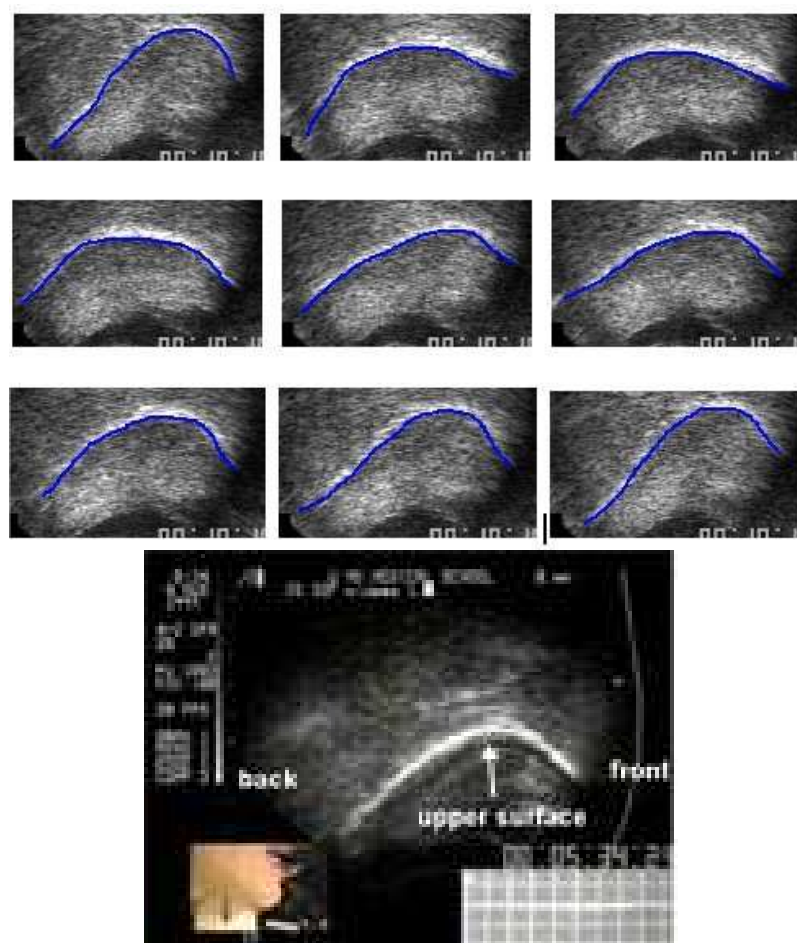


Figura 3. 3 - Dorso da língua obtido através de ultrasonografia (9 cortes), em cima e colocação da sonda sob o queixo do falante (em baixo). Fonte: <http://www.speech.umaryland.edu>.

3.2 Outras técnicas Laboratoriais

Apresentamos de seguida, outras técnicas ou métodos de estudo que têm sido utilizadas para estudar o tracto vocal. Surgem agrupadas apenas para as distinguir das técnicas de imagem descritas na secção anterior. Outros autores, apresentam as técnicas laboratoriais pelo tipo de informação que fornecem, por exemplo, informação 3D ou informação dinâmica (Engwall, 2002) ou métodos internos ou métodos externos (Mohammad, 1999), etc. A ordem com que surgem não é cronológica nem reflecte nenhuma hierarquia. Dá-se especial destaque às três técnicas mais utilizadas e referenciadas na literatura revista: *X-Ray Microbeam*, Articulografia Electromagnética (EMA) e Electroplatógrafia (EPG).

3.2.1 X-Ray Microbeam

Este sistema, específico para investigação na área das ciências da fala, foi desenvolvido na Universidade de Tóquio, com o intuito de minimizar as doses de radiação (mais baixas com este sistema do que com a radiografia simples e cineradiografia) recebidas pelos informantes nos estudos de produção de fala e por outro lado, evitar o gasto de tempo necessário para a análise da informação, proveniente das imagens cineradiográficas, efectuada imagem por imagem (Honda, 2002). As doses de radiação com este sistema são muito baixas, uma vez que são utilizadas voltagens muito elevadas (450 kV e 600 kV) e o feixe é extremamente colimado (0.4 mm espessura). Este facto é importante pois estruturas radiosensíveis como os olhos, não são irradiadas, e a radiação dispersa é mínima (Stone, 1999). Na Figura 3.4 (à esquerda) pode observar-se um sistema de segunda geração desenvolvido por Westbury (1994) na Universidade de Wisconsin (Westbury citado por Araújo, 2007) e à direita, planos de referência e localização dos *pellets* a nível da língua (Araújo, 2007).

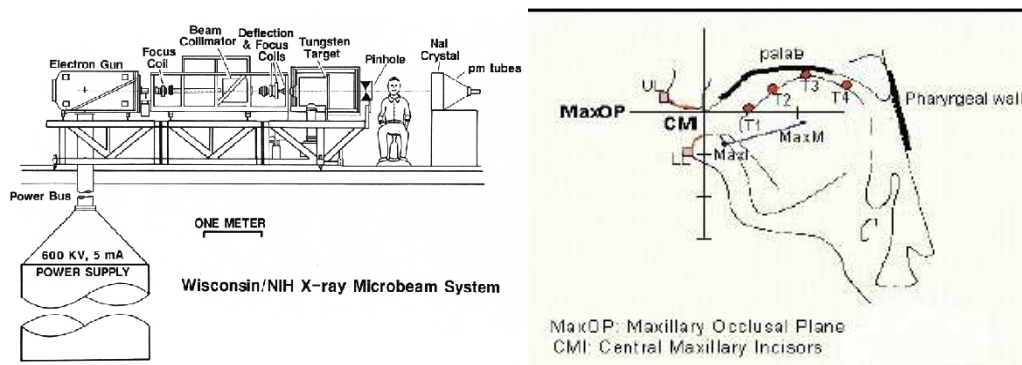


Figura 3. 4 - X-Ray Microbeam (gerador) à esquerda, Fonte: <http://www.medsch.wisc.edu/ubeam/gen/generator.html>. Informação extraída da base de dados XRMB, Fonte: Araújo (2007).

O *X-Ray Microbeam* (XRMB) permite velocidades de amostragem muito elevadas (40 a 180 amostras/ segundo) podendo esta razão de amostragem ser diferencial. É um excelente método para obter informação dinâmica relativa ao movimento dos articuladores e estudar fenómenos de coarticulação (Mohammad, 1999; Stone, 1999). Permite ainda, obter parâmetros cinemáticos como a velocidade, aceleração e inter-coordenação entre os articuladores. Como a dose de radiação é baixa permite obter grande quantidade de informação para um mesmo sujeito. No entanto, apesar da dose absorvida ser baixa, a utilização de radiação ionizante é considerada uma desvantagem. Os custos de instalação e manutenção são, também, significativamente elevados, sendo o número de unidades disponíveis muito escasso. Desta forma, uma investigação que recorra a esta ferramenta torna-se extremamente dispendiosa se uma equipa de investigação tiver que se deslocar para recolher informação (Stone, 1999). Permite apenas a obtenção de informação bidimensional (de alguns pontos no plano sagital) e a colocação dos eléctrodos na zona posterior da língua, velo e faringe é uma tarefa difícil e morosa, sendo mesmo impossível de colocar em alguns indivíduos (Mohammad, 1999; Stone, 1999)

3.2.2 Sistemas de Articulografia Electromagnética (EMA)

Esta técnica de avaliação do tracto vocal, é baseada no fenómeno da indução de Faraday: numa bobine, colocada sob acção de um campo magnético variável no tempo, é induzida uma força electromotriz. No sistema EMA, o campo magnético é criado quando se aplica uma corrente alternada numa bobine - chamada bobine transmissora ou emissora. A bobine onde é induzida a corrente denomina-se receptora. As bobines receptoras são colocadas nos diferentes articuladores (em pontos específicos como se mostrará adiante) no plano médio sagital. As bobines transmissoras são fixas, variando o seu número com o sistema utilizado, sendo cada uma delas sujeita a uma corrente alternada de diferente frequência e originando um campo magnético específico. A intensidade do campo magnético (sinal) na bobine receptora diminui à medida que aumenta a distância entre a bobine receptora e a transmissora (varia na razão inversa do cubo da distância) permitindo estudar os movimentos dos articuladores num sistema bidimensional. Idealmente seriam suficientes apenas duas bobines emissoras para determinar, por triangulação, as coordenadas, mas como existem desalinhamentos torna-se necessária uma terceira bobine emissora (Hoole & Nguyen, 1999; Shiga, 2005; Stone, 1999; Teixeira, Moutinho, & Coimbra, 2001). Estas bobines são montadas num capacete (*Helmet*) colocado em torno da cabeça do informante (ver Figuras 3.5 e 3.6).



Figura 3.5 - Sistema AG 200 - *Helmet* onde se encontram as bobines transmissoras. Fonte: www.articulograph.com.

Existem vários sistemas de articulografia electromagnética, o sistema *Electro Magnetic Mid-Sagittal Articulography* (EMMA) desenvolvido por Perkell et al, 1992 (MIT), os sistemas Carstens AG 100, AG 200 e AG 500 (Carstens Medizinelektronik, Göttingen, Alemanha), e o sistema Movetrack (Botronic, Hägersten, Suécia) que fundamentalmente variam no número de bobines transmissoras e receptoras utilizadas (Engwall, 2000c). Dos sistemas referidos, todos permitem a obtenção de informação 2D, a excepção é o sistema AG 500 que permite a obtenção de informação tridimensional (Figura 3.6) , mas só recentemente estão a ser divulgados os primeiros resultados da sua utilização (Pfitzinger, 2005).

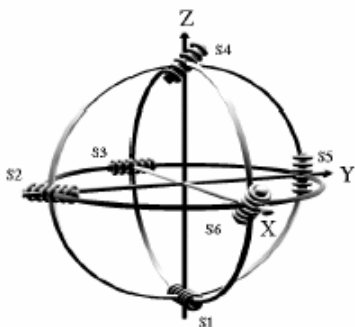


Figura 3.6 - Colocação esférica das seis bobines transmissoras utilizadas no sistema 3D AG 500 (a) e *Helmet* em forma de cubo no mesmo sistema, não sendo necessária a fixação da cabeça. Fonte: Zierdt et al. (2000).

Existe já uma base de dados EMA, para o Português Europeu, cujo corpus foi constituído com o intuito de aprofundar os conhecimentos ao nível da produção de vogais orais e nasais do PE. Esta base de dados foi obtida em Fevereiro de 2001 por Teixeira (Teixeira & Vaz, 2001), no Institut für Phonetik und Sprachliche Kommunikation da Universidade de Munique, utilizando um sistema

Carstens AG 100. O corpus utilizado no presente trabalho inclui alguns item retirados do corpus EMA criado pelos autores referidos, de forma a ser possível conjugar informação diferente, proveniente dos vários métodos. As bobines receptoras ou sensores utilizados foram nove (de forma semelhante ao utilizado noutros estudos para outras línguas - corpora MOCHA, cf. Shiga (2005)), colocados da seguinte forma: três na língua (próximo do ápice da língua, parte posterior da língua e outro a meio caminho entre os dois - dorso da língua), um no lábio inferior, outro no velo (que levantou algumas dificuldades na sua colocação), e um na mandíbula do qual não foi obtida informação independente, devido a um erro de calibração. Com os dados recolhidos pelos sensores obteve-se informação articulatória. Além destes sensores, foram ainda colocados mais dois sensores, para referência, um no nariz e outro nos incisivos superiores, que permitiram corrigir os movimentos da cabeça do falante (Teixeira, Moutinho, & Coimbra, 2001; Teixeira & Vaz, 2001).

As principais vantagens deste método são a capacidade de obter informação relativa a vários articuladores (lábio inferior, maxilar, língua e velo) em simultâneo e em tempo real. Uma vez que existe informação de vários articuladores, é possível avaliar também a inter-acção entre estes (Engwall, 2000c; Stone, 1999; Teixeira & Vaz, 2001). A articulografia electromagnética é também um dos métodos de estudo mais adequados para estudar coarticulação uma vez que permite a obtenção de um vasto número de palavras numa única sessão (Hoole & Nguyen, 1999; Teixeira, Moutinho, & Coimbra, 2001). Como a informação obtida é relativamente simples de analisar torna possível alargar o número de informantes numa determinada investigação (Hoole & Nguyen, 1999). É também possível, com a maior parte dos sistemas EMA, conjugar outras técnicas instrumentais como a EPG, a Ultrasonografia (US) ou outros métodos (OPTOTRAK), se bem que possa existir, com alguns sistemas, algum tipo de interferência (Hoole & Nguyen, 1999).

Uma das limitações deste método reside no facto de só ser possível obter informação bidimensional, na linha média e relativa a alguns pontos, sendo o comportamento do articulador na sua totalidade, inferido e não medido. Este facto, é particularmente condicionante no estudo dos lábios, língua e velo que apresentam movimentos mais fluidos do que rígidos (Stone, 1999). Contudo, com o desenvolvimento e validação da articulografia 3D (3D EMA) algumas destas limitações poderão ser ultrapassadas, nomeadamente no estudo dos movimentos (mais complexos) da língua e do velo, uma vez que é possível recolher informação fora do plano médio-sagital e assim captar os movimentos laterais destes articuladores. O sistema permite ainda o movimento da cabeça e do corpo, não sendo necessária a sua fixação como nos modelos percussores (Figura 3.5) e a orientação dos sensores não é restritiva (Zierdt, Hoole & Tillmann, 1999; Zierdt et al, 2000; Pfitzinger, 2005). Uma outra desvantagem, associada a todos os sistemas, está relacionada com o tempo e esforço dispendido na colocação dos sensores, a maior parte das vezes colados com cola cirúrgica ou utilizando outras opções mais invasivas (Mohammad, 1999). Em termos de segurança e apesar de existir exposição a campos magnéticos é considerado um método relativamente inócuo, com efeitos biológicos mínimos. Não existem, no

entanto, muitos estudos epidemiológicos que avaliem a exposição a campos magnéticos como os utilizados nestes estudos. Ainda assim, não é recomendável a sua utilização em mulheres grávidas (Stone, 1999) nem em indivíduos portadores de *pace maker* (Hoole & Nguyen, 1999).

3.2.3 Electropalatografia

A electropalatografia (EPG) ou palatografia dinâmica é a técnica instrumental utilizada quando se pretende gravar e analisar um aspecto particular da actividade da língua - o contacto da língua com o palato durante a produção de fala (ver Figura 3.7). A EPG permite observar, não só, o padrão de contacto da língua com o palato (obtido também com a palatografia estática) mas também, obter informação dinâmica como a sua variação no tempo (Gibbon & Nicolaidis, 1999; Stone, 1999).

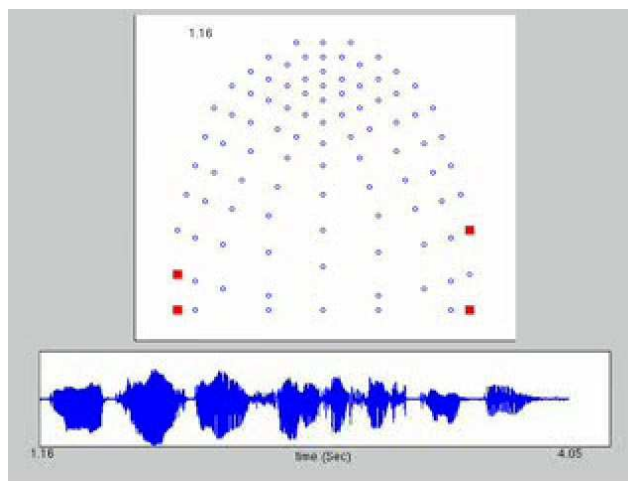


Figura 3. 7 - Imagem referente a uma frame duma palatografia (cima) e respectivo sinal acústico (baixo). Fonte: <http://www.speech.umaryland.edu>

Existem pelo menos 3 sistemas de EPG: (1) o Reading EPG3 desenvolvido na Universidade de Reading (UK), (2) o Rion DP01 da Rion Corporation (Japão) e (3) o Palatometer fabricado pela Kay Elemetrics Corporation (EUA). Os sistemas apresentam algumas diferenças, tais como a construção dos palatos artificiais, o número e construção dos eléctrodos e algumas especificações de *soft* e *hardware*.

Os palatos artificiais ou pseudo palatos utilizados (moldados para se adaptarem perfeitamente ao palato do informante) são manufacturados com materiais acrílicos, não tóxicos, sendo mais rígidos nos sistemas Rion e Reading e mais flexíveis no Palatometer (cf. Gibbon & Nicolaidis, 1999 pp. 230-231).

Ao palato artificial e aos dentes encontram-se ligados um número (variável em cada sistema - desde 62 para o Reading a 96 para Kay Palatometer) de pequenos eléctrodos. Os eléctrodos ou sensores são activados sempre que a língua toca no palato, podendo o padrão de contacto língua-

palato ser inferido pelo sinal de resposta dos eléctrodos. O sinal é depois conduzido através de cabos até uma unidade exterior de processamento. O processamento através de um computador permite a observação, em tempo real, do contacto palato-língua e em simultâneo os dados relativos ao sinal acústico recolhido por um microfone (Gibbon & Nicolaidis, 1999; Stone, 1999; Engwall, 2000c).

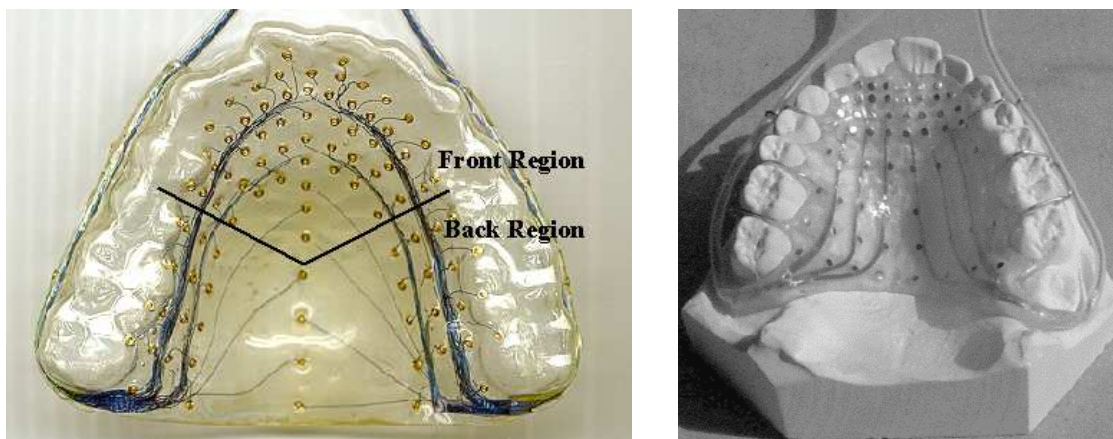


Figura 3. 8 - Palato artificial e disposição dos eléctrodos no sistema Palatometer da Kay Corporation (a) e sistema Reading (b). Adaptado de <http://www.linguistics.ucla.edu/faciliti/facilities/physiology/epg.html>

A disposição dos eléctrodos também não é igual em todos os modelos, por exemplo, o sistema Reading distribui os 62 eléctrodos (1.4 mm diâmetro) pelo palato tendo em consideração determinados pontos anatómicos e o interesse fonético. Assim, os eléctrodos distribuem-se por 8 linhas horizontais com 6 eléctrodos na fileira mais anterior e 8 em cada uma das outras sete linhas (Nguyen, 2000; Engwall, 2000c).

O palato é ainda dividido em zonas fonéticas e sub-zonas (geralmente correspondentes aos pontos de articulação), de tal forma que, tendo como conhecimentos de base a anatomia da língua e os mecanismos de produção de fala é possível inferir qual a parte da língua envolvida na produção de determinado som (Gibbon & Nicolaidis, 1999). No sistema Reading a divisão em zonas inclui a zona alveolar, pos alveolar, palatal e velar. Os movimentos da língua com o palato mole não são registados uma vez que a fileira mais posterior de eléctrodos é anterior à transição palato duro/mole (Nguyen, 2000; Engwall, 2000c).

Os dados provenientes da EPG são sequências bidimensionais (ver Figura 3.9), denominadas palatogramas ou *frames* de onde é possível retirar informação espacial, temporal e estudar a coarticulação lingual. Os palatogramas são lidos da esquerda para a direita e a zona mais anterior corresponde à zona alveolar (Gibbon & Nicolaidis, 1999).

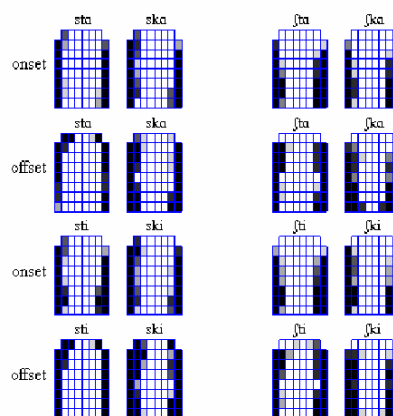
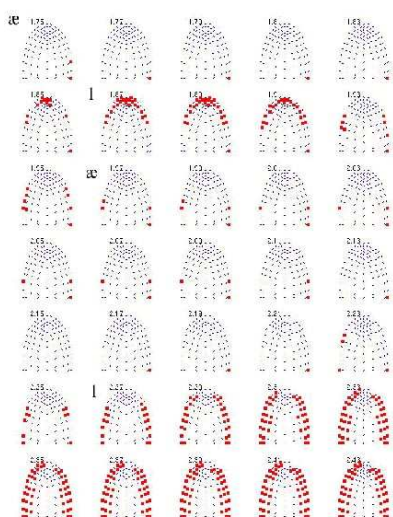


Figure 2: Average EPG patterns at the onset and offset of [s] and [ʃ] for each FCV utterance, computed over 5 repetitions. See text for details.

Figura 3. 9 - EPG frames ou palatogramas: a) sistema Kay Palatometer (disponível em: <http://www.speech.umaryland.edu> e b) Reading EPG 3 (à direita) retirado de Nguyen, Gibbon & Hardcastle (1996)

Em termos de frequência de amostragem, com os sistemas Kay Palatometer e Reading EPG3 é possível uma razão de amostragem de 100 Hz (Nguyen, Gibbon & Hardcastle, 1996; Gibbon & Nicolaidis, 1999; Engwall, 2000c) ou seja uma resolução temporal de 10ms e para o sistema Rion DP01 cerca de 40 Hz (Gibbon & Nicolaidis, 1999).

As principais vantagens da EPG são: (1) a possibilidade de obtenção de informação em tempo real relativamente ao ponto de articulação e ao padrão de contacto palato-língua, (2) simplicidade conceptual (3) fácil manuseamento e (4) e relativamente pouco invasivo (Gibbon & Nicolaidis, 1999; Stone, 1999; Engwall, 2000c). Devido a estas características tem sido largamente utilizada quer em indivíduos sãos (estudos demográficos relativos ao crescimento do palato, dados normativos em crianças e adultos) quer em indivíduos com patologias (as alterações estudadas incluem: fenda palatina, mordida aberta, fala de surdos, etc) e em investigação na área da Fonética e da Língua. Nestas áreas, destacam-se estudos com fricativas e de coarticulação (Nguyen-Trong, Hoole, & Marchal 1991; Hoole, 1993; Engwall, 2000c), e na investigação da fisiologia dos sons, que envolvem o contacto da língua com o palato, em diversas línguas (Japonês, Italiano, Catalão, Alemão e Inglês) (cf. Stone 1999). Como foi já referido é facilmente associável a outras técnicas instrumentais.

Como limitações do método pode referir-se a incapacidade de obter informação quantitativa quando a língua não está em contacto com o palato, o que acontece por exemplo quando a língua se afasta do contacto com o palato, na produção das vogais médias e baixas. Nestes casos, a informação proveniente da EPG é em grande parte inferida. Não fornece informação relativa a outros articuladores e a colocação de um palato artificial poderá também causar alguma interferência na articulação (Mohammad, 1999; Stone, 1999).

3.2.4 Optopalatógrafo

Este sistema permite a obtenção directa da distância entre o palato duro e a língua, mediante a utilização de sensores (*infrared-LED sensors*) colocados num palato artificial. Permite obter informação 3D, a partir de dois cortes ortogonais, utilizando cerca de 16 sensores. Apresenta potencialidade para obtenção de informação referente às vogais, uma vez que, consegue medir directamente a forma e dimensão da cavidade oral, configurada pelo palato duro e língua. A razão de amostragem é da ordem dos 100 Hz (Wrench et al, 1998).

3.3 Sumário

Neste capítulo, apresentou-se uma revisão relativa a algumas técnicas laboratoriais utilizadas na área das Ciências da Fala, para medir e modular a geometria e a dinâmica dos articuladores durante a produção de fala. Como pudemos constatar todos os métodos apresentam algum tipo de limitação, em maior ou menor grau. A técnica de estudo ideal para estudar o tracto vocal, deveria permitir obter de forma directa e tridimensional todo o tracto vocal em tempo real, e sem riscos para o falante. Alguns métodos apresentam boas potencialidades para vir a preencher tais requisitos.

Capítulo 4

Ressonância Magnética

Neste capítulo pretendemos abordar de forma sucinta e compreensível esta técnica de imagem, uma vez que, e dado o enquadramento deste trabalho, uma grande parte dos possíveis interessados na sua leitura, não serão da área da imagem. Para todos os que estão mais próximos da imagiologia, este capítulo surge apenas como uma oportunidade para rever alguns conceitos. Assim, ao longo do capítulo descrevem-se alguns conceitos físicos fundamentais subjacentes a este método de imagem, faz-se uma revisão relativa às sequências de pulso mais frequentemente utilizadas em RM e às técnicas de aquisição rápida. Referem-se os parâmetros que condicionam a qualidade de imagem e os compromissos que têm que ser estabelecidos tendo em atenção os objectivos da aquisição. Uma revisão dos estudos de Ressonância Magnética, no âmbito da produção de fala, será efectuada no capítulo 5 desta dissertação.

4.1 História

Os princípios da Ressonância Magnética Nuclear (RMN), não são recentes. O conhecimento que alguns núcleos atómicos teriam propriedades magnéticas e que estas poderiam ser manipuladas remonta aos anos 20, após os estudos de Wolfgang Pauli (Huettel, Song, & McCarthy, 2004). No entanto, o fenómeno da RMN só viria a ser descrito, de forma independente em 1946, por Felix Bloch e Edward Purcell, que viram reconhecido o mérito da sua descoberta ao serem galardoados com o Prémio Nobel da Física, em 1952. Antes da sua aplicação na área da medicina, no início da década de 80, foi utilizada na área da química (espectroscopia por Ressonância Magnética) para obtenção de informação química e física relativa a diferentes moléculas (Huettel, Song, & McCarthy, 2004; Marques, 2003).

Em 1971, Raymond Damadian demonstrou que os tempos de relaxação de tecidos normais e tumorais variavam, o que o levou a considerar que a ressonância magnética poderia ter utilização no estudo de processos biológicos, em particular no estudo de patologia tumoral (Hornack 1996-2006; Huettel, Song, & McCarthy, 2004). No início da década de 70, foi introduzida por Hounsfield a Tomografia Axial Computorizada (TAC), hoje designada apenas Tomografia Computorizada (TC). No ano de 1973, Paul Lauterbur (Nobel da Medicina em 2003) introduzindo o conceito de gradiente conseguiu efectuar a localização espacial de objectos e, usando uma técnica de retro projecção semelhante à utilizada em TC, apresentou a primeira imagem de RM relativa a um par de tubos teste preenchidos por água (Hornack 1996-2006; Huettel, Song, & McCarthy, 2004;

Marques, 2003). Em 1975, Richard Ernst (Nobel da química em 1991, pelos seus trabalhos em espectroscopia) propõe métodos de codificação da informação em função da fase e da frequência e a utilização da Transformada de Fourier (TF). Em 1977, Peter Mansfield (Nobel da Medicina em 2003) apresentou os fundamentos da imagem eco-planar (EPI) que viriam a permitir, anos mais tarde, adquirir imagens em tempo real. Apesar das primeiras tentativas de obtenção de imagens do corpo humano não terem sido bem sucedidas, a primeira imagem do corpo humano surge em 3 de Julho de 1977 por Damadian e colaboradores, demorando cerca de 4 horas (Huettel, Song, & McCarthy, 2004; Marques, 2003). Em 1980, Edelstein e colaboradores obtêm uma imagem em aproximadamente 5 minutos utilizando a técnica de codificação descrita por Ernst (Hornack 1996-2006).

Em 1987, é utilizada a técnica EPI para obter um vídeo em tempo real referente a um ciclo cardíaco. No mesmo ano, desenvolve-se a Angiografia por Ressonância Magnética e em 1992 surge a Ressonância Magnética Funcional (fRM), que veio permitir a obtenção de informação funcional relativa a diferentes tecidos (Hornack 1996-2006). Muitos outros investigadores e aperfeiçoamentos a nível tecnológico contribuíram para que a Ressonância Magnética seja, nos dias de hoje, uma das técnicas de maior valor na medicina e das mais promissoras no futuro, nomeadamente na obtenção de informação a nível molecular (Weissleder, 1999).

Alguns nomes têm sido atribuídos ao processo que permite medir a absorção e emissão de energia por um núcleo quando sob a acção de um campo magnético. Desde Indução Nuclear, a Ressonância Paramagnética Nuclear até RMN, o nome preferido e mais utilizado nas áreas da química e da física, desde os finais da década de 50 (Elster & Burdette, 2001). Em contexto clínico perdeu o nome nuclear, conotado negativamente com a radioactividade nuclear (Elster & Burdette, 2001; Hashemi & Bradley, 1997; Hornack 1996-2006; Huettel, Song, & McCarthy, 2004; Marques, 2003) e passou a ser designada por Imagem por Ressonância Magnética (IRM), MRI no Inglês, ou somente Ressonância Magnética (RM).

A RM é uma técnica de imagem tomográfica que à semelhança de outras técnicas consegue obter informação que representa a distribuição espacial de uma grandeza física. Apresenta capacidade multiplanar, ou seja, permite obter informação bidimensional (2D) de forma directa, em qualquer orientação, informação volumétrica (3D) e informação 4D se for considerada uma distribuição espectro-espacial. Permite obter uma boa resolução de contraste, em particular a nível dos tecidos moles (Liang & Lauterbur, 2000). É considerado um método de imagem relativamente inócuo, uma vez que utiliza o espectro electromagnético de forma não ionizante, recorrendo a ondas de radiofrequência (RF).

4.2 Equipamento

Para a obtenção de imagens por RM é necessário: (1) a presença de um campo magnético estático, com uma intensidade B_0 , produzido por um magnete, (2) um sistema de RF de emissão/recepção (antenas de RF, também designadas por bobines ou *coils*) e (3) bobines de gradientes que originam um campo magnético variável no tempo (dB/dt), que se sobrepõe a B_0 , e que permite efectuar a codificação do sinal. É necessário ainda um sistema computadorizado, *hardware* e *software* adequados para aquisição e processamento da informação (Ester & Burdette, 2001; Huettel, Song, & McCarthy, 2004; Liang & Lauterbur, 2000; Marques, 2003).

4.2.1 Magnete

Origina um campo magnético que deverá ser estático, uniforme e homogéneo, com uma intensidade de campo que, para fins clínicos, se situa geralmente entre os 0,2 T⁹ e os 1,5 T. No presente, os magnetes de 1,5 T são os mais divulgados na área do diagnóstico médico. No entanto, estão já em grande expansão os magnetes de 3T e na área da investigação estão já a ser utilizados magnetes de 7T e 8T, em seres humanos (Elster & Burdette, 2001). Em função da intensidade de campo que permitem costumam designar-se como sendo de baixo, médio ou alto campo, havendo ainda quem considere categorias como ultra-baixo (0,002T) e muito alto campo (7 a 9T). Geralmente os magnetes com uma intensidade de campo superior a 1T são designados como campos altos (Elster & Burdette, 2001; Marques, 2003). As configurações mais comuns são do tipo; (1) aberto e (2) fechado, em formato de túnel. Os magnetes podem ainda ser do tipo permanente, resistivo ou superconductor. Os magnetes que criam campos altos são, em geral, supercondutores e apresentam formato em túnel. No entanto, recentemente, foram apresentados magnetes de 0,7 e 1T com configuração aberta (Marques, 2003).

Um factor importante a considerar relativamente ao campo magnético principal é a homogeneidade espacial, que pode ser definida pelo desvio na intensidade do campo ao longo de um determinado volume, dentro da região de interesse:

$$\text{Homogeneidade} = \frac{B_{0,\text{máx}} - B_{0,\text{mín}}}{B_{0,\text{médio}}} \quad (4.1)$$

Num magnete para fins clínicos e para obter imagens de qualidade é necessária uma homogeneidade moderada dentro de um volume razoavelmente grande. Em geral, 10 a 50 partes por milhão (ppm) em 30 a 50 cm de um volume esférico (Liang & Lauterbur, 2000). Para efectuar espectroscopia e difusão as exigências relativas à homogeneidade de campo são ainda superiores e será necessário ter em atenção a estabilidade temporal do campo (Elster & Burdette, 2001).

⁹ Tesla (símbolo T) é a unidade de densidade de fluxo magnético (ou indução magnética)

Existem vantagens e desvantagens associadas a cada um destes tipos de magnete quer relacionados com a intensidade de campo, com a configuração ou com o tipo de magnete. As vantagens e desvantagens de cada um deverão ser enquadradas no tipo de aplicação que se pretenda em cada situação. Referindo apenas algumas características dos diferentes magnetes, pode dizer-se que, em geral, a relação Sinal-Ruído (SNR ou S/R) e a homogeneidade de campo é superior nos altos campos, sendo também mais elevada a resolução espacial e a espectral (Liang & Lauterbur, 2000; Marques, 2003). No entanto, com o aumento de B_0 , é também maior o desvio químico, a susceptibilidade magnética e os artefactos de fluxo o que poderá interferir com a qualidade da imagem de forma negativa em determinadas circunstâncias mas que poderá ter utilidade diagnóstica noutras. Com os campos altos, os campos marginais são mais elevados sendo mais difícil confinar o campo magnético, levantando questões de instalação e de segurança.

De um modo geral, os altos campos permitem todas as aplicações como, por exemplo, RM cardíaca, espectroscopia, fRM entre outras, o que nem sempre é possível com campos de menor intensidade. A maior parte destes magnetes apresentam configuração em túnel, como já se referiu, o que representa uma desvantagem em crianças, em indivíduos com claustrofobia ou quando se pretendem efectuar procedimentos de intervenção.

4.2.2 Sistema de RF

O sistema de RF é constituído por uma bobine ou antena transmissora capaz de gerar um campo magnético rotacional, referido vulgarmente por B_1 , e por uma bobine receptora que transforma a magnetização num sinal eléctrico. Em alguns sistemas, estas duas funções podem ser efectuadas pela mesma bobine denominando-se nesse caso por emissora/receptora ou transmissora /receptora. Outras antenas são apenas transmissoras e outras só receptoras. Existem antenas dos mais diversos formatos, tamanhos e tipos em função das diferentes aplicações. Podem ser rígidas ou flexíveis, grandes ou pequenas, de volume ou de superfície, de polarização linear ou de polarização circular (em quadratura), recebendo sinal por mais de um canal (Lafuente, 2000)

Recentemente desenvolveu-se a tecnologia *phased array* que permite uma maior sensibilidade na detecção, permitindo uma boa S/R sem comprometer a resolução espacial. As antenas *phased array* ou em arranjo de fase são um conjunto de antenas que, apesar de juntas, recebem sinal individualmente e funcionam de forma independente. Os sinais individuais de cada uma são combinados para se obter uma imagem com maior S/R e maior cobertura. Combinam-se as vantagens das antenas pequenas (alta S/R e resolução espacial) com a possibilidade de obter um campo de visão (FOV) grande (sem aumento do tempo de exame) através do agrupamento das antenas (Lafuente, 2000). Este tipo de antenas permitiu também a implementação de técnicas de aquisição paralela, de que falaremos mais adiante, com conseqüente diminuição do tempo de

aquisição (TA) e aumento da resolução temporal, fundamentais a algumas aplicações (Glokner, Hu, Standley, Angelos, & King, 2005; Heidemann et al., 2003; Pruessmann, Weiger, Scheidegger, & Boesiger, 1999).

4.2.3 Gradientes de campo magnético

O campo magnético principal é importante num sistema de RM, mas em termos de qualidade da imagem final e das aplicações possíveis com um determinado equipamento o sistema de RF e de gradientes, são igualmente importantes. Quer a resolução espacial, quer a rapidez com que se consegue adquirir, dependem do desempenho (*performance*) do sistema de gradientes (Elster & Burdette, 2001). Num equipamento de RM poderão estar instalados diferentes tipos de gradientes para uma mesma intensidade de campo. Habitualmente, os gradientes são caracterizados pela intensidade (amplitude) máxima que permitem. A maior parte dos equipamentos está equipado com gradientes da ordem dos 15 a 20 mT/m de intensidade máxima, mas para determinadas aplicações (EPI e outras sequências para aquisição rápida) são necessários gradientes com melhor desempenho (alta *performance*), com intensidade máxima entre os 20 e os 40 mT/m. Em investigação são utilizados gradientes com intensidade de 50 mT/m. No que diz respeito à intensidade dos gradientes, é habitualmente referido que quanto maior esta for, melhor, uma vez que permitirá um maior leque de aplicações (Elster & Burdette, 2001; Liang & Lauterbur, 2000). Outras características dos gradientes, extremamente importantes, mas a que por vezes é dispensada menor atenção são: (1) o *rise time* (ms) ou tempo de subida do gradiente (representa o tempo que o gradiente demora a atingir a intensidade máxima) e (2) o *slew rate*, que se define como:

$$Slewrate = \frac{Amplitude_{Máx}}{Risetime} \quad (4.2)$$

O *slew rate* é dado por T/m/s. O *slew rate* influencia o TR e TE mínimos e nas sequências TSE e EPI determina o espaçamento mínimo entre os ecos (Elster & Burdette, 2001). Para EPI, são ainda necessários tempos de subida inferiores a 0,1 ms.

De forma genérica, os gradientes determinam a espessura de corte mínima, a dimensão da matriz a FOV constante e o FOV mínimo para matriz constante. A título de exemplo, podemos referir que, as especificações para uns gradientes de alta *performance* poderão ser: 40 mT/m de máxima amplitude, *rise time* de 0.2 ms, *slew rate* de 200 T/m/s para um FOV de 40 cm.

4.3 Conceitos fundamentais em RM

4.3.1 Spin, precessão e frequência de precessão

Como já referimos a RMN é um processo físico que envolve a absorção e emissão de energia por um núcleo. No que diz respeito à imagem para fins clínicos (IRM ou só RM) utiliza-se essencialmente o núcleo do átomo de Hidrogénio, em grande parte devido à sua abundância nos diferentes tecidos do corpo humano (cerca de 60% do corpo humano é água) e ao seu momento magnético elevado. O **spin**, é um termo que tem origem na descrição quântica da RMN, em que cada partícula sub-atômica, incluindo o seu núcleo, é definida por um número quântico de *spin* (I). Este número, chamado *spin* nuclear ou apenas *spin*, é directamente proporcional ao seu momento angular e em consequência à sua capacidade em exibir ressonância. Núcleos com número par de protões, ou seja, em que $I = 0$ não apresentam momento angular. Os núcleos com um número ímpar de protões, de que o hidrogénio (^1H) é um exemplo, têm um número quântico de *spin*. Este número para o H é $I = 1/2$. Na mecânica quântica o valor de I para um determinado núcleo determina o número de estados de energia (discretos) possíveis quando este é sujeito à acção de um campo magnético externo B_0 . Os níveis de energia permitidos podem ser dados pela equação 4.3:

$$E_m = -m\gamma\hbar B_0 \quad (4.3)$$

em que γ representa a razão giromagnética, um valor constante para um determinado núcleo, \hbar é a constante de Planck dividida por 2π , e m o número quântico que pode assumir valores de $-I$ a $+I$. Assim, existem $(2I + 1)$ níveis de energia de energia separados por:

$$\Delta E = \gamma\hbar B_0 \text{ e que, como } \Delta E = \hbar\omega \text{ vem}$$

$$\omega = \gamma B_0 \quad (4.4)$$

Esta última fórmula representa a **frequência de Larmor** ou **frequência de precessão**, a que voltaremos, um pouco mais adiante neste texto.

Para o núcleo do átomo de hidrogénio ($I = 1/2$) existem 2 níveis de energia (níveis de Zeeman), separados pela frequência de Larmor, $\omega = \gamma B_0$. Estes estados de energia correspondem aos dois estados principais quantizados do protão, denotados por $|1/2\rangle$ e $|-1/2\rangle$, a que se chamam estado paralelo e anti-paralelo, respectivamente. De forma simplificada aparecem com frequência na bibliografia referente à RM, esquemas ilustrativos com os protões orientados para cima e para baixo, o que pode criar a ideia, errada, de que os protões só poderão estar num desses dois estados, ou que, após absorverem energia do pulso "rodam" simplesmente de um estado para

outro. O que a mecânica quântica prevê é que um próton apresentará em determinado momento uma combinação linear dos dois estados. Se assim não fosse, não seria possível explicar a existência de pulsos de 10° ou 90° , ou explicar porque é que o vector de magnetização transversal, imediatamente após o pulso de 90° , tem exactamente a mesma magnitude da magnetização longitudinal antes da aplicação desse pulso (Elster & Burdette, 2001).

Considerando os dois estados de energia principais $|1/2\rangle$ e $|-1/2\rangle$, para o átomo de hidrogénio, a sua função de onda pode ser definida por:

$$\psi(t) = a|1/2\rangle + b|-1/2\rangle \quad (4.5)$$

onde a e b , representam amplitudes quânticas, variáveis com o tempo, que incorporam as oscilações harmónicas para cada estado principal. Para um único próton, colocado sob acção do campo magnético externo B_0 , cada termo a e b contém uma frequência harmónica simples respectivamente $(\omega_0/2)$ e $(-\omega_0/2)$, e cuja diferença é a frequência de Larmor, $\omega = \gamma B_0$. Quando um pulso de RF é aplicado à frequência de Larmor, a sua acção sob $\psi(t)$ é induzir “rotação” ou mistura das componentes a e b , com preservação das suas amplitudes totais $(|a|^2 + |b|^2)$. Desta maneira, explica-se porque é que a magnetização transversa depois do pulso de RF é igual, em magnitude, à componente longitudinal antes da aplicação do pulso.

A temperaturas próximas do zero absoluto, poder-se-ia esperar que a maior parte dos *spins* se mantivessem no estado de menor energia $(|a|^2 > |b|^2)$. No entanto, à temperatura corporal, esta tendência para os *spins* “preferirem” o estado de menor energia é contraposta pela agitação térmica que tende a igualar os dois níveis de energia (paralelo e anti-paralelo). O equilíbrio resultante representa um compromisso previsto pela distribuição de Boltzmann:

$$N^+/N^- = \exp[-\Delta E/kT] \quad (4.6)$$

onde N^+ e N^- representam o número de *spins* que se espera encontrar no estado paralelo ou anti-paralelo, ΔE a diferença de energia entre um estado e o outro, k a constante de Boltzmann (1.38×10^{-23} Joule/K), e T a temperatura do sistema em graus Kelvin. À temperatura corporal e para um campo de 1.0 T, esta equação prevê que N^+ e N^- sejam aproximadamente iguais, existindo apenas um pequeno excesso de prótons ($\sim 3 \times 10^{-6}$) no estado de menor energia (Figura 4.1). Este número depende da Intensidade de B_0 , aumentando com o aumento desta (Brown & Semelka, 2003; Bushberg, Seibert, Leidholdt & Boone, 2002; Elster & Burdette, 2001; Liang & Lauterbur, 2000).

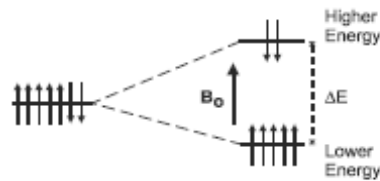


Figura 4. 1 - Estados de energia principais para o átomo de Hidrogênio. O alinhamento paralelo a B_0 corresponde ao estado de menor energia. A diferença de energia entre os dois níveis (ΔE) é proporcional a B_0 . Fonte: Brown & Semelka (2003).

Quando os *spins* não estão sob acção de um campo magnético, giram em torno do seu próprio eixo, dando origem ao seu próprio campo magnético. Sob a acção de um campo magnético externo, além deste movimento giratório em torno de si próprio, os *spins* precessam em torno da direcção de B_0 . A razão a que os *spins* precessam em torno de B_0 , é dada pela equação de Larmor $\omega = \gamma B_0$, referida no ponto anterior. A frequência angular ω , pode ser expressa em radianos por segundo. A constante de proporcionalidade γ é fixa para determinado núcleo, tendo o valor de 42,6 para o Hidrogênio. Assim, num campo magnético de 1T, a frequência de precessão será $\omega = 42.6 \text{ MHz}$ e num campo de 1.5 T será aproximadamente 63 MHz .

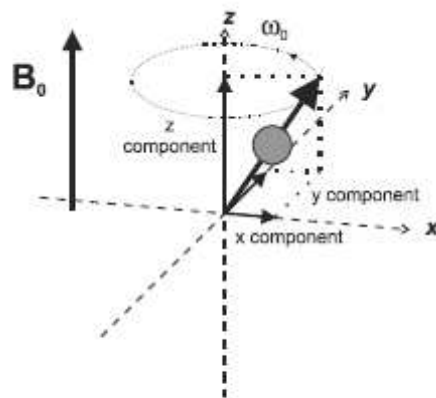


Figura 4. 2 - Sob a acção de um campo magnético, os prótons precessam em torno deste. O eixo precessional é paralelo ao campo magnético principal B_0 . A componente z não altera a sua direcção ou magnitude mas as componentes x e y, variam no tempo com uma frequência proporcional a B_0 . Fonte: Brown & Semelka (2003).

O pulso de RF, geralmente utilizado para obtenção de imagens por RM, denominado B_1 , é aplicado por um curto período de tempo, num plano perpendicular a B_0 . Para que haja inclinação do vector de magnetização (M), este campo B_1 , terá que girar precisamente com uma frequência

igual à frequência de precessão dos *spins* para cada B_0 . Só desta forma é possível a troca de energia entre os sistemas e a ocorrência de ressonância. Se B_1 fosse aplicado a qualquer outra frequência, os *spins* estariam alternativamente em fase e fora de fase com ele, não seria possível manter o torque magnético induzido por B_1 , e não haveria ressonância. Em termos de mecânica quântica isto equivale a dizer que a frequência a aplicar deverá ser coincidente com a frequência harmônica da função de onda principal que descreve o estado desse particular conjunto de prótons.

Com a aplicação do pulso de RF, com uma frequência igual à frequência de precessão dos *spins*, o sistema vai absorver energia passando de um estado de menor energia para um estado de maior energia, anti-paralelo relativamente a B_0 . Quando o pulso cessa, o retorno ao estado de equilíbrio é feito com emissão ou perda de energia, que veremos com mais detalhe no ponto seguinte. O ângulo de rotação (α) também chamado Flip Angle (FA) é determinado por, $\alpha = \gamma B_1 t_p$, onde t_p representa o intervalo de tempo em que se aplica o pulso de RF e B_1 , a intensidade do mesmo. Grande parte das vezes, α é próximo dos 90° (Brown & Semelka, 2003; Elster & Burdette, 2001; Hashemi & Bradley, 1997; Pooley, 2005)

4.3.2 Tempos de relaxação T1, T2 e T2*

Depois da aplicação de um pulso de excitação de RF os prótons recebem energia e começam a precessar no plano XY criando uma componente de magnetização transversal (M_{xy}), que oscila em torno do eixo longitudinal (Z), com todos os prótons inicialmente em coerência de fase. Este estado, não corresponde ao estado de energia mínima e como tal, logo que cessa o pulso de RF, duas situações ocorrem:

- (1) - Os prótons ou *spins* regressam ao estado de energia mínima - a componente de magnetização (M_z) é recuperada de forma lenta ao longo do eixo Z;
- (2) - Os prótons começam a perder a coerência de fase (desfasamento) - a componente de magnetização transversal (M_{xy}) decai rapidamente;

Estes acontecimentos apesar de ocorrerem separadamente, são coincidentes.

T1 - Tempo de relaxação longitudinal ou tempo de recuperação da magnetização longitudinal

O tempo de relaxação longitudinal (T1) representa o tempo necessário para que os prótons recuperem 63% da magnetização longitudinal M_z (ao longo do eixo Z). Uma vez que a energia

recebida pelo pulso de RF é libertada para a rede envolvente, de forma a restabelecer-se o estado de equilíbrio, também é frequente chamar ao tempo T1, o tempo de relaxação *spin*-rede (do inglês, *spin-lattice*). O processo de recuperação da magnetização longitudinal é exponencial e pode ser representado matematicamente pela equação 4.7 onde M_0 representa o valor da magnetização para o tempo $t = 0$, e graficamente pela curva da figura 4.3 (Bushberg, Seibert, Leidholdt & Boone, 2002; Elster & Burdette, 2001; Hashemi & Bradley, 1997).

$$M_z(t) = M_0 \left(1 - e^{-t/T_1}\right) \quad (4.7)$$

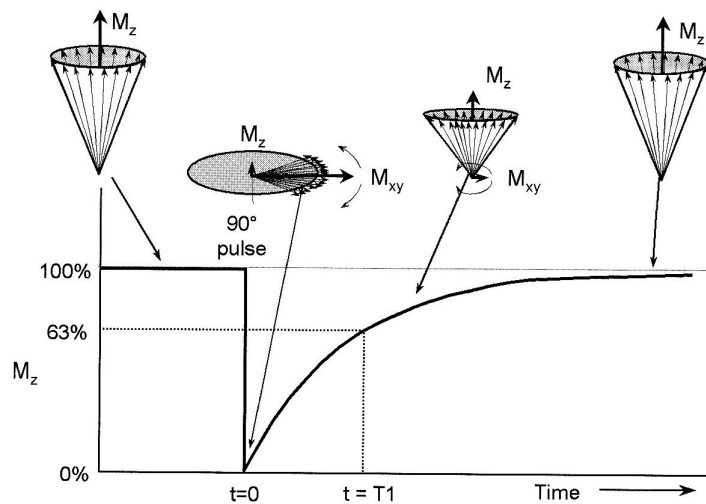


Figura 4.3 - Tempo de recuperação da magnetização longitudinal - Após aplicação de um pulso de excitação de 90° , a magnetização longitudinal (M_z) é recuperada exponencialmente sendo caracterizada pela constante temporal T1. Depois de um período de tempo igual a T1, 63% da magnetização longitudinal é recuperada. Fonte: Bushberg et al. (2002).

T2 - Tempo de relaxação T2, decaimento T2, Relaxação spin-spin

O tempo de relaxação T2 corresponde ao tempo que os *spins* ou prótons demoram a decair de $\frac{1}{e}$ ou a perder 37% da magnetização transversal (Figura 4.4). Depois de cessar o pulso de RF, todos os *spins* estão em fase, o que significa que estão alinhados e precessam à mesma frequência ω_0 . No entanto, devido às interações *spin-spin* e à falta de homogeneidade do campo magnético (que existe sempre, mesmo que o campo seja considerado estável), os prótons precessam a frequências ligeiramente diferentes, o que faz com que rapidamente fiquem fora de fase, conduzindo ao rápido decaimento da magnetização transversal. A razão a que ocorre este decaimento é cerca de 5 a 10 vezes superior à da recuperação da magnetização longitudinal. O

decaimento M_{xy} pode ser definido pela equação 4.8, onde se verifica que M_{xy} decai a uma razão caracterizada pela constante temporal T2.

$$M_{xy}(t) = M_0 \left(e^{-t/T_2} \right) \quad (4.8)$$

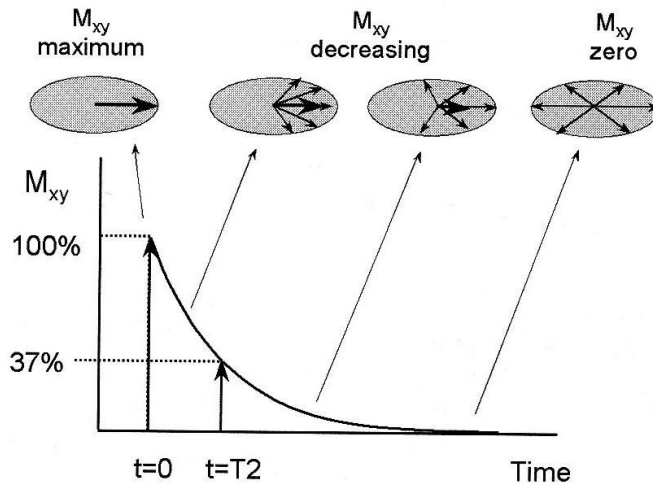


Figura 4. 4 - Tempo de relaxamento T2. Fonte: Bushberg et al. (2002).

O decaimento T2 depende essencialmente das interações *spin-spin*, característica inerente de cada tecido, sendo esta a principal causa do desfaseamento dos prótons. Os tempos de relaxação T1, T2, representam propriedades inerentes dos tecidos e como tal são constantes para cada tecido, para uma determinada intensidade de campo magnético (ver tabela 4.1).

T2* além das características intrínsecas aos tecidos é susceptível a factores externos como, por exemplo, a não homogeneidade do campo magnético principal B_0 , efeitos de susceptibilidade magnética, desvio químico, etc.

Na Tabela 4.1 podem observar-se os valores de relaxamento T1, T2 (em ms) para diferentes tecidos e para intensidade de campo de 0,5 T e de 1,5 T.

TECIDO	T1, (0,5T) (ms)	T1 (1,5 T) (ms)	T2 (ms)
Gordura	210	260	80
Músculo	550	870	45
Substância Branca	500	780	90
Substância Cinzenta	650	900	100
LCR	1800	2400	160

Tabela 4. 1 - Valores referentes ao tempo de relaxamento T1 e T2 (em ms) para Intensidade de campo de 0,5 T e 1,5 T. Fonte: Bushberg et al. (2002).

4.3.3 Free Induction Decay (FID) ou decaimento de indução livre

Numa situação ideal, em que se considera que o campo magnético é perfeitamente uniforme, o sinal recebido numa bobine, ao longo do tempo poderia ser caracterizado por uma curva como a que se representa na figura 4.5, no entanto, e como vimos atrás, imediatamente depois de cessar o pulso de RF ($t = 0$) os *spins* devido às interacções *spin-spin* e à falta de homogeneidade do campo magnético externo, começam a desfasar e o sinal começa a decair em relação ao seu valor original.

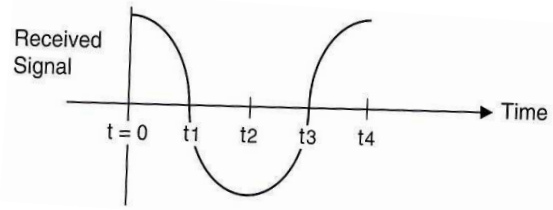


Figura 4. 5 - Sinal recebido numa bobine, num campo uniforme para uma situação ideal. Fonte: Hashemi & Bradley (1997).

Uma forma mais realística de representar o sinal recebido pelo receptor, ao longo do tempo, pode ser observado na figura 4.6.

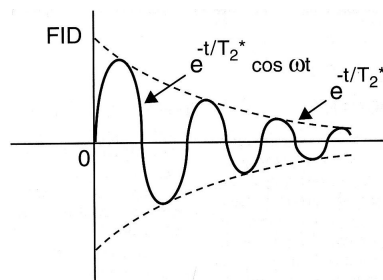


Figura 4. 6 - Decaimento de Indução Livre (FID). Fonte: Hashemi & Bradley (1997).

A este decaimento, chama-se decaimento de indução livre, pois imediatamente após a cessação do pulso de RF, os *spins* começam a precessar livremente, havendo um decaimento do sinal ao longo do tempo, com indução de corrente numa bobine receptora. Matematicamente, o FID é definido pela equação 4.9:

$$M_{xy}(t) = M_0 e^{-t/T_2^*} (\cos \omega_0 t) \quad (4.9)$$

- 1) e^{-t/T_2^*} pois o sinal decai de forma exponencial ao longo do tempo e neste caso a constante temporal desta função é T_2^* e
- 2) $\cos \omega_0 t$ uma vez que se trata de um onda sinusoidal com frequência ω_0 .

T_2^* difere de T_2 essencialmente porque além das interações *spin-spin*, o T_2^* conta também com a falta de homogeneidade do campo magnético (figura 4.7). Enquanto que T_2 é constante para cada tecido, T_2^* varia com o campo magnético e é um decaimento mais rápido que T_2 (Brown & Semelka, 2003; Elster & Burdette, 2001; Hashemi & Bradley, 1997). A equação 4.10 relaciona as duas constantes temporais.

$$1/T_2^* = 1/T_2 + \gamma\Delta B \quad (4.10)$$

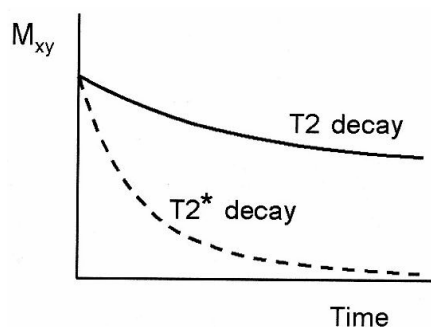


Figura 4. 7 - Decaimento T_2 e T_2^* . Fonte: Hashemi & Bradley (1997).

Resumindo, vimos que $T_1 > T_2 > T_2^*$, e que os tempos de relaxação variam em função das características dos tecidos. A densidade protônica, e os tempos de relaxamento T_1 e T_2 são propriedades fundamentais dos tecidos que podem ser exploradas para obtenção de contraste nas imagens RM. A intensidade de campo influencia o tempo de relaxamento T_1 , mas muito pouco o tempo de relaxamento T_2 (Bushberg, Seibert, Leidholdt & Boone, 2002).

4.3.4 Tempo de Repetição e Tempo de Eco (TR e TE)

É importante definirem-se os termos TR e TE, uma vez que, para compreender os diagramas das sequências de pulsos de que falaremos mais adiante, no ponto 4.7, é necessário saber o que representam, e como podem ser modulados de forma a maximizar a diferenciação entre tecidos. Depois da aplicação de um pulso de excitação RF de 90° , seguir-se-á a aplicação de um outro pulso de 90° . O **Tempo de Repetição** é, por definição, o intervalo de tempo que decorre entre a aplicação de dois pulsos sucessivos RF de 90° .

Considerando $t = 0$ o momento em que se aplica o pulso RF, no instante imediatamente antes, o vector de magnetização está sobre o eixo dos ZZ (vector M_0), após o pulso de RF existe magnetização ao longo do plano XY (M_{xy}), sem componente ao longo do eixo longitudinal. À medida que t se aproxima de TR , há uma recuperação gradual de M_z e perda de M_{xy} . Se, quando $t = TR$ a magnetização ao longo do plano transversal for muito pequena e for aplicado

um novo pulso de RF, a magnetização longitudinal existente vai de novo, sofrer uma inclinação para o plano transversal. O vector de magnetização M_z , para $t = TR$, é menor que o vector original uma vez que o segundo pulso foi aplicado antes de existir uma recuperação total do vector de magnetização longitudinal. Se na equação 4.7, fizermos $t = TR$, vem:

$$M_z (TR) = M_0 \left(1 - e^{-TR/T_1} \right) \quad (4.11)$$

Por sua vez, a magnetização no plano transversal, que está directamente relacionada com o sinal detectável na bobine será dado por:

$$M_{xy} (t) = M_0 \left(1 - e^{-TR/T_1} \right) e^{-t/T_2} \quad (4.12)$$

Esta equação demonstra que o sinal RM depende, não só da magnetização original M_0 , (que depende de $N(H)$) mas também de características dos tecidos, tal como expresso pelas constantes temporais T1 e T2.

Tempo de Eco (TE) - Intervalo de tempo entre a aplicação do pulso RF de 90° e a sua medição. O momento de medição será tal que o sinal recebido apresentará a sua maior amplitude, sendo considerado um eco da transmissão original. Se na equação anterior se substituir t por TE, obter-se-á o sinal RM numa imagem, para um particular TE.

$$M_{xy} (t) = M_0 \left(1 - e^{-TR/T_1} \right) e^{-TE/T_2} \quad (4.13)$$

Esta equação é a base que permite modular o sinal de um determinado tecido, controlando os parâmetros TR e TE.

Em RM interessa essencialmente comparar o sinal proveniente de diferentes estruturas, para assim as caracterizar. A diferença de sinal entre dois tecidos ou estruturas reflecte o contraste¹⁰ entre elas. Assim, o contraste entre as duas estruturas é dado pela diferença entre o sinal proveniente de cada uma delas (Huettel, Song, & McCarthy, 2004). Matematicamente, podemos representar o contraste entre 2 estruturas pela fórmula:

$$C_{AB} = M_{0A} \left(1 - e^{-TR/T_{1A}} \right) e^{-TE/T_{2A}} - M_{0B} \left(1 - e^{-TR/T_{1B}} \right) e^{-TE/T_{2B}} \quad (4.14)$$

Para conhecer a intensidade do sinal (IS) de cada tecido, incluindo já os efeitos da não homogeneidade do campo magnético, pode recorrer-se à fórmula 4.15:

$$SI \propto N(H) \left(e^{-TE/T_2} \right) \left(1 - e^{-TR/T_1} \right) \quad (4.15)$$

¹⁰ Definição de contraste - O contraste numa imagem é extremamente importante, dadas as dificuldades do sistema visual humano em discriminar diferenças de luminância. Apesar da percepção do contraste numa imagem depender de inúmeros factores, tecnicamente é definido em termos de diferenças de Intensidade de Sinal (Liang & Lauterbur, 2000).

4.3.5 Ponderação DP, T1 e T2

Densidade Protónica

A forma mais simples de contraste em RM é baseada na densidade de prótons ou densidade protónica (DP) dos tecidos. A imagem baseada na densidade de prótons, como o próprio nome indica, fornece contraste com base no número de prótons $N(H)$ presente num determinado voxel, o que é variável nos diferentes tecidos. Para maximizar o contraste baseado em DP, utilizam-se sequências de pulso em que se minimizam os efeitos T1 e T2, de forma a que a IS seja principalmente dependente de $N(H)$, o que se consegue seleccionando TR longos e TE curtos.

Ponderação T1

Se na equação, 4.15 se fizer $TR \rightarrow \infty$, então $(1 - e^{-TR/T1}) \rightarrow 1$ e $SI \rightarrow N(H) \left(e^{-TE/T2^*} \right)$. Isto quer dizer que, com um TR longo, se eliminam os efeitos T1 da imagem. Se por outro lado, o TR for demasiado curto, $TR \approx 0$, vem $e^{-0/T1} = e^0 = 1$ e $1 - e^{-TR/T1} = 1 - 1 = 0$ ou seja, deixa de haver sinal, o que também não é desejável.

O ideal será ter um TR que não seja muito diferente do T1 dos tecidos a estudar. Se observarmos a figura 4.8 e se escolhermos um $TR = TR_1$, o TR não é suficientemente longo para eliminar o termo T1 da equação, e teremos:

$$SI(\text{tecidoA}) / SI(\text{tecidoB}) = \left(1 - e^{-TR_1/T1A} \right) / \left(1 - e^{-TR_1/T1B} \right) \quad (4.16)$$

Como o tempo T1 dos dois tecidos é diferente, a escolha de um TR pequeno realça a diferença entre eles, podendo dizer que foram discriminados com base no seus tempos de relaxamento T1.

Em relação ao TE, interessa minimizar os efeitos T2*, assim se $TE \approx 0$, vem $e^{TE/T2^*} \rightarrow e^0 = 1$ ou seja $SI \propto N(H) \left(1 - e^{-TR/T1} \right)$. Daqui se pode concluir que, para se obter uma imagem baseada nas características T1 dos tecidos, se deve escolher um valor de TR e TE curtos.

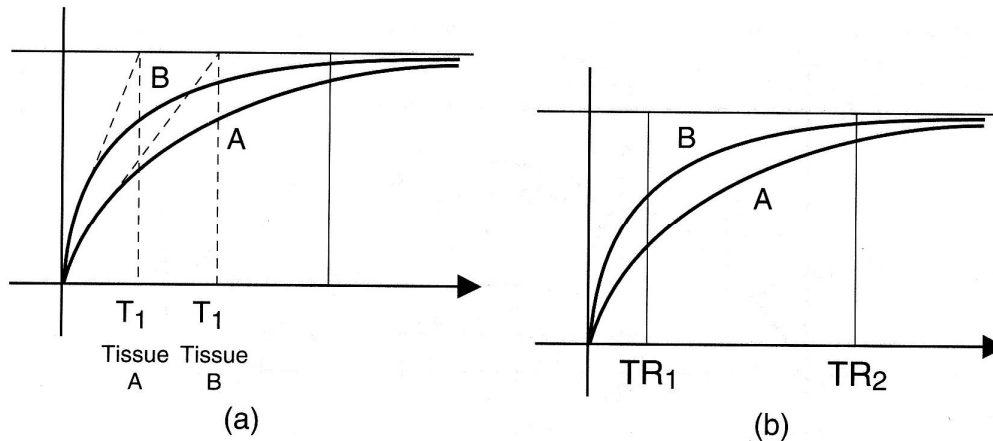


Figura 4. 8 - TR adequado para obtenção de ponderação T1. Fonte: Hashemi & Bradley (1997).

Ponderação T2 e T2*

Para obter imagens baseadas nas características T2 e T2* dos diferentes tecidos devem-se minimizar os efeitos T1 na imagem. Como vimos atrás, se for seleccionado um TR longo, ou seja um TR, 4 ou 5 vezes o valor do T1 dos tecidos, será suficiente para que o efeito T1 seja negligível. Se por outro lado, for escolhido um TE longo maximiza-se o contraste baseado nas diferenças T2 e T2*. Como nas sequências *Spin Eco* (SE), que veremos no ponto 4.7, existe um pulso de refocalização de 180°, este vai compensar o decaimento derivado à não homogeneidade de B_0 e as características da imagem devem-se essencialmente às interações *spin-spin*, ou seja, ao decaimento T2. Falamos então em ponderação T2 que conseguimos com selecção de valores de TR e TE longos. Nas sequências Eco de Gradiente, como não existe um pulso de refocalização de 180°, o decaimento depende, não só, das interações *spin-spin* como também da heterogeneidade de campo magnético, das diferenças de susceptibilidade magnética e de desvio químico, falando neste caso numa ponderação em T2*.

O contraste numa imagem RM poderá ser, como acabamos de ver, dependente de características dos tecidos (N(H), T1, T2), da constituição de determinada molécula (Transferência de Magnetização-TM) ou do conteúdo químico (Espectroscopia) que, no seu conjunto, são considerados mecanismos de contraste estáticos. Há também tipos de contraste em RM, que são sensíveis ao movimento dos *spins*, como por exemplo, a difusão e a perfusão, muito importantes no estudo da função tecidual (Huettel, Song, & McCarthy, 2004), mas que, por não se enquadrarem no âmbito deste trabalho, não serão aqui abordados.

4.4 Codificação da informação e formação de imagem

4.4.1 Gradientes

Os gradientes de campo magnético fazem com que haja uma variação do campo magnético ao longo de um determinado eixo. Geralmente, essa variação é linear, crescente ou decrescente produzindo uma variação inferior a 1% (Brown & Semelka, 2003). A intensidade do campo magnético associada a um conjunto de prótons, numa determinada localização espacial (x, y, z) em função do tempo, pode ser dada pela equação seguinte:

$$B(t) = B_0 + G_x(t)x + G_y(t)y + G_z(t)z \quad (4.17)$$

A localização espacial do sinal RM é efectuada, exclusivamente, através da aplicação dos gradientes ao longo das três direcções cartesianas x, y e z. Os gradientes, dependendo da sua função designam-se por:

- Gradiente de selecção de corte
- Gradiente de leitura ou de codificação da frequência
- Gradiente de codificação de fase

Dependendo da orientação dos cortes (coronal, axial, sagital ou oblíquos) cada um dos gradientes é utilizado para efectuar a selecção de corte ou para efectuar a codificação da fase ou da frequência.

A Tabela 4.2, resume a acção de cada gradiente para as diferentes orientações. Para os cortes oblíquos serão utilizados, em simultâneo, dois gradientes quando da aplicação do pulso de RF.

	SELECÇÃO DE CORTE	CODIFICAÇÃO DE FASE	CODIFICAÇÃO DA FREQUÊNCIA
Sagital	X	Y	Z
Axial (corpo)	Z	Y	X
Axial (crânio)	Z	X	Y
Coronal	Y	X	Z

Tabela 4. 2 - Gradientes seleccionados para cada um dos planos ortogonais. Fonte: Hashemi & Bradley (1997) e Westbrook & Kaut (1998).

4.4.1.1 Selecção de corte

Um dos passos iniciais numa sequência de imagens RM é dirigir pulsos de RF para uma determinada zona do espaço, utilizando pulsos de excitação selectivos, em conjunto com a aplicação de gradiente de selecção de corte (Brown & Semelka, 2003; Huettel, Song, & McCarthy, 2004). A direcção do gradiente determina a orientação do corte, enquanto que a amplitude do gradiente em conjunto com algumas características do pulso de RF (por exemplo, a largura de banda do transmissor) determinam a posição e a espessura do corte.

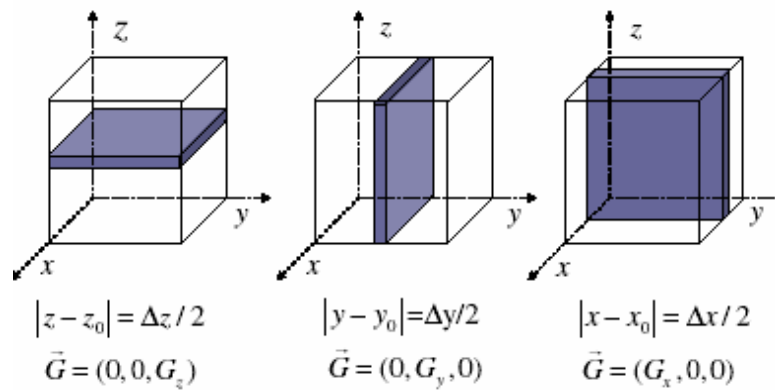


Figura 4. 9 - Coordenadas espaciais para cortes com diferentes orientações. Fonte: Liang & Lauterbur (2000).

Um pulso de RF é caracterizado por uma frequência central e por uma gama de frequências ($\Delta\omega$) designada por largura de banda do transmissor, geralmente situada entre os 1-2 kHz. A frequência central do pulso de RF define a zona a excitar pelo pulso quando o gradiente de selecção de corte é aplicado. Mudando a frequência central é possível efectuar cortes em diferentes posições. Quando se pretendem efectuar vários cortes, a situação mais comum na prática clínica, são utilizados vários pulsos de RF, com a mesma largura de banda mas com frequência central diferente para excitar zonas diferentes do espaço (vários cortes). A espessura de corte é determinada pela amplitude do gradiente e pela largura de banda das frequências contidas no pulso de RF ($\Delta\omega$) utilizado, de acordo com a fórmula:

$$\Delta\omega = \gamma(G_s * Thickness) \tag{4.18}$$

Para uma determinada largura de banda fixa, a espessura de corte pode ser aumentada ou diminuída fazendo variar a amplitude do gradiente, como se pode observar na figura 4.10.

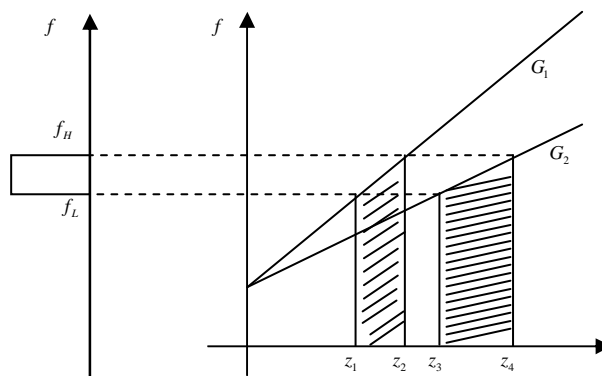
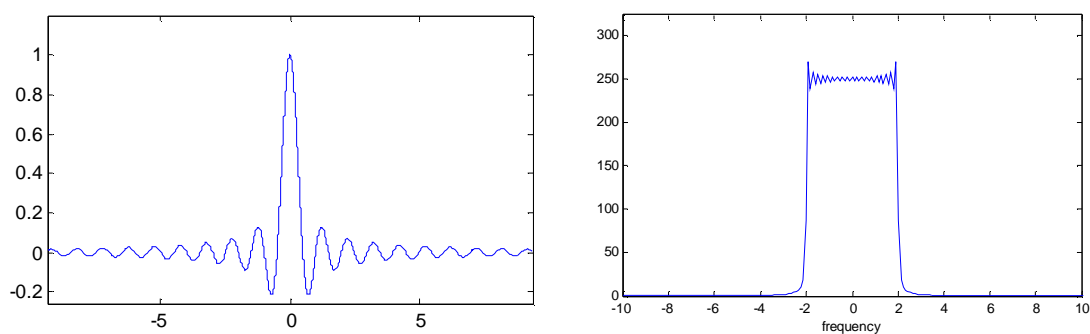


Figura 4.10 - Para uma determinada gama de frequências contidas num pulso de RF, a espessura de corte é determinada pela amplitude do gradiente de selecção de corte. Fonte: Silva (2003).

Para diminuir, por exemplo, uma espessura de corte de 10 mm para 5 mm será necessário uma amplitude do gradiente maior, para a mesma largura de banda. Geralmente, os equipamentos fazem o ajuste da amplitude do gradiente e da largura de banda em função da espessura de corte seleccionada (Westbrook & Kaut, 1998).

Idealmente, seria importante excitar uma área absolutamente rectangular, no entanto daí não resultará um corte absolutamente rectangular como é desejável. Assim, o mais indicado será a utilização de um pulso de RF em que a função modulante seja uma função *sinc*, dado que o seu conteúdo espectral é uma função rectangular. (Brown & Semelka, 2003; Huettel, Song, & McCarthy, 2004; Silva, 2003). A função *sinc*, deverá ser tão próxima do ideal quanto possível, pois só desta forma o perfil espectral associado à selecção de corte será aproximadamente rectangular (Figura 4.11). No entanto, daí derivam alguns problemas práticos uma vez que, obter uma função *sinc* ideal é praticamente irrealizável. Na melhor das hipóteses são possíveis aproximações truncadas, no entanto, se os pulsos forem temporalmente truncados (limitados) não o são espectralmente dando origem a sobreposições espectrais que resultam em artefactos na imagem (*cross-talk*). Estes artefactos poderão ser minimizados recorrendo a um pequeno intervalo (*gap*) entre os cortes ou utilizando modo de aquisição *interleaved*. As oscilações (Fenómeno de Gibbs) no espectro, decorrentes da truncatura do pulso, conduzem a falta de uniformidade nas frequências de excitação ao longo da espessura de corte, conduzindo a limitações na resolução longitudinal (Silva, 2003).



$$F(\text{sinc}(t)) = \text{rect}(f)$$

$$F(\text{rect}(t)) = \text{sinc}(f)$$

Figura 4. 11 - Função *sinc* com um pulso espectral aproximadamente rectangular. Fonte: Silva (2003).

4.4.1.2 Codificação da Frequência e Codificação de Fase

Depois de excitar selectivamente um corte, após definição da sua posição e espessura é necessário fazer a codificação da posição dentro do corte. Para que tal seja possível recorre-se a

duas estratégias de manipulação da oscilação, mediante a aplicação de gradientes lineares no espaço cartesiano (x, y) : (1) o gradiente de codificação da frequência e (2) o gradiente de codificação de fase.

Codificação da frequência

Como se depreende da própria designação, a codificação em frequência cria uma dependência linear entre a localização espacial e a frequência de oscilação. Numa situação ideal em que se considera um objecto unidimensional com uma distribuição $\rho(x)$, o campo magnético sentido, após um pulso de excitação, representa a soma de B_0 e da variação linear do gradiente na direcção x . A variação da frequência de Larmor na posição x será dada por:

$$\omega(x) = \gamma(B_0 + G_x x) = \omega_0 + \gamma G_x x \quad (4.19)$$

Na figura 4.12, ilustra-se o processo através do qual é efectuada a codificação em frequência, que passa por fazer variar a frequência de precessão através da aplicação de um gradiente numa determinada direcção.

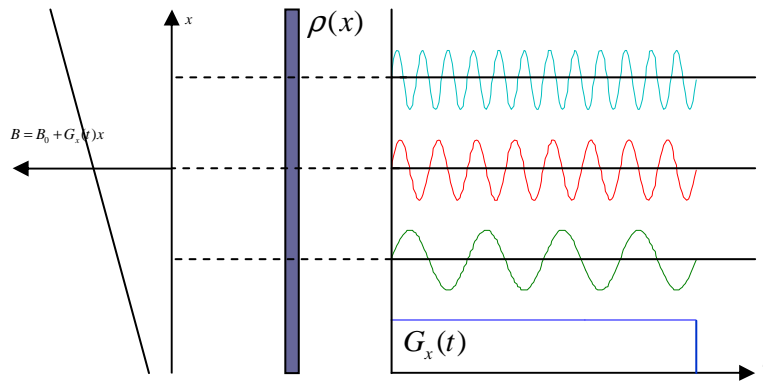


Figura 4. 12 - Codificação em frequência, para um objecto unidimensional com uma densidade de spins $\rho(x)$. Fonte: Silva (2003).

O sinal FID gerado localmente a partir dos spins localizados numa região infinitesimal Δx , centrada em x_n , e omitindo o efeito da relaxação transversal, será dado por:

$$\Delta S(x_n, t) \propto \rho(x_n) \Delta x e^{-i\gamma(B_0 + G_x X_n)t} \quad (4.20)$$

Integrando todas as contribuições obtém-se a expressão 4.21, que representa o sinal codificado em frequência, uma vez que a frequência de oscilação é linearmente dependente da localização espacial:

$$S(t) \propto \sum_{n=1}^N \rho(X_n) \Delta x e^{-i\gamma(B_0 + G_x X_n)t} \quad (4.21)$$

O sinal recebido a partir de todo o objecto na presença do gradiente é uma portadora com amplitude modulada, representada por:

$$S(t) \propto \left[\sum_{n=1}^N \rho(x_n) \Delta x e^{-i\gamma G_x x_n t} \right] e^{-i\omega_0 t} \quad (4.22)$$

Retirando a portadora ao sinal, ou seja desmodulando ter-se-á:

$$S(t) \propto \sum_{n=1}^N \rho(x_n) \Delta x e^{-i\gamma G_x X_n t} \quad (4.23)$$

$$S(K_x(t)) \propto \sum_{n=1}^N \rho(x_n) \Delta x e^{-i2\pi K_x x} \quad \text{e se } K_x(t) = \frac{\gamma}{2\pi} G_x t \quad (4.24)$$

em que K_x , designa a frequência de codificação espacial dependente de t e de G_x . Assim teremos:

$$S(t) \propto T_{\text{Fourier}}(\rho(x)) \quad (4.25)$$

Ou seja, o perfil de $\rho(x)$ pode ser obtido através da função inversa da TF.

Quando se tem um objecto bidimensional este processo é repetido ao longo da direcção y .

Codificação de fase

Como é necessário obter informação numa outra direcção, para codificar um objecto bidimensional é necessário combinar a codificação da frequência na direcção x , com outro tipo de codificação, mediante aplicação de um gradiente G_y na direcção y , durante um intervalo de tempo T_{PE} . Esta codificação designa-se por codificação de fase mas é formalmente idêntica à codificação em frequência. A codificação de fase resulta de uma pré codificação de frequência durante um intervalo de tempo T_{PE} .

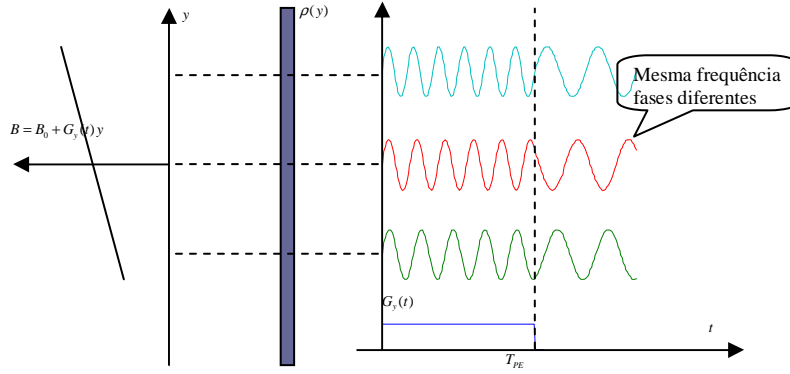


Figura 4. 13 - Codificação de fase, para um objecto unidimensional com uma densidade de spins $\rho(y)$. Fonte: Silva (2003)

De forma idêntica ao observado para a codificação da frequência, consideraremos novamente a situação ideal de um objecto unidimensional. A representação matemática da contribuição para o sinal de uma região Δy , centrada em y_m será:

$$\Delta S(y_m, t) \propto \rho(y_m) \Delta y e^{-i\gamma(B_0 + G_y y_m)t} \quad 0 \leq t \leq T_{PE} \quad (4.26)$$

$$\Delta S(y_m, t) \propto \rho(y_m) \Delta y e^{-i\gamma(G_y y_m T_{PE})} e^{-i\gamma B_0 t} \quad T_{PE} \leq t \quad (4.27)$$

Torna-se evidente que durante o intervalo $0 \leq t \leq T_{PE}$, o sinal está codificado em frequência. Em consequência desta codificação, os sinais provenientes de diferentes posições ao longo da direcção y , acumulam diferentes desvios de fase, depois do intervalo de tempo T_{PE} , como se evidencia na figura 4.13.

Desta maneira, qualquer sinal medido após T_{PE} , terá uma fase inicial $\phi(y)$ que é proporcional a y , e que pode ser representada por:

$$\phi(y) = -\gamma G_y y T_{PE} \quad (4.28)$$

Uma vez que $\phi(y)$ está linearmente relacionado com o sinal na localização y , podemos dizer que o sinal está codificado em fase.

Integrando agora as contribuições de todo o objecto temos que:

$$S(t) \propto \left[\sum_{m=1}^M \rho(y_m) \Delta y e^{-i\gamma G_y y_m T_{PE}} \right] e^{-i\omega_0 t} \quad (4.29)$$

Desmodulando, ficará:

$$S(t) \propto \left[\sum_{m=1}^M \rho(y_m) \Delta y e^{-i\gamma G_x y_m T_{PE}} \right] \quad (4.30)$$

Após T_{PE} , não existirá dependência temporal e desta forma:

$$S(K_y) \propto \sum_{m=1}^M \rho(y_m) \Delta y e^{-iK_y y_m} \text{ e como } K_y = \frac{\gamma}{2\pi} G_y T_{PE} \quad (4.31)$$

(K_y , designa uma frequência decorrente da codificação de fase, dependente de T_{PE} e de G_y)

O que resultará em:

$$S(K_y) \propto T_{Fourier}(\rho(y)) \quad (4.32)$$

que é também uma transformada de Fourier e que permitirá, através da sua inversa, a obtenção do perfil de $\rho(y)$.

De forma resumida, podemos dizer que, a codificação de frequência permite medir os sinais associados ao espaço K, de acordo com as coordenadas $K_x(t) = \frac{\gamma}{2\pi} G_x t$, apresentando dependência temporal, ou seja, durante a aplicação de G_x obtém-se um intervalo de valores K_x .

A codificação de fase permite medir sinais associados ao espaço K de acordo com as coordenadas $K_y = \frac{\gamma}{2\pi} G_y T_{PE}$. Para cada combinação (G_y, T_{PE}) existe um único valor de coordenada K_y . Ou seja, percorre-se o espaço K, linha a linha, como se observa na figura 4.14.

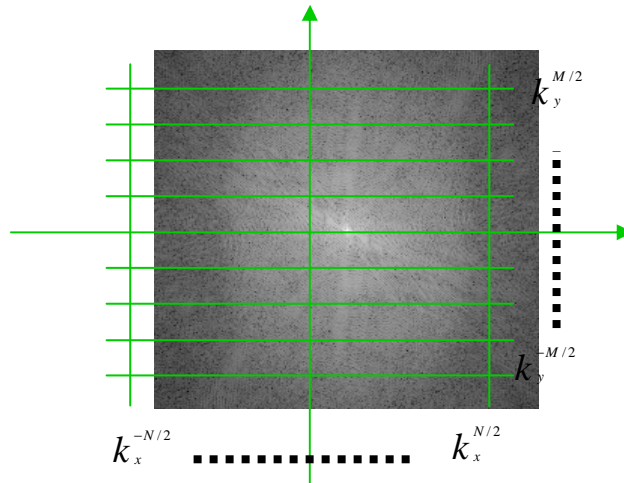


Figura 4. 14 - Preenchimento do espaço K, seguindo uma trajetória linear. Fonte: Silva (2003).

Numa sequência SE, o gradiente de codificação da fase é aplicado no intervalo que decorre entre a aplicação do pulso de excitação de 90° e o Eco. No modo de aquisição convencional, geralmente, é efectuado um passo de codificação de fase por TR. Para discriminar 256 linhas será necessário aplicar 256 passos de codificação de fase (N_y) o que levará um tempo equivalente $256 \times TR$. O gradiente de codificação da frequência é accionado posteriormente durante a fase de leitura do sinal.

O sinal com codificação sequencial em fase e frequência pode ser representado pela expressão

$$S(K_x, K_y) \propto \sum_{n=1}^N \sum_{m=1}^M \rho(x_n, y_m) \Delta x \Delta y e^{-2\pi i(K_y y_m + K_x x_n)} \quad (4.33)$$

$$S(K_x, K_y) \propto T\text{Fourier}2D(\rho(x, y)) \quad (4.34)$$

Ao efectuar-se a amostragem do espaço K, estamos a recolher amostras da TF 2D, relativa à distribuição espacial de $\rho(x, y)$, sendo a reconstrução obtida a partir da implementação numérica da transformada inversa 2D.

Aquisição 3D

Na técnica de aquisição 3D (imagem volumétrica) é necessária a aplicação de um pulso de excitação não selectivo, com o intuito de excitar simultaneamente um grande volume de spins. O volume é frequentemente dividido em partições, aplicando sucessivos passos de codificação da fase nessa direcção, por exemplo, ao longo do eixo dos Z. É aplicada TF3D, de forma semelhante ao que acontece na aquisição 2D. Os volumes obtidos podem ser isotrópicos (mesma dimensão nas três direcções, ou anisotrópicos em que pelo menos uma das dimensões é diferente). Na primeira situação, a resolução é superior permitindo reconstruções com cortes extremamente finos e de igual qualidade em qualquer plano. Ao invés de adquirir (excitar) um corte ou conjunto de cortes, adquire-se informação directa relativa a um volume (ou slab) e a partir deste volume reconstroem-se os cortes nos diversos planos e com as inclinações que forem necessárias. A resolução no eixo dos Z, depende do número de passos de codificação da fase (N_z) nessa direcção (Lafuente, 2000; Bushberg, Seibert, Leidholdt & Boone, 2002).

4.4 Espaço K e o espaço da imagem (FOV)

Depois do preenchimento do espaço K, é necessário a aplicação da TF para que se convertam os dados brutos do espaço K^{11} no espaço da imagem $M(x, y)$. Nestes dois espaços os parâmetros de amostragem são inversamente proporcionais, ou seja, no espaço da imagem a unidade de

¹¹ Note-se que o espaço K é uma outra designação para espaço de Fourier.

amostragem é a distância, no espaço K a unidade de amostragem é dada pelo inverso da distância. Falando de um ponto de vista qualitativo podemos dizer que: a uma maior cobertura no domínio do espaço K corresponde a uma melhor resolução espacial na imagem (menor dimensão do voxel). A periferia do espaço K, ou seja, as altas frequências contribuem para a nitidez da imagem, enquanto que as baixas frequências (centrais) determinam o contraste, como se pode observar na imagem 4.15. Um ponto no espaço K não corresponde directamente a um pixel na imagem. Cada zona do espaço K contém informação relativa a toda a imagem.

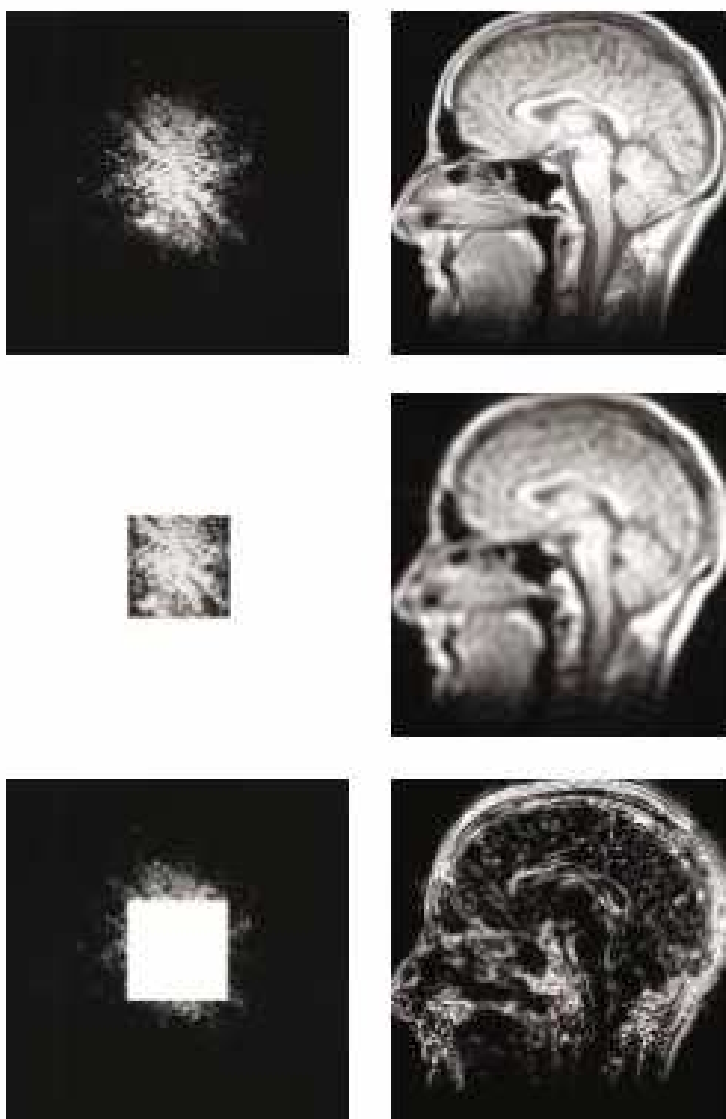


Figura 4. 15 - Espaço K e espaço da imagem. Fonte: Siemens (2003)

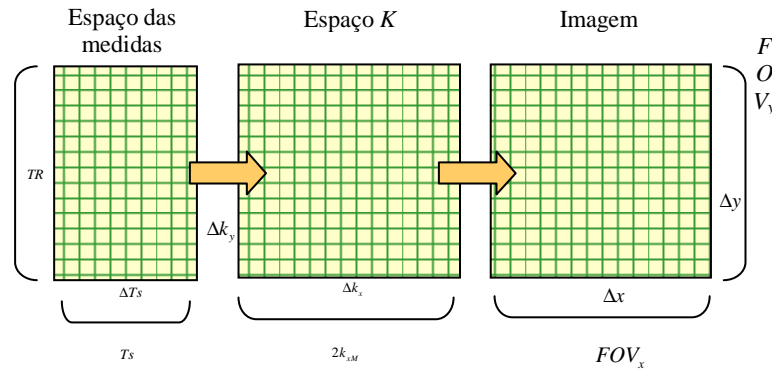


Figura 4. 16 - Relação entre o espaço das medidas, o espaço K e a imagem. Fonte: Silva (2003)

Há uma relação directa entre a gama de frequências de precessão associadas à codificação de frequência, o gradiente e o FOV, como se pode depreender da expressão 4.35.

$$FOV = \frac{BW}{\gamma G_x} = \frac{1}{\gamma G_x \Delta T_s} \quad (4.35)$$

Como, ΔT_s , é o intervalo de amostragem na direcção x , virá:

$$\frac{1}{FOV} = \gamma G_x \Delta T_s \text{ e como, } K_x = \gamma G_x \Delta T_s \text{ virá } FOV_x = \frac{1}{\Delta K_x} \quad (4.36)$$

Ou seja, existe uma relação inversa entre o espaço da imagem e o espaço K.

Uma revisão aprofundada a nível de processamento de sinal em RM, não cabe no contexto deste trabalho, mas pode ser encontrada em Liang & Lauterbur (2000).

4.5. Parâmetros de aquisição e qualidade da imagem

Existem inúmeros parâmetros (primários) que o operador pode seleccionar e que influenciam a qualidade de imagem. A escolha dos parâmetros primários influencia outros parâmetros chamados secundários e que geralmente se utilizam para caracterizar uma imagem (exemplos são a relação Sinal/Ruído (S/R), o contraste na imagem, a resolução espacial, o TA e a cobertura (Hashemi & Bradley, 1997; Westbrook & Kaut, 1998)). De um modo geral, têm que se estabelecer compromissos, manipulando estes parâmetros em função dos objectivos a atingir. Isto porque regra geral, quando melhoramos um parâmetro degradamos outro que também poderá ser importante. No caso particular dos estudos de produção de fala, temos que estabelecer compromissos entre a resolução temporal e a resolução espacial para uma relação S/R adequada. Quando pretendemos estudar a dinâmica dos articuladores ao longo do tempo, precisamos de ter

uma resolução temporal elevada que terá que ser conseguida às custas da degradação da resolução espacial e do contraste na imagem.

Dos **parâmetros primários** podemos referir o TR, o TE, o TI e o FA como parâmetros que influenciam o contraste na imagem. A espessura de corte e o intervalo entre os cortes influenciam a cobertura. A dimensão do FOV nas duas direcções, o número passos de codificação da fase (N_y) e da frequência (N_x) determinam a resolução espacial. Estes últimos em conjunto com o número de aquisições (NEX - Number of Excitations), e com a BW influenciam a S/R.

4.5.1 Relação Sinal-Ruído (S/R)

É dos parâmetros mais importantes na imagem. De pouco adiantará um boa resolução espacial se a S/R for muito baixa. O ruído tem diversas fontes mas, para um determinado paciente, ele pode considerar-se aproximadamente constante (Westbrook & Kaut, 1998). Assim, a alteração de alguns parâmetros conduzirá ao aumento ou diminuição da S/R pelo aumento ou diminuição do sinal. Existe uma fórmula importante em RM (4.37), que nos permite observar: (1) que parâmetros e (2) como influenciam a S/R (Hashemi & Bradley, 1997).

$$S/R \propto (\text{volume voxel}) \sqrt{(N_y)(NEX)/BW} \quad (4.37)$$

O voxel tem as dimensões Δx , Δy e Δz , onde Δz representa a espessura de corte nas aquisições 2D. Quanto maior o número de *spins* num voxel, maior o sinal, e em consequência quanto maior o tamanho do voxel, maior a S/R. Mantendo constantes todos os outros parâmetros, haverá com o aumento do voxel uma melhoria na S/R e uma degradação da resolução espacial que poderá ser de maior ou menor impacto em função do que se pretenda avaliar. Quando aumentamos o NEX, ou o número de passos de codificação da fase, aumenta-se também a S/R. Duplicando NEX e (N_y), aumentamos a S/R por um factor de $\sqrt{2}$. Saliente-se, no entanto, que o aumento do número de passos de codificação da fase, mantendo o FOV constante, resultará numa diminuição da S/R. Algum cuidado se deverá ter quando se analisa a fórmula sem que se tenha em conta, também, o volume do voxel. Com a largura de banda, e decorrente da fórmula, verifica-se uma relação inversa com a S/R. Assim, diminuindo a BW por um factor de 2, aumenta-se a S/R por um factor de $\sqrt{2}$. A diminuição da BW, provoca um aumento no sinal, mas terá como consequência um aumento do desvio químico, o que poderá não ser desejável. O TE terá que ser aumentado pois, o tempo de amostragem terá que ser maior, para diminuir a BW ($BW = N_x/T_s$). O número de cortes permitido, para um determinado TR, diminui pois o TE é maior. Além destes factores, também a DP na área a examinar e o tipo de antena receptora

utilizada têm influência na IS e conseqüentemente na S/R (Elster & Burdette, 2001; Hashemi & Bradley, 1997).

Numa aquisição 3D, em regra, a S/R também deverá ser superior, uma vez que, além de todos os outros factores existe ainda um passo de codificação de fase adicional ao longo da direcção Z (N_z). Numa aquisição 3D a S/R será dada pela equação 4.38.

$$S/R \propto \sqrt{N_z} \times S/R(2D) \quad (4.38)$$

4.5.2 Resolução espacial

A resolução espacial que poderemos definir, de uma forma simplificada, como a capacidade de um sistema em discriminar dois pontos próximos na imagem é determinada essencialmente pela dimensão do pixel e pelo número de passos de codificação da fase. À distância mínima (d) a que discriminamos dois pontos na imagem chamamos distância de resolução ($d = 1/r$). De um modo geral, os parâmetros que conduzem à melhoria da resolução espacial conduzem à degradação da S/R (como já referimos) ou a um aumento do tempo de aquisição. Assim, e por exemplo se para melhorar a resolução espacial (pixel menor) aumentamos o número de passos de codificação da fase (N_y), mantendo o FOV haverá uma diminuição da S/R. Não esquecer que dimensão do pixel é dada pela fórmula 4.39.

$$pixel_y = FOV_y / N_y \quad (4.39)$$

Se aumentarmos N_y e o FOV, a resolução mantém-se, a S/R pode ser aumentada, mas haverá um aumento do TA, proporcional a N_y .

4.5.3 Cobertura e Tempo de aquisição

A cobertura pode ser definida como a distância abrangida com uma técnica de aquisição multicorte. Esta depende do número de cortes, da espessura de corte e do intervalo (*gap*) entre os cortes. Como o número de cortes possível, depende do TR e do TE, segundo a fórmula 4.40.

$$N^{\circ} \text{ cortes} = TR / (TE + T_s / 2 + T_o) \quad (4.40)$$

assim, também a cobertura será dada por:

$$\text{Cobertura} = TR / (TE + T_s / 2 + T_o) \times (\text{espessura} + \text{gap}) \quad (4.41)$$

(Hashemi & Bradley, 1997; Westbrook & Kaut, 1998).

Em relação ao **tempo de aquisição** já vimos que depende do TR, do NEX e de N_y , numa sequência convencional. Numa sequência TSE este tempo é dividido pelo factor turbo e, numa aquisição 3D corresponde a $N_z \times TA$ (2D).

4.6 Artefactos

Os artefactos são uma presença frequente nas imagens RM sendo necessário identificá-los. É importante ter conhecimento da sua origem de forma a poder minimizá-los. Alguns artefactos podem ser eliminados, outros apenas minimizados alterando e corrigindo alguns parâmetros. Os artefactos podem ser agrupados em função da sua origem. Assim, podem resultar de: (1) Perturbações do campo magnético, (2) Artefactos de RF e relacionados com os gradientes, (3) Movimento e Fluxo, (4) Processamento e mapeamento e ainda relacionados com outras causas. No âmbito deste trabalho abordaremos aqueles que no contexto dos estudos relacionados com a produção de fala poderão ser mais frequentes ou que condicionem a qualidade de imagem.

4.6.1 Artefactos de susceptibilidade magnética

A susceptibilidade magnética (χ) representa a capacidade de uma substância se magnetizar na presença de um campo magnético. Os tecidos apresentam diferentes graus de magnetização dando origem a diferenças na frequência de precessão (Hashemi & Bradley, 1997; Caseiro Alves, 1998). Os artefactos de susceptibilidade magnética ocorrem como consequência de variações microscópicas na Intensidade de campo magnético na interface de substâncias com diferente susceptibilidade magnética (interfaces ar/tecido e tecido/gordura como, por exemplo, no pulmão, na naso-faringe, seios face, base crânio etc). Podem ocorrer também quando há objectos ferromagnéticos dentro de materiais diamagnéticos (p.e. alguns implantes, corpos estranhos). Surgem também em áreas de hemorragia recente, devido à concentração local elevada de hemoglobina podendo nestes casos contribuir para o diagnóstico da hemorragia. Quando se utilizam altas concentrações de contrastes superparamagnéticos (a ingestão de óxidos de ferro como agentes de contraste oral) pode também existir perturbação do campo magnético local, dando origem a artefactos. Os objectos ferromagnéticos têm uma susceptibilidade magnética muito elevada e a consequência para a imagem traduz-se pelo aparecimento de áreas muito brilhantes e áreas escuras (vazio de sinal), com distorção espacial das estruturas anatómicas circunvizinhas, com degradação da imagem (Hashemi & Bradley, 1997; Ester & Burdette, 2001).

Estes artefactos são mais acentuados em sequências com TE muito elevado e nas sequências EG, uma vez que a reversão do gradiente não compensa a diferença de fase na interface das estruturas. Este artefacto poderá, nos estudos de produção de fala mostrar-se bastante condicionante na medida em que a escolha dos informantes estará dependente do tipo de material presente nos tratamentos dentários. Se nas sequências mais frequentemente utilizadas para os

estudos estáticos (por exemplo, SE, TSE) estes artefactos são mínimos, já nas sequências de aquisição mais rápida para obtenção de imagens em tempo real (geralmente EG) podem ser de uma ordem de grandeza que não permita a visualização das estruturas em estudo, e em particular de alguns articuladores cujos movimentos têm que ser observados.

4.6.2 *Aliasing* ou *Wrap around* (2D e 3D)

É possível reconstruir um sinal contínuo a partir de um determinado número discreto de amostras. O número de amostras mínimo ou frequência de amostragem é determinado pelo Teorema de *Nyquist*. Segundo este teorema, a frequência mínima de amostragem, não deverá ser inferior a pelo menos o dobro da frequência mais alta contida no sinal (Elster & Burdette, 2001; Hashemi & Bradley, 1997; Liang & Lauterbur, 2000). Se esta frequência de amostragem, não for respeitada dará origem a um artefacto denominado de *aliasing* de frequência e ocorre ao longo desta direcção. Nos equipamentos mais recentes dificilmente é perceptível uma vez que se utilizam técnicas de sobre-amostragem (*oversampling*) e filtros passa-banda (Elster & Burdette, 2001).

O artefacto de *aliasing* na prática diária é mais frequente e sério ao longo da direcção de codificação da fase e resulta da selecção de um FOV, na direcção da fase, que é mais pequeno do que as estruturas que se querem estudar. Em termos físicos, o artefacto resulta do facto de ser necessário um determinado número de ciclos de fase para cada passo de codificação da fase, para abarcar o FOV seleccionado. Para o primeiro passo de codificação existem desvios de fase entre 0° e 360° , ao longo do FOV. Qualquer estrutura anatómica que esteja fora do FOV, apresenta um desvio de fase menor que 0° , ou superior a 360° . Se, por exemplo, tiver um desvio de fase entre os 361° e 450° e, como todas as frequências com significado, foram definidas entre 0° e 360° , um desvio de fase de 361° corresponderá a uma posição espacial de 1° , e um desvio de 450° corresponderá a $450^\circ - 360^\circ$, ou seja, 90° . Desta maneira, a parte anatómica adjacente aos bordos do campo de visão é projectada (“enrolada”) do outro lado da imagem (Elster & Burdette, 2001; Hashemi & Bradley, 1997).



Figura 4. 17 - Artefacto de *aliasing* (2D FT) na direcção da fase onde se observa o nariz projectado posteriormente.

A correcção deste artefacto passa por aumentar o FOV na direcção de codificação da fase (diminui a resolução espacial) ou *oversampling* (aumentar Ny) às custas do aumento do tempo de aquisição, trocar a direcção da fase com a frequência se possível, utilização de bandas de saturação exteriores ao FOV ou quando adequado, utilizar antenas de superfície.

Aliasing 3D

Nas aquisições em volume (3D FT) também pode ocorrer *aliasing* ao longo da direcção de codificação da fase (na direcção do *slab*). O artefacto ocorre entre cortes na periferia da partição 3D, geralmente com sobreposição dos primeiros com os últimos cortes do volume. Para minimizar este artefacto, tem que se seleccionar um FOV um pouco maior do que a dimensão das estruturas a observar nessa direcção. Para manter a resolução espacial, com o aumento do FOV é necessário aumentar os passos de codificação da fase o que se reflecte num aumento nem sempre desejado do TA (Elster & Burdette, 2001; Hashemi & Bradley, 1997).

Na Figura 4. 18, pode observar-se o artefacto entre os cortes periféricos de um volume adquirido no âmbito deste estudo, no falante PAA.

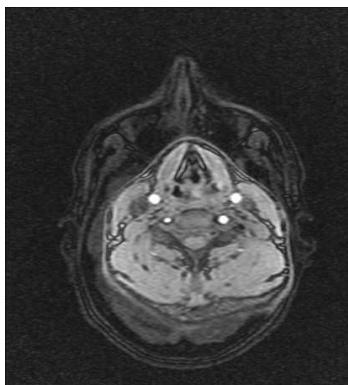


Figura 4. 18 - Artefacto de "enrolamento" numa aquisição 3D - observa-se sobreposição dos últimos cortes do volume obtido, sobre os primeiros (Falante PAA).

4.6.3 Slice overlap

Semelhante ao artefacto de excitação cruzada¹². Resulta numa perda de sinal quando se efectuam múltiplas ranges de cortes, com diferentes orientações. A perda de sinal atravessa a imagem horizontalmente, sendo mais pronunciada na parte posterior. Nos estudos de produção de fala aparece com frequência quando se efectuam, na mesma aquisição cortes axiais, coronais e coronais oblíquos (cortes perpendiculares ao tracto vocal) para obtenção das áreas transversais.

¹² O artefacto de excitação cruzada resulta do facto dos pulsos de excitação não serem perfeitamente rectangulares, no domínio espacial, conduzindo a uma sobreposição espectral a nível dos cortes adjacentes. Para minimizar este artefacto deve-se utilizar um pequeno intervalo entre os cortes.

4.7 Sequências de Pulso

Uma sequência de pulsos é um conjunto pré-definido de aplicação de pulsos RF e de gradientes, repetido um determinado número de vezes durante a aquisição das imagens. O tempo entre os pulsos, a amplitude e a forma do gradiente controlam o sinal RM e afectam as características das imagens. As sequências de pulso não são mais do que programas que controlam todos os aspectos *do hardware* relacionados com a aquisição (Elster & Burdette, 2001; Hashemi & Bradley, 1997; Hornack 1996-2006).

Para quem trabalha com RM, é fundamental ter conhecimento das sequências de pulso pois é através da sua manipulação que se conseguem diferentes tipos de contraste nas imagens RM. Usualmente, para descrever as sequências de pulso são importantes parâmetros como o TR e TE, que já definimos atrás, o Tempo de Inversão (TI) nas sequências de Inversão da Recuperação (IR) e o *Flip Angle* (FA) nas sequências eco de gradiente (EG). Existem, hoje em dia, um vasto número de sequências sendo por vezes difícil a sua categorização e a sua normalização, uma vez que cada fabricante utiliza nomes diferentes, os acrónimos, para sequências que são basicamente idênticas (Marques, 2003).

Podemos dividir as sequências de pulso em três grandes grupos: (1) Sequências Eco de *Spin* (ES) ou *Spin* Eco (SE), (2) Sequências EG e (3) Sequências híbridas (Marques, 2003). As sequências IR poderão ser consideradas como casos especiais das SE (Brown & Semelka, 2003; Pooley, 2005). Tendo por base estas sequências desenvolveram-se modos de aquisição mais rápidos existindo, nos dias de hoje, sequências Turbo ou Fast Spin Eco (TSE ou FSE), sequências Turbo GE ou Fast GE e sequências Turbo ou Fast IR.

4.7.1 Sequência Spin-Eco Convencional

A sequência SE é uma das sequências mais conhecidas, tendo sido das mais utilizadas na prática clínica. Apesar de, hoje em dia, existirem estratégias de aquisição mais rápidas, ainda são utilizadas, devido a algumas vantagens sobre outras sequências a que nos referiremos mais adiante.

Basicamente, numa sequência SE, após a aplicação do pulso de excitação de 90° são aplicados um ou mais pulsos de refocalização de 180° . Usualmente, utilizam-se sequências com um pulso de 180° para uma ponderação em T1 e dois pulsos para obter imagens com ponderação DP e T2 (Lafuente, 2000). O efeito do pulso de refocalização de 180° é fazer com que os *spins*, que após a cessação do pulso de 90° , desfasam rapidamente no plano transversal, voltem a precessar em fase, atingindo o máximo do sinal (eco ou ES). Ao intervalo de tempo entre a aplicação do pulso de 90° e o pulso de 180° decorre o tempo τ e ao intervalo de tempo entre o pulso de 180° e o

momento de máximo refasamento ou eco corresponde também a um tempo τ (Figura 4.19). O Tempo de Eco corresponde portanto a 2τ , uma vez que, e por definição o TE é o intervalo entre a aplicação do pulso de 90° e a medição do sinal. (Hashemi & Bradley, 1997).

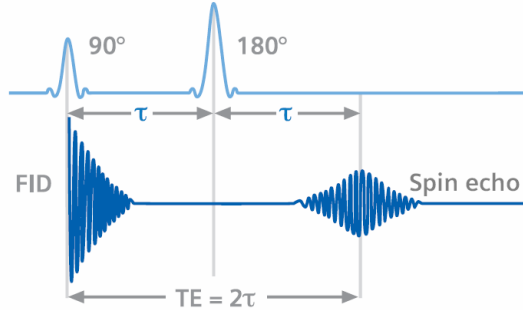


Figura 4.19 - Tempo de Eco. Fonte: Siemens (2003).

Na Figura 4.20, pode observar-se o diagrama relativo a esta seqüência de pulso onde surge também o esquema de aplicação dos gradientes.

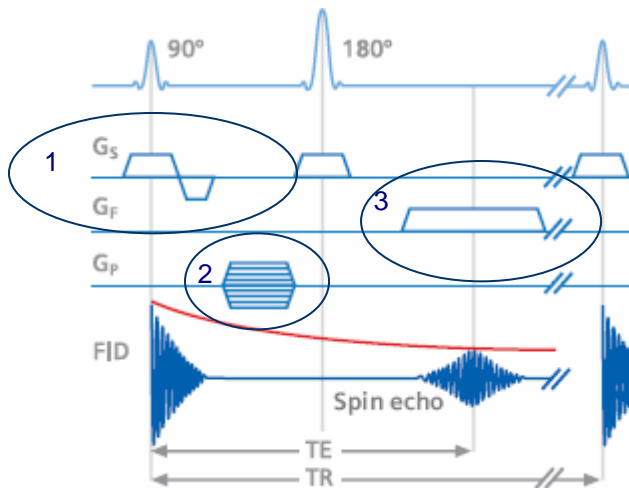


Figura 4.20 - Diagrama de pulso de uma seqüência SE. Adaptado de: Siemens (2003).

Imediatamente a seguir à aplicação do gradiente de selecção de corte (G_s) é introduzido um gradiente com direcção negativa mas com metade da duração, com o intuito de refocalizar os *spins* (1). Sempre que se aplica um gradiente há desfaseamento, neste caso o desfaseamento é efectuado com o intuito de definir ou seleccionar um corte. Com a aplicação do pulso de 180° este gradiente volta a ser aplicado, desta maneira só os *spins* dentro do corte previamente estimulado sentem o efeito do pulso de 180° . O gradiente de codificação da fase G_y ou G_p (2), é aplicado entre o gradiente de selecção de corte (após o pulso de 90°) e o ES. O gradiente de leitura ou de codificação da frequência é aplicado para obter informação necessária para preencher uma linha do espaço K, no corte seleccionado. Este gradiente é aplicado durante o TE, ou seja durante a leitura do sinal (3). No entanto, se o gradiente fosse simplesmente aplicado enquanto se faz a

leitura do sinal, o mais provável seria um desfasamento total dos *spins* no momento da leitura (meio do eco), com perda de sinal. Na realidade, antes da aplicação do gradiente de leitura é aplicado um gradiente com direcção negativa e com uma área igual a metade do gradiente de leitura. Este gradiente faz com que os *spins* fiquem desfasados, ficando em fase no centro do eco quando é medido o sinal. A duração do gradiente de leitura corresponde ao intervalo de tempo em que se faz a colheita do sinal $-T_x$ (Hashemi & Bradley, 1997).

Na sequência CSE todo este esquema decorre dentro do tempo TR, e é repetida tantas vezes quanto o número de passos de codificação da fase N_y ou N_p . Por cada TR é preenchida uma só linha do espaço K, pelo que o TA numa sequência CSE, é dado pela fórmula:

$$TA = TR \times N_y \times NEX \quad (4.42)$$

Uma das grandes vantagens desta sequência é a sua grande versatilidade permitindo obter imagens com ponderação T1, T2 e DP, sendo o contraste nas imagens facilmente reconhecido. Nas imagens ponderadas em T1 (TR e TE curtos) a gordura apresenta hipersinal, os líquidos sinal intermédio ou hiposinal (preto), os músculos e órgãos como o fígado, baço e rins apresentam sinal intermédio (tons de cinzento) ou hiposinal. Com as sequências ponderadas em T1 obtém-se uma boa discriminação das estruturas anatómicas. Nas sequências com ponderação T2 (TR e TE longos) a gordura perde sinal, os líquidos brilham intensamente com IS superior à da gordura. As estruturas musculares e o fígado diminuem ligeiramente a IS em T2, o baço e os rins aumentam de sinal. De um modo geral, os tecidos patológicos (tumores por exemplo) apresentam hiposinal em T1 e diversos graus de hipersinal em T2, pois apresentam T1 e T2 longos. Estruturas como ligamentos e tendões, a cortical óssea, estruturas calcificadas (como os dentes) e o ar apresentam vazio de sinal em ambas as ponderações (Marques, 2003).

Na sequência SE a relação S/R é, de um modo geral, boa. Com a aplicação do pulso de refocalização de 180° elimina-se o desfasamento causado pela heterogeneidade do campo magnético externo. É também, pouco sensível a pequenas diferenças de susceptibilidade magnética quando comparada com as sequências EG. Como grande desvantagem aponta-se o TA demasiado longo, em particular nas ponderações DP e T2 (geralmente 7 a 10 minutos). Os TR utilizados são demasiado longos, tornando esta sequência incompatível com o modo de aquisição 3D, com estudos dinâmicos ou estudos em apneia muito utilizados, por exemplo, nos exames às áreas abdominais. É também muito sensível ao movimento e fluxo o que pode originar alguns artefactos. O SAR (*Specific Absorption Rate*), que está relacionado com a absorção de RF pelos tecidos do corpo humano, é mais elevado nas sequências SE do que nas sequências EG devido ao pulso de refocalização de 180° (Lafuente, 2000; Marques, 2003).

4.7.2 Sequências de Inversão da Recuperação

A sequência IR começa com a aplicação de um pulso de 180° ao qual se segue, após um intervalo de tempo (Tempo de Inversão ou TI), um pulso de excitação RF de 90°. A escolha do TI, determina o intervalo de tempo que é permitido para a relaxação, possibilitando assim, a supressão do sinal de alguns tecidos, em particular, a gordura e o LCR. As variantes mais utilizadas das sequências de IR são o STIR (*Short TI Inversion Recovery*) para a saturação da gordura e o FLAIR (*Fluid Attenuated Inversion Recovery*) para o LCR. No entanto, qualquer tecido poderá ser saturado se o TI for seleccionado de acordo com a fórmula (4.43) e se o TR for significativamente superior ao T1 do tecido a saturar.

$$TI(null) = 0,693 \times T1_{\text{tecido}} \quad (4.43)$$

Para a saturação da gordura utiliza-se um TI curto uma vez que o T1 da gordura é também curto. Para campos de 1,5 T o TI é de aproximadamente 150 ms. Não é adequada a sua realização após administração de produtos de contraste paramagnéticos pois pode haver supressão do sinal da lesão (T1 curtos).

O FLAIR (Figura 4.21) é muito utilizado na área da neuro-radiologia sendo hoje uma sequência de rotina na avaliação da esclerose múltipla, fundamental na detecção de pequenas lesões periventriculares e em lesões da substância branca na coluna vertebral etc (Marques, 2003).

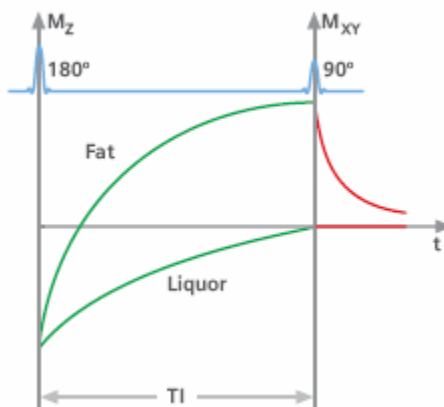


Figura 4. 21 - Diagrama de pulso referente a uma sequência IR, neste caso a sequência FLAIR ou *dark-fluid*. Fonte: Siemens (2003).

Ao aplicar-se o pulso de 180°, vai haver uma inversão da magnetização longitudinal. Os diferentes tecidos recuperam a magnetização a diferentes velocidades. Ao fim de um intervalo de tempo adequado (TI) o sinal proveniente do tecido que se pretende saturar (LCR, neste caso), vai passar pelo eixo do zero e nesse momento é aplicado um pulso de 90°. Após a aplicação do pulso de 90° os *spins* de todos os outros tecidos precessam no plano transversal à excepção do tecido que foi saturado, não contribuindo para o contraste na imagem (Pooley, 2005; Siemens, 2003).

A sequência IR propriamente dita permite obter imagens com forte ponderação em T1 permitindo uma excelente discriminação anatômica sendo muito utilizada, por exemplo, em pediatria uma vez permite melhorar o contraste substância branca/cinza. A grande limitação destas sequências é o tempo de aquisição demasiado longo (o TR é longo) para uma cobertura muito limitada (número cortes permitido é pequeno). Hoje em dia, são geralmente aplicadas em conjunto com um esquema TSE (Turbo IR) de modo a diminuir o TA (Marques, 2003).

4.7.3 Sequências Eco de Gradiente (EG)

As sequências de EG são, no presente, uma componente essencial de um exame de RM. Existe uma grande variedade de sequências EG, sendo muito versáteis. A sua aplicação é vasta, podendo incluir a imagem cardíaca, aquisição volumétrica (3D), EPI, estudos dinâmicos após contraste, imagem em tempo real e muitas outras. A sua classificação não é simples havendo alguma controvérsia na bibliografia. As características da imagem, em termos de ponderação, nem sempre são óbvias (Elster & Burdette, 2001). Nas sequências EG, ao contrário das SE, não se aplicam pulsos de refocalização de 180° . Após o pulso de excitação de RF, é aplicado um gradiente de reversão ou inversão que causa o desfasamento dos *spins* (gradiente negativo) no plano transversal e permite, posteriormente, o seu refasamento (gradiente positivo) para que se produza um eco, neste caso um eco de gradiente (Figura 4.22).

A ação deste gradiente só se faz sentir nos *spins* que ele próprio desfasou, não compensando diferenças de fase causadas pela heterogeneidade do campo, susceptibilidade magnética ou efeitos de desvio químico. Nestas sequências falamos em decaimento $T2^*$ e não T2 uma vez que o desfasamento não se deve apenas às interações *spin-spin* (Elster & Burdette, 2001; Marques, 2003; Bitar et al, 2006).

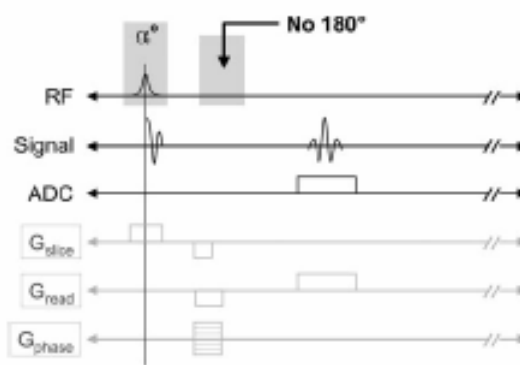


Figura 4. 22 - Diagrama de pulso de uma sequência EG. Fonte: Brown & Semelka (1999).

Como só é aplicado um pulso de excitação de RF, o eco pode ser produzido mais rapidamente. Desta maneira, o TE pode ser significativamente mais curto nestas sequências do que nas SE. Outra particularidade, é que nestas sequências utilizam-se FA parciais (em geral, FA menores que 90°). Com a utilização de FA muito pequenos (10° , por exemplo) é possível ter TR muito curtos.

Com TR e TE curtos é possível diminuir o TA. De salientar, que a diminuição significativa dos tempos de aquisição representa uma das maiores vantagens destas sequências. Nas sequências EG o valor do FA é, também, determinante na ponderação obtida (Elster & Burdette, 2001; Lafuente, 2000).

A utilização de FA parcial faz com que o sinal RM seja mais forte. Parece paradoxal como é que, utilizando um FA pequeno (menor fracção da componente da magnetização longitudinal deflectida no plano transversal) se pode vir a obter um sinal mais forte do que o conseguido com pulsos de 90°. Se se utilizasse apenas um pulso de excitação assim seria. No entanto, numa sequência utilizam-se n pulsos de RF. Estes pulsos, principalmente se forem muito próximos, tendem a reduzir a fracção da magnetização longitudinal M_z utilizada para criar o sinal. Ou seja, a fracção M_z deflectida no plano transversal é menor, mas diminui menos a magnetização longitudinal. Do equilíbrio entre estes dois factores resulta um sinal mais forte nestas sequências (Elster & Burdette, 2001).

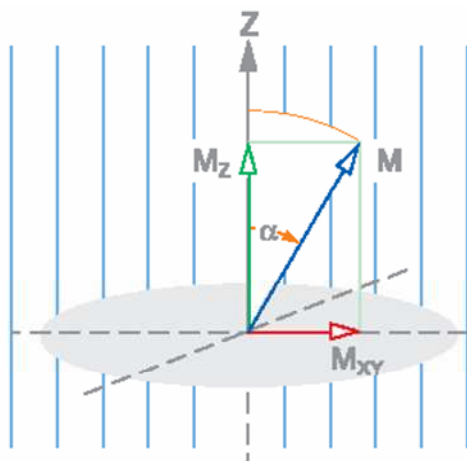


Figura 4. 23 - Flip angle parcial (α). Fonte: Siemens (2003).

O valor do FA que maximiza o sinal para um determinado tecido depende do tipo de sequência de pulsos utilizada e dos parâmetros de imagem. Nas sequências em que não se estabelece o *Steady State*, veremos mais adiante quando se estabelece este estado, o sinal é maximizado para um ângulo, chamado Ernst angle (α_E). Para determinar esse valor pode-se recorrer à fórmula seguinte:

$$\cos \alpha_E = \exp(-TR/T1) \quad (4.44)$$

Como se pode observar, o FA ideal será 90°, apenas se $TR \gg T1$ (Elster & Burdette, 2001; Lafuente, 2000).

Com estas sequências (modulando o TR, TE e FA) pode obter-se contraste semelhante ao das sequências SE (T1, DP) com tempos de aquisição significativamente mais curtos (Marques, 2003). Nesta situação é mais correcto falar em contraste T2* e não T2. Há uma maior sensibilidade a efeitos de susceptibilidade magnética, que poderá ser uma desvantagem quando em presença de alguns implantes metálicos, mas que poderá ser utilizada de forma vantajosa no diagnóstico de pequenos cavernomas ou focos hemorrágicos.

4.7.4 Sequências Híbridas : GRASE

A sequência GRASE - o acrónimo de Gradient and *Spin* Echo é uma técnica híbrida em que é possível obter uma série alternada de Ecos de Gradiente e Ecos de *Spin* a partir de um conjunto de pulsos de RF. Pode ser considerada uma variante da sequência TSE, de que falaremos no ponto seguinte, em que alguns ES's são substituídos por EG permitindo ainda maior rapidez na aquisição. Por cada pulso de refocalização de 180° é colectado um ES e vários EG. O contraste é determinado pelo TE efectivo e pela razão de EG e ES. As características da imagem são mais próximas das sequências EG, sendo mais sensíveis a artefactos de susceptibilidade magnética e de desvio químico. No entanto, o SAR é menor do que nas sequências TSE. Se a homogeneidade de campo não for boa, esta técnica tem aplicação limitada devido à presença de artefactos (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001).

4.8 Sequências e técnicas para aquisição rápida (Fast Imaging)

A rapidez com que se consegue adquirir uma imagem é, cada vez mais, um factor crucial em grande parte das aplicações clínicas e em investigação, quer com RM quer com outras técnicas de imagem. Torna-se necessário obter um número elevado de *frames* ou imagens por segundo para monitorizar com precisão alguns processos fisiológicos dinâmicos tais como a contracção cardíaca ou a perfusão tecidual (Sodickson, 2006). Como veremos no próximo capítulo desta dissertação, também, para estudar o movimento dos diversos articuladores durante a produção de fala é necessária uma resolução temporal elevada o que implica um número elevado de imagens por segundo e um aproveitamento quase até ao limite das capacidades técnicas dos equipamentos dentro dos limites de segurança estabelecidos.

4.8.1 Sequência Turbo ou Fast Spin Eco (TSE ou FSE)

Esta sequência é uma versão optimizada da técnica RARE (*Rapid Acquisition Relaxation Enhancement*) originalmente proposta por Hennig e colaboradores em 1986 (Elster & Burdette, 2001; Hashemi & Bradley, 1997; Marques, 2003). Após um pulso de excitação de RF de 90° segue-se a aplicação de um conjunto de pulsos de refocalização de 180° que dão origem a um trem de ecos.

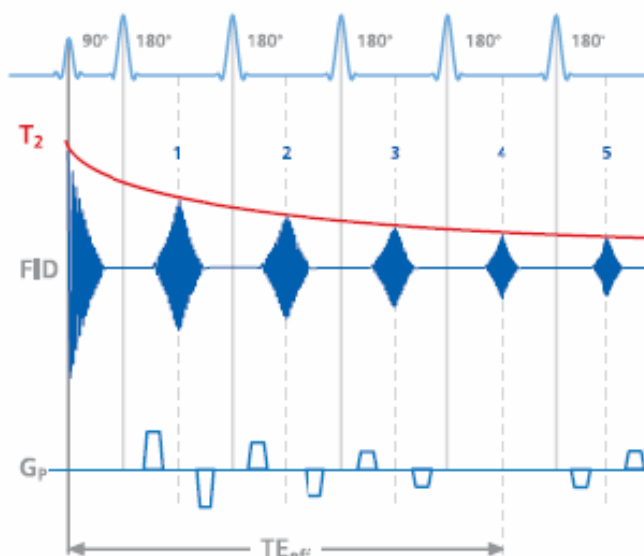


Figura 4. 24 - Diagrama de pulso de uma sequência TSE. Fonte: Siemens (2003).

É uma técnica multi-eco em que cada ES produzido é codificado com fase diferente. Como resultado da mudança de fase (pela aplicação do gradiente) entre ecos, podem ser preenchidas várias linhas do espaço K por TR, encurtando o TA. O número de ecos por TR designa-se por *Echo Train Length* (ETL) ou *turbo Factor* (Brown & Semelka, 2003; Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001; Marques, 2003). Se o factor turbo for 8, o tempo de aquisição será reduzido por um factor de 8. O factor turbo poderá ir de 3 a 256, no entanto, quanto maior o factor turbo, menor o número de cortes por TR permitidos. Os valores para o TR são em geral bastante mais elevados do que na sequência CSE. Valores de TR de 4000ms e de TE de 120 ms são frequentes nas sequências TSE. O TA nas sequências TSE é dado por:

$$TA = TR \times N_y \times NEX / ETL \quad (4.45)$$

O espaçamento entre os ecos, *echo spacing*, é da ordem dos 4 a 12 ms. Com um espaçamento entre ecos mais curto, a duração do trem de ecos é menor, diminuindo o esborratamento na imagem característico desta sequência (Carroll, Sakaie, Wielopolski, & Edelman, 2006).

O contraste na imagem é determinado pelo TE efectivo, sendo possível obter ponderação T1, DP e T2 de uma forma semelhante às sequências SE. Uma grande diferença para as sequências CSE é o hipsinal da gordura na ponderação T2 (resultante de interacções intramoleculares conhecidas como *J coupling*) razão por que frequentemente se efectua, em conjunto, uma saturação selectiva da gordura (*Fat Sat*). As grandes vantagens das sequências TSE, além da óbvia diminuição do TA, são a diminuição dos efeitos de susceptibilidade magnética quando comparadas com as sequências EG e o facto de tornarem possível modo de aquisição 3D TSE com ponderação em T2, com um TA aceitável, algo completamente inviável com as sequências ES convencional. O modo de aquisição 3D, torna possível a obtenção de cortes mais finos (menor

volume parcial) e voxel isotrópico com capacidade para obter reconstruções de qualidade em qualquer plano.

Como se referiu, nas sequências TSE e uma vez que o preenchimento do espaço K é mais rápido, pode-se diminuir o TA ou em alternativa para um tempo idêntico melhorar em muito a resolução espacial com o aumento da matriz (é possível utilizar matrizes de 512x512) e diminuição da espessura de corte. Como se pode aumentar as NEX, a S/R é boa mesmo com cortes finos. Uma das grandes desvantagens, e que pode ser condicionante, é um aumento da deposição de RF, com aumento do SAR, devido ao maior número de pulsos de 180° utilizados. Os efeitos transferência de magnetização (TM) e o esborratamento observado, em particular com TE curto são outras das desvantagens (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001; Hashemi & Bradley, 1997).

HASTE (Half- Fourier single shot turbo spin echo)

Variante da sequência TSE, com a particularidade de o espaço K ser adquirido após a aplicação de um único pulso de excitação, seguido de um número elevado de pulsos de 180° (*single-shot*). Para tornar a aquisição ainda mais rápida só metade do espaço K (half-Fourier) é adquirido (são necessários metade dos passos de codificação da fase). Se for necessário preencher 128 linhas serão necessários 128 ETL (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001). Com esta técnica é possível obter imagens em menos de 1s. Técnica muito utilizada no estudo das vias biliares - colangiopancreatografia por RM (CPRM) para obtenção de cortes espessos (30 a 80 mm em 2s) ou multicortes de espessura fina. Estas sequências são fortemente ponderadas em T2 aparecendo o líquido nas vias biliares com hipersinal e os tecidos adjacentes sem sinal (Marques, 2003).

4.8.2 Sequências EG Rápidas

Com base na utilização de TR curtos e FA parciais desenvolveram-se as sequências EG rápidas muito utilizadas nos dias de hoje. Antes de apresentar uma divisão possível destas sequências e da sua descrição é importante esclarecer alguns conceitos, nomeadamente o conceito de *Steady State Free Precession* e o conceito de *Spoiling*.

Steady State Free Precession

O “estado estacionário de precessão livre” ou estado de equilíbrio estável *Steady State Free Precession (SSFP)* - pode ocorrer sempre que uma série de pulsos de RF, pouco espaçados no tempo, é aplicada a uma amostra. Nestas condições, o sinal FID ocorrerá após aplicação de um pulso de RF e serão produzidos Ecos de *Spin* por cada par sucessivo de pulsos de RF. Cada conjunto de 3 ou mais pulsos dará origem a um outro eco (eco estimulado - STE) que será coincidente com o SE se os pulsos forem muito próximos e não forem aplicados gradientes.

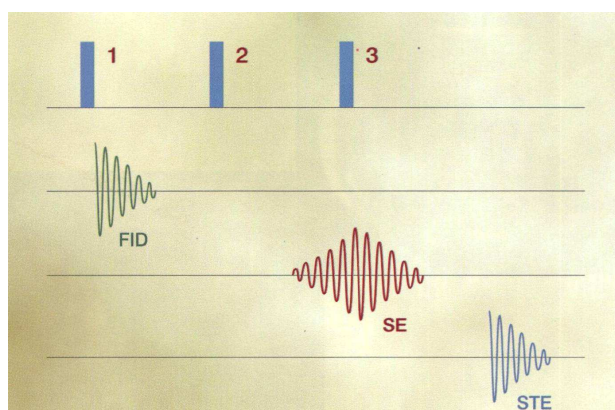


Figura 4. 25 - FID, SE e STE . Fonte: Lafuente (2000).

Se os pulsos de RF forem aplicados muito rapidamente ($TR \ll T_2$), os sinais FID e SE misturam-se de forma que é produzido um sinal contínuo de amplitude variável estabelecendo-se o SSFP.

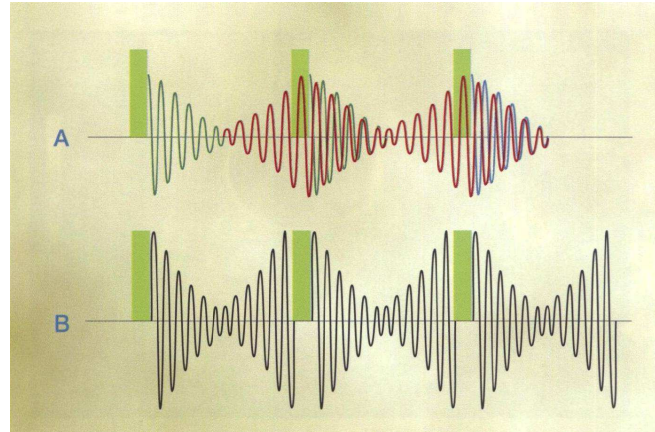


Figura 4. 26 - Sinal composto pela sobreposição do FID, SE's e ecos estimulados. Fonte: Lafuente (2000).

O estado estacionário só ocorre quando: (1) o TR é muito inferior ao T2 (se assim não fosse o decaimento natural destruiria a coerência transversal), (2) os desvios de fase causados pelos gradientes são mantidos constantes de ciclo para ciclo (pulsos com a mesma fase em todos os ciclos) e (3) os *spins* estão estacionários ou for efectuada compensação do fluxo (Elster & Burdette, 2001; Lafuente, 2000).

Spoiling

Destruição da magnetização transversa residual, que permanece de ciclo para ciclo, quando o TR é muito curto, geralmente menor que 50 ms. Desta forma, não existirá componente de magnetização transversa imediatamente antes da aplicação de um novo pulso de RF. O *spoiling* pode ser efectuado por gradientes ou através da aplicação de RF. O *spoiling* de RF, de que é exemplo a sequência SPGR da GE, é mais eficiente do que o *spoiling* de gradientes (sequência FLASH - Fast Low Angle Shot, original da Siemens).

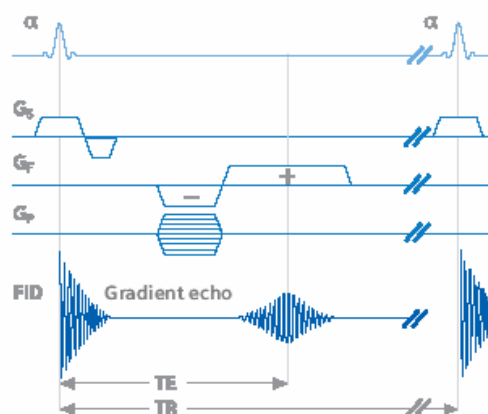


Figura 4. 27 - Sequência EG *Spoiled* - FLASH. Fonte: Siemens (2003)

As sequências EG rápidas dividem-se geralmente em sequências *Steady State* coerente e *Steady State* incoerente. Na verdade, e de acordo com Carr e Hahn, citados por Elster & Burdette (2001), o termo *Steady State* só deveria ser utilizado em sequências em que a coerência transversal dos *spins* é mantida entre os pulsos de RF. Assim, as sequências EG *spoiled* não deveriam ser consideradas SSFP uma vez que a magnetização transversa é destruída antes da aplicação de um novo pulso de RF. No entanto, na bibliografia frequentemente são referidas como sequências SSFP incoerentes, sendo as “verdadeiras” sequências SSFP denominadas como coerentes (Elster & Burdette, 2001). Os termos coerente e incoerente estão relacionados com a presença ou não da componente transversa da magnetização antes da aplicação de um novo pulso de RF.

Assim, as sequências SSFP coerente podem, por sua vez, ser divididas em três grupos: (1) as que medem fundamentalmente a componente FID (SS-GRE-FID) do sinal, (2) as que medem a contribuição do ES e STE (SS-GRE-SE/STE) e as que contam com a componente FID do sinal e do Eco (SS-GRE-FID/SE) (Elster & Burdette, 2001; Lafuente, 2000; Marques, 2003). Como exemplos do primeiro grupo temos a sequência FISP - *Fast Imaging with Steady Precession* da Siemens, GRASS (*Gradient Recalled Acquisition in Steady State*) da GE Medical, FFE (*Fast Field Echo*) da Philips. Como exemplos do segundo grupo temos a sequência PSIF (*Reversed FISP*) da Siemens, CE-FAST (*Contrast Enhancement - FAST*) o nome original desta sequência e que manteve o nome na Picker, SSFP o nome utilizado na GE e T2 - FFE o acrónimo utilizado pela Philips. Do terceiro grupo falaremos mais adiante.

As sequências FISP/GRASS são técnicas EG muito utilizadas na prática. No entanto, as equações relativas à IS destas sequências são extremamente complexas, onde se incluem combinações lineares e exponenciais de T1, T2, T2*, T2/T1 e de α (Lafuente, 2000). O contraste nestas imagens nem sempre é óbvio uma vez que uma só imagem pode apresentar diferentes tipos de contraste. Além disso, o sinal da água e do LCR não podem ser utilizados da mesma forma que nas sequências SE, para determinar a ponderação na imagem. A água poderá apresentar uma IS semelhante numa ponderação DP, T2* ou T2/T1 (Elster & Burdette, 2001).

O segundo grupo de sequências que inclui a sequência PSIF ou CE-FAST apresenta uma forte ponderação em T2. O grau de ponderação em T2 depende exponencialmente do valor do TR. Os líquidos apresentam-se hipertensos, mas há uma grande sensibilidade aos movimentos.

Balanced SSFP

No terceiro grupo das sequências SSFP coerente, as duas componentes do sinal (FID e SE/STE) são utilizadas, obtendo-se um eco combinado simples. Os acrónimos para esta sequência são True FISP (Siemens), balanced FFE (Philips) e FIESTA (GE) (Elster & Burdette, 2001; Lafuente,

2000; Marques, 2003; Nayak et al., 2005; Scheffler & Lehnhardt, 2003). Muitas vezes esta sequência é apenas designada por SSFP o que poderá não estar muito correcto, pois como vimos atrás, outras sequências EG rápidas são também designadas simplesmente por SSFP (Scheffler & Lehnhardt, 2003). Alguns autores, designam estas sequências como balanced SSFP (Nayak et al., 2005; Scheffler & Lehnhardt, 2003) para as distinguir de outras sequências nas quais se verifica o SSFP. À semelhança da técnica EPI, descrita por Mainsfield em 1977, também esta sequência foi descrita em 1958 por Carr e proposta por Oppelt em 1986 (referidos por Scheffler & Lehnhardt, 2003), mas só nos últimos anos se tornou possível a sua aplicação, uma vez que só com a evolução a nível da *performance* dos gradientes foi possível diminuir o TR a poucos ms (3 a 6 ms) permitindo minimizar as bandas de interferência (*dark-phase dispersion bands*) características nesta sequência, resultantes de falta de homogeneidade de campo induzida pelo paciente (Scheffler & Lehnhardt, 2003). O diagrama de pulsos desta sequência é apresentado na Figura 4.28.

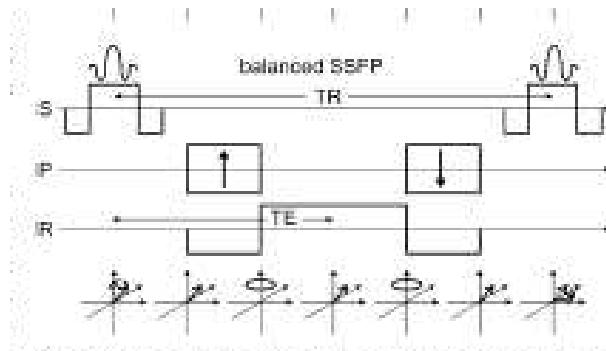


Figura 4. 28 - Diagrama de pulsos de uma sequência b-SSFP (TrueFISP). Nesta sequência, os três gradientes são refocalizados e balanceados conduzindo a um único vector de magnetização, no final do TR. Fonte: Scheffler & Lehnhardt (2003).

O contraste nas sequências balanced SSFP (b-SSFP) resulta da contribuição dos efeitos T1,T2, TR, FA, sendo essencialmente baseado na relação T2/T1, o que nem sempre é o ideal para fins de diagnóstico (Scheffler & Lehnhardt, 2003). Nestas sequências, a gordura e o líquido apresentam IS elevada, ou seja, ambos apresentam hipersinal, com valores de T1 e T2 muito diferentes mas em que a relação T2/T1, é semelhante. Outra das características importantes das sequências b-SSFP é apresentarem a S/R mais elevada, por unidade de tempo, de todas as sequências conhecidas (Marques, 2003; Nayak et al., 2005; Scheffler & Lehnhardt, 2003). Para tecidos em que os valores de T1 e T2 sejam semelhantes (LCR e gordura) e para um FA optimizado o máximo de sinal será 50% de M_0 , o que é praticamente único e conduz a uma S/R tão elevada. No entanto, para uma S/R desta natureza é necessária a utilização de FA elevados (50 a 80°) o que poderá ultrapassar os limites estabelecidos para o SAR, em particular, em campos altos (Scheffler & Lehnhardt, 2003). Para valores em que $T1 \gg TR \gg T2$, o contraste é

semelhante ao de uma sequência EG *spoiled* (FLASH). Para $TR \ll T_2$, o sinal é independente do valor de TR, verificando-se assim, que utilizando TR muito curtos a S/R não se altera e o tempo de aquisição é menor (Liang & Lauterbur, 2000). Esta sequência é muito sensível a pulsos *off-resonance* e requer uma excelente homogeneidade de campo, sendo necessário fazer uma regulação *shim* (3D) antes da sua realização. É também muito susceptível a artefactos de susceptibilidade magnética podendo inviabilizar alguns estudos, no entanto é muito pouco sensível a artefactos de movimentos e fluxo. Está indicada para estudos cardíacos (consegue-se um elevado contraste líquido/miocárdio), estudos abdominais em apneia, em doentes pouco colaborantes e crianças. O TA poderá ser significativamente curto permitindo imagens em tempo real. A sequência b-SSFP poderá ser combinada com imagem espiral com aplicações na RM cardíaca em tempo real (Nayak et al., 2005) e muito útil em estudos de produção de fala permitindo a obtenção de um número elevado de imagens por segundo (Narayanan, Nayak, Lee, & Byrd, 2004). Informação mais detalhada sobre os processos físicos, contraste, S/R e aplicações clínicas pode ser encontrada em (Scheffler & Lehnhardt, 2003).

CISS - Constructive Interference Steady State

Produz imagens fortemente ponderadas em T2 com contraste similar ao obtido com a sequência trueFISP. No fundo são duas sequências trueFISP em que os pulsos de excitação têm fase alternada. Os sinais são combinados após a aquisição de dados (o que aumenta o tempo de reconstrução) mas os erros de fase são cancelados eliminando as bandas de interferência (Elster & Burdette, 2001; Scheffler & Lehnhardt, 2003). O CISS 3D permite obter alto sinal dos fluidos (LCR) e excelente resolução espacial. Contrariamente às sequências FISP há uma compensação inerente do fluxo devido a um balanceamento perfeito dos gradientes. Geralmente utilizada para estudo dos ângulos ponto-cerebelosos, estudo dos nervos cranianos, tumores do acústico, avaliação da cóclea e labirinto (avaliar permeabilidade da cóclea antes colocação de implante coclear) e cerebelo (Elster & Burdette, 2001).

DESS - Dual Echo Steady State

A sequência DESS combina dois ecos no mesmo TR, um FISP e um PSIF, como se pode observar na Figura 4.29. A sequência FISP apresenta ponderação T1/T2* enquanto que a PSIF, como vimos atrás apresenta forte ponderação em T2. Desta maneira a sequência DESS, que combina os dois ecos apresenta um melhor contraste T2 do que a sequência FISP. As imagens têm S/R elevada apresentando os líquidos IS elevada. Por esta razão é indicada para estudos ortopédicos (em geral efectuados com aquisição 3D) onde é necessária boa discriminação entre a cartilagem hialina e o líquido sinovial.

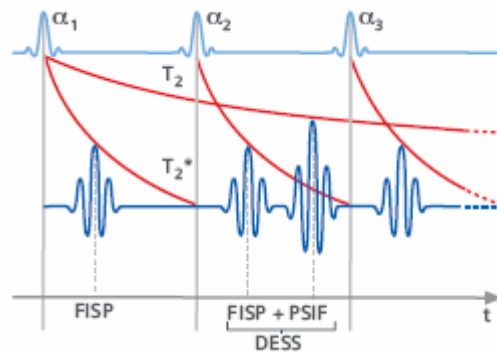


Figura 4. 29 - Diagrama de pulso de uma sequência DESS. Fonte: Siemens (2004).

Magnetization Prepared Gradiente Echo (MP-GRE)

Nas sequências EG, com valores de TR muito pequenos (TR aproximadamente 10 ms), é possível obter tempos de aquisição abaixo do segundo, o que é importante, mas com baixo contraste. Adicionando um pulso de inversão com TI adequado é possível manipular o contraste nas imagens tornando-as clinicamente úteis. São exemplos, a sequência turboFLASH e nas aquisições 3D a sequência MP-RAGE (*Magnetization Prepared - Rapid Gradient Echo*). Na sequência turbo FLASH há uma fase de preparação em que é aplicado um pulso de 180° e que é determinante para o contraste a obter. Com a escolha adequada do TI é possível anular o sinal de alguns tecidos (e.g fígado com hipersinal e lesão com hiposinal) aumentando o contraste entre eles. Durante a fase de aquisição, a matriz é preenchida de uma só vez como numa sequência EG rápida.

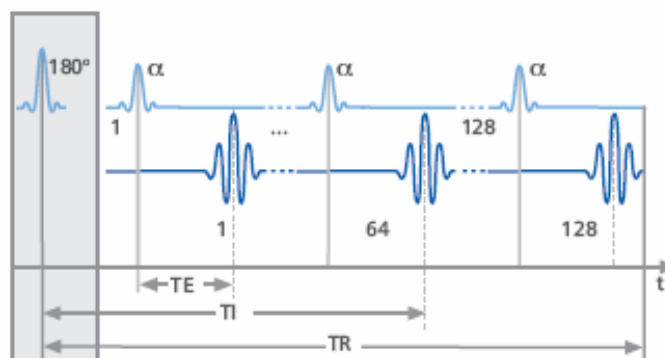


Figura 4. 30 - Diagrama de pulso de uma sequência com preparação da magnetização: Turbo FLASH. Fonte: Siemens (2004).

4.8.3 Técnica Eco Planar ou Imagem Eco planar

A imagem eco-planar que designaremos por EPI (*Echo Planar Imaging*) é uma das estratégias de aquisição de imagem por RM mais antigas. Esta técnica foi proposta por Peter Mainsfield (Universidade de Nottingham) em 1977 (Mainsfield viria a receber o prémio Nobel de Fisiologia e Medicina devido ao desenvolvimento desta técnica em 2003) mas por dificuldades técnicas (relacionadas com os gradientes) só nos anos noventa foi aplicada à prática clínica (Brown & Semelka, 2003; Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001; Huettel, Song, & McCarthy, 2004; Marques, 2003). Esta técnica é considerada das mais rápidas (ultra-rápida) com tempos de aquisição da ordem dos décimos de segundo, tornando fiáveis algumas aplicações clínicas, das quais as mais importantes estão relacionadas com técnicas funcionais para estudo cerebral como a perfusão, difusão e activação cortical (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001).

Com a técnica EPI, o espaço K é preenchido, na sua totalidade, com um único pulso de excitação RF (*single shot*) desde que acompanhado por uma variação extremamente rápida dos gradientes. Para que isto seja possível são necessários gradientes de alta *performance* - actualmente gradientes de 20 a 40 mT/m, com tempos de subida inferiores a 300 μ s e *slew rate* elevado (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001). Os tempos de subida devem ser curtos para que a duração do trem de ecos seja tão curta quanto possível (para minimizar a perda de sinal resultante do decaimento T2*). É necessário que a amplitude máxima atingida seja também elevada, uma vez que a resolução espacial que se consegue é determinada pelo integral da forma de onda do gradiente. A utilização de *slew rate* elevado poderá trazer algumas vantagens, mas aumenta o risco de estimulação (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001).

Na sequência EPI, na direcção de leitura G_x (direcção de codificação da frequência), os gradientes oscilam rapidamente entre valores positivos e negativos, correspondendo a movimento em direcção oposta ao longo do eixo K_x . A trajectória para preencher o espaço K poderá ser em *zig-zag* (utiliza um gradiente de pequena intensidade mas constante na direcção de codificação da fase) ou em *blip* (gradiente forte, aplicado durante um muito curto espaço de tempo (< 200 μ s) quando o gradiente de leitura está no zero - na extremidade do eixo K_x). Cada *blip* muda a trajectória do espaço K de uma linha para outra ao longo da direcção G_y como se fosse um degrau. O modo *blipped EPI* é o que é mais frequentemente utilizado, nos dias de hoje, uma vez que é mais fácil a aplicação da TF com esta trajectória (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001). A resolução espacial com a técnica *single shot* ronda os 2 a 3 mm (dimensão do pixel) com FOV de 250 mm e matrizes de aquisição de 128 x 128, podendo melhorar recorrendo à utilização de imagem paralela ou técnica *multi-shot* (aquisição de

informação à custa de vários *shots*, cada um dos quais corresponde a um pulso de excitação (Carroll, Sakaie, Wielopolski, & Edelman, 2006)). De salientar, que a técnica EPI é apenas um esquema de leitura que permite preencher rapidamente o espaço K, sendo possível obter imagens SE-EPI, EG-EPI ou IR-EPI (figura 4.31) e com diferentes ponderações em função dos pulsos de excitação e da sequência utilizada (Elster & Burdette, 2001; Hashemi & Bradley, 1997). Assim, para obter uma SE-EPI, é aplicado um esquema (90°- 180°- EPI). Com esta técnica pode obter-se, numa sequência SE-EPI, contraste T2 muito semelhante ao obtido com uma sequência SE convencional, manipulando o TE. Como o TR efectivo na técnica *single-shot* é infinitamente grande, não existe uma ponderação T1 inerente (Hashemi & Bradley, 1997; Liang & Lauterbur, 2000; Marques, 2003).

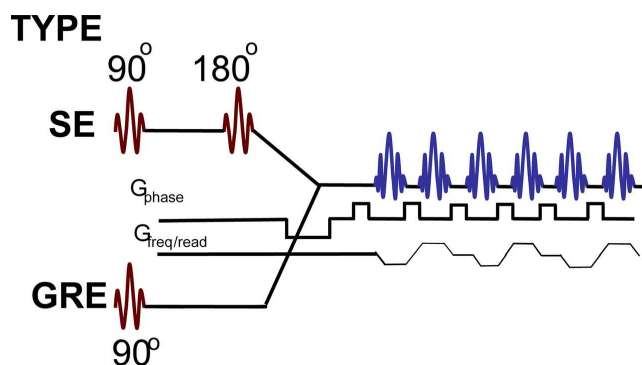


Figura 4. 31 - Esquema de uma sequência EPI: em cima SE-EPI em baixo EG-EPI. Fonte: Bitar et al. (2006).

Como os tempos de aquisição são extremamente curtos, a degradação de imagem resultante de movimentos fisiológicos é praticamente nula (Liang & Lauterbur, 2000).

No entanto, associam-se a esta técnica alguns artefactos, nomeadamente: (1) artefacto N/2 *ghost*, imagens fantasma da imagem principal, ao longo da direcção de codificação da fase, causadas por correntes de *eddy*, imperfeições nos gradientes, deficiente homogeneidade do campo ou ainda por coincidência entre o tempo dos ecos pares e ímpares; (2) artefactos de susceptibilidade magnética e de (3) desvio químico. Os artefactos de desvio químico com a técnica EPI são bastante mais pronunciados do que com a técnica SE convencional e manifestam-se ao longo da direcção de codificação da fase. Para minimizar este artefacto é necessário utilizar uma eficiente técnica de supressão da gordura (Carroll, Sakaie, Wielopolski, & Edelman, 2006).

4.8.4 Outras estratégias para aquisição rápida

4.8.4.1 Time-resolved imaging

Técnicas cartesianas

Uma das estratégias para adquirir de forma mais rápida é aumentar a rapidez da aquisição utilizando EPI, Single-Shot TSE, etc. Outra alternativa viável será reduzir a quantidade de informação adquirida. Vulgarmente designa-se esta técnica por undersampling ou sub-amostragem. As técnicas de imagem paralela também utilizam o princípio da sub-amostragem mas tendo em atenção princípios diferentes como veremos no ponto seguinte (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001).

Partial Fourier Imaging

A técnica *Partial Fourier* é um método de reconstrução robusto e usado com alguma frequência em estudos de difusão ou angio RM. É ainda utilizada em conjunto com outras técnicas rápidas como o HASTE e em conjunto com sequências b-SSFP (Carroll, Sakaie, Wielopolski, & Edelman, 2006). Numa aquisição normal, o gradiente de codificação da fase vai aumentando a partir de valores negativos (de grande amplitude) até valores positivos de grande amplitude ou vice-versa. Utilizando a técnica *Partial Fourier*, são utilizados menos passos de codificação da fase (N_p) mas a partir desta informação é possível reconstruir a restante, para formar a imagem. Frequentemente são obtidos pouco mais que metade dos passos de codificação da fase pelo que se designa esta técnica por Half-Fourier, $\frac{1}{2}$ NEX ou ainda Half-Scan (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001; Hashemi & Bradley, 1997). O termo $\frac{1}{2}$ NEX poderá confundir, uma vez que, o que se reduziu a metade foi N_p , e não o número de excitações por passo (Elster & Burdette, 2001).

Uma imagem pode ser reconstruída a partir de parte dos dados, apenas porque o espaço K apresenta uma simetria Hermetiana ou simetria conjugada (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001; Hashemi & Bradley, 1997). A simetria Hermetiana resume-se ao seguinte: a IS de um ponto na porção crescente de um eco utilizando um passo de codificação da fase positivo corresponde a um ponto na porção descendente de outro eco utilizando o passo de codificação negativo correspondente (imagem em espelho). As aproximações por simetria conjugada podem não ser perfeitas pois existem sempre erros de fase pelo que são sempre adquiridas mais do que metade das linhas (60%) do espaço K. Apesar de reduzir o TA mantendo a resolução espacial, o que é importante, o preço a pagar é uma

diminuição da S/R por um factor equivalente a $\sqrt{2}$ (Elster & Burdette, 2001; Hashemi & Bradley, 1997).

FOV Rectangular e matriz rectangular

O tempo de aquisição pode ainda ser reduzido se forem utilizadas técnicas com FOV rectangular (RecFOV) ou matriz rectangular. O recFOV pode ser utilizado sempre que a dimensão anatómica numa das direcções seja muito maior que na outra. Nesta situação, reduz-se o número de linhas do espaço K a adquirir, sendo o restante espaço K preenchido com zeros. Nesta situação, a resolução espacial mantém-se, a S/R diminui ligeiramente, o equivalente a $1/\sqrt{2}$ e diminui o tempo de aquisição. Com a diminuição do FOV é necessário ter em atenção os artefactos de *aliasing* na direcção da fase (Lafuente, 2000).

Também pode ser utilizada uma matriz rectangular, assimétrica em que se diminuem N_y , adquirindo-se apenas a parte central do espaço K (diminuir 50%). O TA diminui, a S/R aumenta pois o voxel aumenta, mas há degradação da resolução espacial na direcção da fase.

Zoom Imaging

A técnica Zoom foi desenvolvida pela Philips, encontrando-se disponível nos seus equipamentos. Combina uma aquisição e reconstrução rápidas com visualização automática, para obtenção de imagens em tempo real, com FOV rectangular. Frequentemente é utilizada com sequências TSE com um espaçamento muito curto entre os ecos ou com técnica GRASE. É possível utilizar um FOV rectangular extremamente reduzido (25%), podendo ser conjugado com técnica de aquisição parcial (Figura 4.32). Desta forma, consegue-se um *single shot* com um número muito pequeno de passos de codificação de fase (20 a 30). Obtém-se uma aquisição ultra-rápida, sacrificando a cobertura anatómica em prol de uma maior rapidez, mantendo-se a resolução espacial. Para que não exista enrolamento com um FOV tão pequeno é necessário inverter o gradiente na direcção de codificação da fase (Elster & Burdette, 2001). Alguns estudos, (Demolin, Metens & Soquet, 2000; Engwall, 2003) realizados na área da fala recorrem a esta técnica para aquisição de imagens em tempo real.

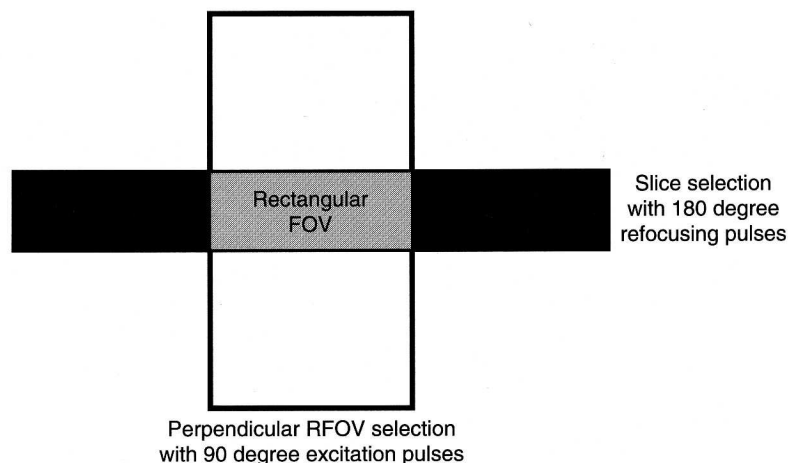


Figura 4. 32 - Técnica TSE Zoom (Philips). Fonte: Elster & Burdette (2001).

4.8.4.2 Técnicas não cartesianas

Spiral imaging

A imagem espiral é também uma estratégia desenvolvida para preencher o espaço K, em técnicas de aquisição de imagem ultra-rápida. De forma semelhante à EPI, este método de preenchimento do espaço K, pode ser conjugado com diferentes tipos de sequência (EG, SE ou IR). Este método recorre a dois gradientes (G_x e G_y) oscilantes accionados durante o período de leitura da sequência (Elster & Burdette, 2001; Liang & Lauterbur, 2000). Os gradientes aplicam uma aceleração constante para que o espaço K seja preenchido numa trajectória espiral. Com esta técnica não é necessário uma rápida variação dos gradientes como acontece na EPI (Liang & Lauterbur, 2000). As espirais iniciam-se no centro do espaço K e terminam na periferia fazendo com que as baixas frequências (responsáveis pelo contraste na imagem) sejam lidas no início do esquema de leitura, quando a perda de sinal devida a efeitos $T2^*$ ainda é pequena. Também os erros de fase devidos a movimento são muito menores do que em outras técnicas de aquisição rápidas (Carroll, Sakaie, Wielopolski, & Edelman, 2006). Pode existir uma ou mais espirais (*interleaves*). A informação obtida a partir das trajectórias espirais necessita de algoritmos especiais para reconstrução da informação, sendo necessários vários passos antes da aplicação da TF (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001; Liang & Lauterbur, 2000). O espaço K não é preenchido de uma forma cartesiana e como tal o tempo e a localização no espaço K não se relacionam de forma linear. Desta maneira, os artefactos devidos a efeitos *off-resonance* não resultam em desvios de fase mas num esborratamento de forma radial na imagem (Carroll, Sakaie, Wielopolski, & Edelman, 2006; Elster & Burdette, 2001; Liang & Lauterbur, 2000).

4.8.4.3 Imagem Paralela

As técnicas de aquisição de imagem paralela, que denominaremos técnicas PAT (acrónimo do termo em inglês *Parallel Acquisition Techniques*) ou iPAT surgiram nos anos noventa e resultam de um novo mecanismo de codificação espacial da informação. Para a sua implementação foi fundamental o desenvolvimento tecnológico ao nível das bobines de RF utilizadas para a captação de sinal. Na realidade, as técnicas PAT utilizam conjuntos de antenas ou bobines de superfície, denominadas antenas *phased array*, para adquirir, em simultâneo, informação relativa a múltiplos pontos, permitindo diminuir significativamente o TA, sem que sejam ultrapassados alguns limites fisiológicos e técnicos associados à rápida variação dos gradientes de campo magnético (Glokner, Hu, Standley, Angelos, & King, 2005; Heidemann et al., 2003; Sodickson, 2006).

Será importante clarificar que as técnicas PAT não são novas sequências, permitem acelerar qualquer sequência de pulsos (EG, SE, EPI) geralmente utilizadas na prática clínica e não alteram o contraste das imagens subjacente à sequência utilizada. Pode-se utilizar PAT meramente para acelerar a aquisição por um factor R, melhorar a resolução espacial para o mesmo TA ou, o mais frequente, uma combinação dos dois (Glokner, Hu, Standley, Angelos, & King, 2005; Heidemann et al., 2003).

De forma muito resumida pode-se dizer que para obter imagem paralela:

- São necessárias múltiplas antenas ao longo do FOV com diferentes perfis de sensibilidade;
- Cada antena tem de ter o seu próprio receptor de forma a receber informação em simultâneo mas independente, de forma a obter informação individual;
- É necessário conhecer o perfil de sensibilidade (obtido por calibração) de cada antena do conjunto (*array*) para obter a reconstrução final, o que demora algum tempo;
- A aquisição é acelerada por um factor R (2 ou mais), diminuindo o TA, mas com diminuição da S/R por um factor equivalente a \sqrt{R} ;
- A direcção de codificação em fase está condicionada quando se utilizam *phased array* convencional (Heidemann et al., 2003).

O princípio destas técnicas tem por base o conhecimento de que o TA, numa aquisição Cartesiana, é proporcional ao número de passos de codificação da fase. Reduzindo o número de passos de codificação da fase, mantendo a resolução espacial, reduz o tempo de aquisição por um factor semelhante. Esta sub-amostragem periódica do espaço K conduz a artefactos de *aliasing* ou *wraparound* na direcção de codificação da fase (ver Figura 4.33). No entanto, os métodos de reconstrução, baseados na sensibilidade das antenas permitem depois a eliminação do artefacto.

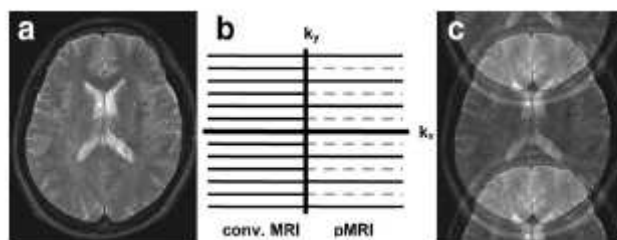


Figura 4. 33 - Efeito da amostragem no FOV. (a) Imagem axial convencional e (b) esquema de amostragem do espaço K para uma aquisição convencional e com PAT, e (c) artefacto derivado da sub-amostragem. Fonte: Heidemann et al.(2003).

Os métodos de imagem paralela existentes, hoje em dia, podem-se dividir ou classificar em 3 grandes grupos: (1) SENSE (Sensitivity Encoding) baseada na composição da imagem, (2) técnicas híbridas (SPACE RIP - Sensitivity Profiles from an Array Coil for Encoding and Reconstruction in Paralell) e (3) métodos de regeneração do espaço K (SMASH - Simultaneous Acquisition of Spacial Harmonics, AutoSMASH e GRAPPA - Generalized Autocalibrating Partially Parallel Acquisition).

A técnica SMASH surgiu em 1997, foi introduzida por Sodickson & Manning (Sodickson, 2006) e utiliza os perfis de sensibilidade da antena como base para obter harmónicos espaciais e posteriormente reconstruir a imagem no domínio de Fourier. A Técnica GRAPPA é uma variante da SMASH em que são adquiridas linhas adicionais do espaço K, não sendo necessária uma aquisição em separado para calibração da sensibilidade das antenas. Em 1999, foi descrita a técnica SENSE (Pruessmann, Weiger, Scheidegger, & Boesiger, 1999) em que é aplicada a TF em primeiro, conduzindo a um artefacto de *aliasing*, sendo depois reconstruída a informação com base nos perfis de sensibilidade das antenas obtidos por calibração. As técnicas híbridas foram introduzidas em 2000 (Glokner, Hu, Standley, Angelos, & King, 2005; Heidemann et al., 2003). Uma explicação e revisão mais aprofundada destas técnicas pode ser encontrada em (Glokner, Hu, Standley, Angelos, & King, 2005; Heidemann et al., 2003; Hunold, Maderwald, Ladd, Jellus, & Barkhausen, 2004; Pruessmann, Weiger, Scheidegger, & Boesiger, 1999; Sodickson, 2006).

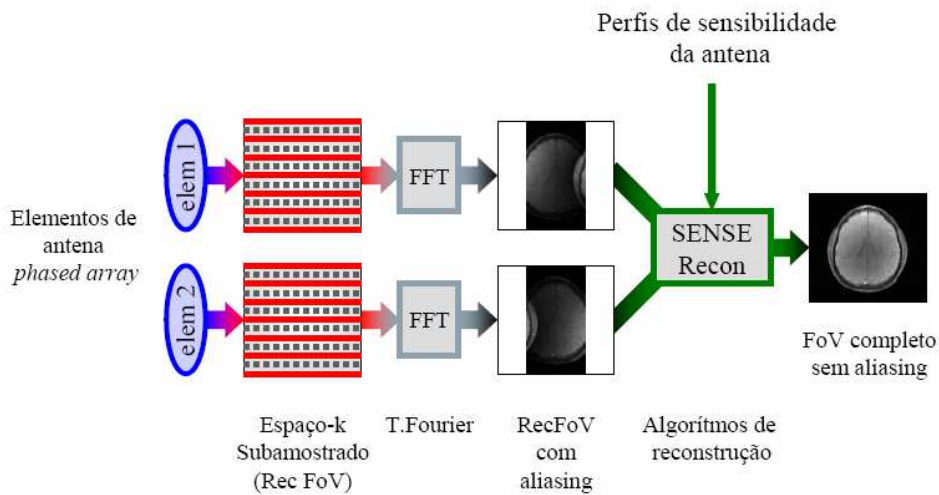


Figura 4. 34 - Esquema simplificado de aquisição paralela utilizando a técnica SENSE. Fonte: Pires (2005).

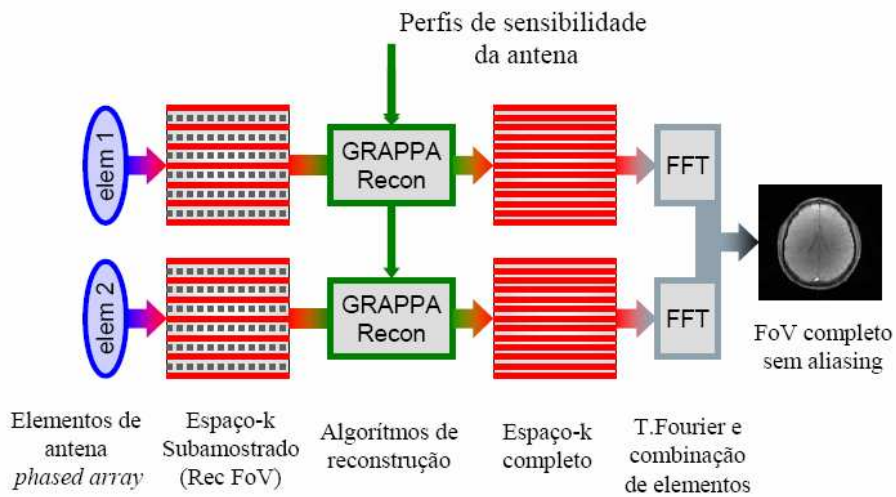


Figura 4. 35 - Esquema simplificado de aquisição paralela utilizando a técnica GRAPPA. Fonte: Pires (2005).

Entre as muitas aplicações das técnicas PAT (estudos cardíacos e vasculares, estudos dinâmicos, estudos em apneia entre outros) gostaríamos de referir a sua aplicação na obtenção de imagens em tempo real dada a elevada resolução temporal que permite. É especialmente indicada para aquisição de imagem quando o movimento a estudar não é periódico não havendo possibilidade de efectuar *trigger*. Movimentos como a deglutição, movimentos dos articuladores durante a fala, entre outros, só poderão ser observados com imagem em tempo real e com um número elevado de imagens por segundo (Heidemann et al., 2003). No presente, com o desenvolvimento das antenas matriciais é possível obter *arrays* em qualquer direcção não condicionando a escolha da direcção de codificação da fase e permitindo factores de aceleração maiores com melhoria da S/R

(Pires, 2005). Os artefactos de reconstrução e a S/R ainda crítica, para factores de aceleração muito elevados, são as principais limitações (Glokner, Hu, Standley, Angelos, & King, 2005).

4.9 Sumário

Neste capítulo, efectuou-se uma síntese das principais técnicas de aquisição de imagem por RM, com o intuito de, nos capítulos seguintes serem compreensíveis algumas descrições e opções que efectuaremos. Dada a complexidade do tema e os permanentes desenvolvimentos na área ficará sempre muito aquém do que será o verdadeiro estado da arte. Também, não foram incluídos tópicos pertinentes como a segurança, dado o já avultado capítulo. No entanto, gostaríamos de referir que a RM sendo considerado um método de imagem seguro, não é isenta de riscos, particularmente, se não forem respeitadas normas e limites que estão relacionados quer com o campo magnético principal (efeito míssil, compatibilidade de implantes, etc), quer com a RF (por exemplo, o Specific Absorption Rate) e com as bobines de gradientes (por exemplo, estimulação periférica e índices de ruído).

Capítulo 5

Ressonância Magnética em Estudos de Produção de Fala

A obtenção de informação relativa à forma e dimensões do tracto vocal é fundamental para uma completa compreensão dos processos acústicos e articulatórios envolvidos na produção de fala (Baer, Gore, Gracco, & Nye, 1991). No capítulo três foram já descritas, algumas das técnicas frequentemente utilizadas para obtenção dessa informação e onde se incluíram algumas técnicas de imagem. Na verdade, informação proveniente de estudos cineradiográficas serviu de suporte à teria acústica de produção de fala (Chiba & Kajiyama, 1941; Fant, 1960). No entanto, e como já se referiu, os riscos decorrentes da exposição à radiação de indivíduos saudáveis, condicionam a sua utilização, nos dias de hoje, neste tipo de estudos (Arnal et al., 2000; Munhall, Vatikiotis-Bateson, & Tohkura, 1995; Narayanan, Nayak, Lee, & Byrd, 2004).

A Ressonância Magnética (RM) sendo considerado um método de imagem relativamente inócuo surge assim como uma das técnicas de imagem que mais poderá contribuir para um melhor conhecimento dos mecanismos envolvidos na produção de fala. Assim, neste capítulo e uma vez que já foram expostos os princípios subjacentes a esta técnica de imagem (ver capítulo quatro) far-se-á apenas um resumo das suas potencialidades e limitações no contexto da produção de fala (secção 5.1). Nas secções seguintes, resumem-se alguns dos estudos que têm utilizado a RM para esse efeito, contemplando diferentes línguas. Dividem-se os estudos em estudos estáticos (secção 5.2) e estudos dinâmicos e em tempo real (secção 5.3) e dentro de cada secção os estudos são apresentados cronologicamente do mais antigo para o mais recente.

5.1 Potencialidades e Limitações da RM no contexto da produção de fala

A RM é um dos métodos de imagem que poderá representar uma alternativa viável aos métodos de imagem com recurso a radiações ionizantes, tendo sido amplamente utilizada nesta área, nos últimos anos, como já foi referido. A RM tem demonstrado ser uma ferramenta poderosa e atractiva nos estudos de fala uma vez que permite obter informação relativa à geometria do tracto vocal sem os riscos associados às radiações ionizantes. A ausência de radiações é importante uma vez que na maior parte destes estudos há necessidade de adquirir um grande número de imagens no mesmo sujeito (Baer, Gore, Gracco, & Nye, 1991; Engwall, 2003b; Engwall & Badin, 1999; Narayanan, Alwan, & Haker, 1995b, 1997; Narayanan, Nayak, Lee, & Byrd, 2004; Story, Titze, & Hoffman, 1996). Outras vantagens da RM incluem a sua capacidade multiplanar, excelente resolução de contraste e boa relação sinal-ruído permitindo excelente discriminação

das estruturas anatómicas, particularmente dos tecidos moles (Baer, Gore, Gracco, & Nye, 1991; Crary, Gorham, Burton, Kotzur, & Gauer, 1996; Moore, 1992; Shadle, Mohammad, Carter, & Jackson, 1999). Permite modelação 3D e cálculo de áreas e volumes do tracto vocal e nasal de forma directa (Narayanan, Alwan, & Haker, 1995b, 1997; Narayanan, Nayak, Lee, & Byrd, 2004). Outra vantagem apresentada pela RM é permitir o estudo das estruturas internas sem necessidade de colocação de eléctrodos ou outro tipo de materiais (palatos artificiais) como acontece com a EMA e EPG e que podem de alguma forma interferir com a articulação (Mohammad, 1999).

No entanto, o tempo de aquisição longo, com a maior parte das sequências utilizadas nos primeiros estudos com RM limitou a sua utilização ao estudo de configurações estáticas do tracto vocal, durante a produção de sons passíveis de serem sustidos por um período de tempo relativamente longo, minutos nos primeiros estudos (Alwan, Narayanan, & Haker, 1997; Baer, Gore, Gracco, & Nye, 1991; Narayanan, Alwan, & Haker, 1995b; Story, Titze, & Hoffman, 1996) a alguns segundos em estudos mais recentes (Demolin, Metens, & Soquet, 1996; Gick, Kang, & Whalen, 2000, 2002; Kröger, Winkler, Mooshammer, & Pompino-Marschall, 2000; Yang, 1999). De salientar, que as articulações produzidas de forma sustida tendem a ser hiperarticuladas e mais difíceis de manter se o tempo de aquisição for demasiado longo (Engwall, 2003b). Ainda que com estas limitações em termos de resolução temporal, foi possível investigar a configuração do tracto vocal durante a produção de vogais (orais e nasais) e algumas consoantes (nasais, fricativas, laterais) e tem sido permitido gerar funções de área do tracto vocal, o que representa um passo importante no estudo da relação entre a geometria do tracto vocal e a acústica na produção de fala (Demolin et al., 1997; Demolin, Metens, & Soquet, 2000).

Recentemente, algumas inovações técnicas ao nível de *soft* e *hardware* têm permitido alargar a aplicação da RM, nos estudos de produção de fala. O desenvolvimento de técnicas de aquisição ultra-rápidas, com resolução temporal da ordem dos 110 ms tem permitido monitorizar em “tempo real” alguns fenómenos dinâmicos envolvidos na produção de fala, como alterações no posicionamento da língua, lábios, e posicionamento do velo ao longo do tempo, mas ainda não suficiente para observar algumas características dinâmicas do movimento da língua, que exige amostragens mínimas da ordem dos 20 Hz (Narayanan, Nayak, Lee, & Byrd, 2004). A utilização da RM com o intuito de evidenciar aspectos dinâmicos da articulação durante a produção das consoantes oclusivas, por exemplo, é também bastante mais exigente devido às características temporais das oclusivas (Mathiak et al., 2000).

Outra das limitações da RM frequentemente referida neste tipo de estudos está relacionada com os elevados índices de ruído durante a aquisição. Se com as sequências utilizadas classicamente nos estudos estáticos ainda é possível a obtenção de um sinal acústico (de má qualidade, é certo) obtido por gravação através do intercomunicador (Mohammad, 1999; Shadle, Mohammad, Carter, & Jackson, 1999) com as sequências em tempo real, dado o ainda maior índice de ruído, isso é completamente inviável, com equipamento comum. Recentemente, essa limitação está a ser

ultrapassada, mediante o recurso a sistemas de microfones ópticos, com sistema de cancelamento de ruído, que por um lado não degradam a qualidade de imagem nem representam um risco para falante, e por outro, permitem a obtenção de um sinal acústico referente à produção efectuada, com uma qualidade aceitável (Bresch, Nielsen, Nayak, & Narayanan, 2006; NessAiver et al., 2006). A obtenção do sinal acústico pode servir vários objectivos: (1) comparar as produções naturais do falante com as obtidas através de processo de síntese, (2) fazer a sincronização das imagens com o sinal acústico e (3) verificar se o falante produziu exactamente o que lhe foi solicitado.

Outro factor também apontado como uma desvantagem da RM reside no facto de, na maior parte dos equipamentos a aquisição ser feita com o falante deitado o que pode causar alguma interferência na produção. Alguns estudos (Engwall, 2003b; Tiede, Masaki, & Vatikiotis-Bateson, 2000; Whalen, Kang, Magen, Fulbright, & Gore, 1999) revelam que a posição deitada pode influenciar a posição e configuração da língua, podendo diminuir a passagem a nível faríngeo, sendo este efeito, mais observável nas articulações sustentadas artificialmente do que nas articulações obtidas em tempo real. Contudo, alguns autores referem que essas alterações gravitacionais são negligíveis podendo ser efectuadas compensações (Magen, Kang, Tiede, & Whalen, 2003).

O facto de estruturas calcificadas, como os dentes, serem mal observadas, em qualquer ponderação nas imagens RM, acarretando dificuldades acrescidas na segmentação a nível da cavidade oral, tem conduzido ao desenvolvimento de estratégias que permitem, através de co-registo introduzir essa informação, quer a partir de outras aquisições RM (Engwall & Badin, 1999; Takemoto, Kitamura, Nishimoto, & Honda, 2004) para as arcadas dentárias, quer a partir de informação adquirida por TC (Serrurier & Badin, 2005ab) ou EBCT (Story, Titze, & Hoffman, 1996).

5.2 Estudos estáticos

1991 - Baer, T., J. C. Gore, et al. "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels."

Não tendo sido o primeiro estudo a utilizar RM nos estudos de produção de fala, foi o primeiro a incluir informação 3D e é considerado um estudo pioneiro nesta área. Foi efectuado por Baer, Gore, Gracco, & Nye (1991), para a Língua Inglesa. Já anteriormente, estes autores tinham sugerido a aplicação da RM neste contexto (Baer et al, 1987).

Período do estudo:

- Estudo decorreu durante um período de 4 anos;

Corpus:

- Foram obtidas configurações do tracto vocal, em toda a sua extensão (da laringe aos lábios), de 2 informantes do sexo masculino, durante a produção de vogais;
- Falantes nativos da língua Inglesa, um com dialecto de Nassau County (NY) e outro com dialecto de Somerset (SW Inglaterra);

Aquisição de Imagem:

Foram efectuadas 2 experiências utilizando 2 equipamentos diferentes.

Estudo 1 (EMR- Experimental Magnetic Resonance)

- Equipamento GE (magneto resistivo de 0.15 Tesla);
- Utilização de estabilizador para a cabeça e gravação de som;
- Utilizada uma antena de cabeça e cefalostato para reposicionar a cabeça entre as várias sessões;
- A aquisição das imagens para cada configuração do tracto vocal demorou cerca de 3 a 4 horas;
- Informantes ouviam uma produção canónica da vogal que produziam;
- Foi efectuada gravação através de microfone colocado, dentro do magneto a poucos cm dos lábios do falante. O microfone conectava-se através de um cabo com 4 m a um amplificador, que por sua vez estava ligado a um gravador fora da sala do magneto;
- Foram efectuadas imagens axiais, sequência SE ponderada em T1, com TR=200ms e TE=11 ms, espessura de corte de 8 mm com intervalo de 5 mm entre cortes, matriz 128x128, obtendo-se boa S/R e bom contraste tecidual. O tempo de aquisição para cada imagem foi de 3.4 minutos;
- Efectuados cortes coronais, com a mesma espessura, marcados desde parede posterior da faringe até aos lábios; total de 19 cortes axiais e 18 coronais;
- Os falantes produziam cada vogal de forma sustida durante 15s, entre inspirações rápidas, durante o tempo total de aquisição.

Estudo 2

- Equipamento SIGNA da GE (supercondutor com 1,5 T), utilizado para fins de diagnóstico, pelo que o tempo disponível para aquisição de imagens e a liberdade de instalar equipamento foram limitados;
- Não foi utilizado cefalostato, nem foi obtida a gravação das produções durante a aquisição, contudo permitiu encurtar o tempo de aquisição, obtendo-se cada configuração do tracto vocal em cerca de 30 minutos. Foram obtidos cortes axiais referentes às 4 vogais cardinais (*point vowels*) e das 5 vogais intermédias da língua inglesa;
- Foram realizados 17 cortes axiais, SE T1 com TR=800ms e TE=20 ms, 256x256, FOV=200mm e 1 NEX;

Discussão:

- Centra-se na comparação dos resultados obtidos para o comprimento do tracto vocal obtido a partir das imagens RM e de raios X, no papel dos seios piriformes na qualidade das vogais sintetizadas e nas grandes diferenças encontradas entre as formantes das produções originais (análise LPC - *Linear Predictive Coding*) e os valores das formantes calculadas a partir das formas de onda sintetizadas a partir dos valores obtidas para as funções de área;
- Relações entre as áreas e as distâncias médio-sagitais;

Limitações:

- Tempo de aquisição extremamente longo, sendo difícil manter a qualidade da produção durante períodos tão longos;
- Fraca qualidade de imagem devido a movimentos;
- Qualidade das gravações longe do ideal, não sendo possível, por exemplo, determinar a presença ou não de nasalização nas produções originais.

1992 - Greenwood, A. R., C. C. Goodyear, et al. "Measurements of vocal tract shapes using magnetic resonance imaging."

Greenwood et al. efectuaram em 1992, um estudo também para a língua Inglesa, em que estudaram, 5 vogais da língua inglesa, num informante. Foi efectuada sequência FLASH, uma imagem no plano médio sagital e vários cortes no plano axial e coronal com um tempo de aquisição de 10 a 12s. Obtiveram as funções de área a partir dos cortes efectuados, tendo como referência o modelo proposto por Mermelstein. Foi efectuado estudo de percepção.

1995 - Narayanan, S., A. Alwan, et al. "An articulatory study of fricative consonants using Magnetic Resonance Imaging".

Estudo efectuado em quatro falantes do Inglês Americano (dois do sexo masculino e dois do sexo feminino) para obtenção de informação 3D relativa à configuração do tracto vocal para oito consoantes fricativas do Inglês Americano. Foi utilizado campo de 1,5 Tesla, com recurso a sequência Spoiled GRASS, com espessura de corte de 3 mm. Foram obtidos cortes coronais (40 a 45 cortes), axiais (40 a 45 cortes) e sagitais (28 a 35 cortes), os primeiros foram marcados desde os lábios até à parede posterior da faringe, os axiais desde a porção acima do palato duro até ao nível de C7-T1. Os coronais e axiais foram utilizados como referência para marcação dos sagitais. Durante a aquisição, os falantes sustentaram cada som por períodos de 13 a 16s permitindo a obtenção de quatro ou cinco cortes. Cada som teve que ser repetido cerca de seis a nove vezes para obter toda a configuração do tracto para cada som, em cada um dos três planos. A aquisição de imagens em cada um dos planos (ex. axial) para o mesmo sujeito foi efectuada na mesma sessão, sendo o conjunto dos 3 planos adquirido em 3 sessões.

Foram obtidas medidas do comprimento do tracto vocal a partir dos cortes sagitais (comprimento total, comprimento da região anterior e posterior do tracto, abertura labial vertical, local de maior constrição etc.), funções de área para cada uma das fricativas (também, a partir dos cortes sagitais), foi também efectuada uma análise morfológica detalhada do tracto vocal e, em particular da língua durante a produção das fricativas (a partir de reconstruções 3D). Foram obtidas também medidas relativas às cavidades sublinguais e seios piriformes.

1996 - Story, B. H., I. R. Titze, et al. "Vocal tract area functions from magnetic resonance imaging."

Com o objectivo de obter informação tridimensional relativa à configuração do tracto vocal, e obtenção das funções de área, estes autores obtiveram para um falante da zona oeste dos EUA (Inglês Americano), a configuração do tracto vocal para dezoito fones (doze vogais, três consoantes nasais e três oclusivas). Foram utilizadas palavras de referência para identificação dos fones. As imagens foram adquiridas num campo de 1,5 T e utilizada sequência FSE. Foram obtidos 26 cortes axiais paralelos e contíguos com modo de aquisição "interleaved", marcados desde a parte superior do palato duro até à parte inferior do 1º anel traqueal. O tempo de aquisição foi de 4 minutos e 16s. A aquisição tinha início quando, através do intercomunicador, se ouvia o início da fonação e parava a aquisição ao fim de 8 s para permitir que o falante respirasse, desta forma o tempo total de aquisição para cada configuração foi de aproximadamente 10 minutos. Durante a aquisição não foi adquirido o sinal de voz, mas foi efectuada gravação de alta qualidade posteriormente, com o mesmo falante a produzir os mesmos sons e com configuração estática igual à adquirida com a RM. A finalidade destas gravações foi permitir a comparação da fala natural do indivíduo com as simulações conseguidas a partir das áreas calculadas.

Foram também obtidas, utilizando EBCT, as arcadas dentárias do falante com o intuito de calcular o volume ocupado pelos dentes e permitir uma melhor segmentação do tracto vocal. Os autores compararam os seus resultados (funções de área de algumas vogais) com outros estudos já publicados. Segundo estes autores, as diferenças encontradas podem ser devidas a diferenças a nível da técnica de aquisição de imagem utilizada ou a diferentes protocolos, e mais importante a diferenças anatómicas e fisiológicas entre os sujeitos.

1996 - Demolin, D., T. Metens, & Soquet "Three-dimensional measurement of the vocal tract by MRI."

Este estudo foi efectuada com o intuito de obter medidas do tracto vocal a partir de cortes obtidos nos planos sagital, coronal, coronal oblíquo e axial e a partir destes obter áreas precisas em diferentes pontos do tracto vocal para falantes da língua Francesa. Participaram no estudo quatro

falantes (dois do sexo masculino e dois do sexo feminino) residentes em Bruxelas. O corpus incidiu sobre as dez vogais da língua Francesa, sendo efectuado com os falantes a suster a articulação durante os 14s do tempo de aquisição. Foi utilizado campo de 1,5 Tesla e sequência TSE com ponderação DP e espessura de corte de 4 mm. Antes da sessão foram efectuados 11 cortes no plano sagital, sendo utilizado o corte ao nível do plano médio sagital para marcar os cortes coronais (5) coronais oblíquos (3) e axiais (6), num total de 14 cortes, conforme se pode observar na Figura 5.1.

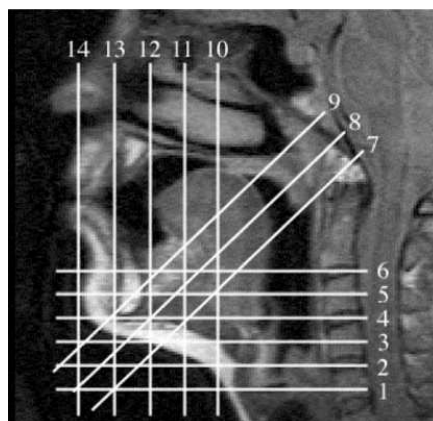


Figura 5. 1 - Marcação das ranges no plano axial, coronal e oblíquo. Fonte: Demolin, Metens & Soquet (1996).

Os autores referem dificuldades na determinação das interfaces ar-tecido e na dificuldade em distinguir ar/dentes na região alveolar e do palato duro, devido à falta dos moldes dentários dos falantes.

1997 - Alwan, A. A., S. S. Narayanan, et al. "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics."

1997 - Narayanan, S., A. Alwan, et al. "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals."

Estes dois estudos tiveram como objectivo caracterizar a geometria tridimensional das variantes *dark* e *light* da lateral /l/ (Parte I) e das vibrantes (rhotics) - (parte II) do inglês americano, de forma a permitir o desenvolvimento de modelos acústicos mais perfeitos. A informação utilizada foi obtida por RM e EPG. Não foram estudados efeitos de coarticulação nos modelos 3D nem características dinâmicas, dadas as limitações da RM e a pouca informação obtida por EPG.

O corpus foi obtido com quatro falantes (dois do sexo feminino e dois do sexo masculino), todos com treino fonético, e cuja proveniência de dentro dos EUA foi identificada. As variantes da lateral /l/ foram sustentadas durante 13 a 16 s permitindo a aquisição de 4 a 5 imagens, num determinado plano. As variantes *light* e *dark* foram produzidas no contexto de vogais neutras, tendo o mesmo

som sido repetido cerca de 6 a 9 vezes (com pausa de 3 a 10 s entre repetições) para permitir a aquisição de imagens ao longo de todo o tracto vocal. A aquisição das vibrantes seguiu metodologia semelhante. O protocolo de aquisição de imagem foi o mesmo utilizado em Narayanan, 1995 , já descrito.

1999 - Yang, B. "Measurement and synthesis of the vocal tract of Korean monophthongs by MRI."

O estudo de Yang, obtido num falante (masculino) da língua coreana teve como objectivos: (1) a obtenção de imagens médio-sagittais durante a produção das vogais da língua coreana, (2) obter as áreas transversas do tracto vocal desde os lábios até à glote, (3) obter um coeficiente de correlação entre os valores das formantes, estimados a partir das funções de área e das produções reais, (4) testar perceptualmente as vogais sintetizadas a partir dos valores das formantes, calculadas com base nas áreas obtidas e (5) averiguar quanto varia cada vogal em relação aos valores médios das vogais extremas (*corner* ou *point vowels*)

Foi utilizado um equipamento de 1,5 Tesla, e efectuada sequência FLASH 2D com espessura de corte de 10 mm e obtidas imagens nos planos coronal (quatro cortes na cavidade oral), coronal oblíquo (2 a 3 cortes na oro-faringe e velo), médio-sagittal (para referência) e transverso (3 cortes na faringe e laringe) durante a produção dos monotongos Coreanos. O tempo de aquisição foi de 19s por som.

1998 - Badin, Pouchoy et al. "Un Modele Articulateur Tridimensionnel du Conduit Vocal Base Sur des Donnees IRM".

Este estudo foi efectuado num falante da língua francesa:

- Para cada articulação do corpus 3D, foram efectuados 54 cortes ortogonais em relação ao plano sagittal, distribuídos por 3 ranges, cada uma das quais foi orientada de forma a obter imagens, tanto quanto possível, perpendiculares ao tracto vocal. Desta maneira os autores obtiveram imagens do tracto vocal, com sobreposição parcial, desde a laringe até aos lábios;
- A sequência utilizada foi Eco de *Spin*, com espessura de corte de 3,6 mm, e espaçamento de 4 mm entre os cortes, matriz 256x256, com resolução final de 1 mm/pixel;
- O tempo de aquisição foi de 43 s, durante os quais o falante susteve de forma artificial a articulação, respirando suavemente;
- Para o corpus 2D o tempo de aquisição foi de 11s para cada articulação, matriz 256x256.

1999 - Engwall, O. and P. Badin. "Collecting and analysing two and three dimensional MRI data for Swedish."

Um estudo muito completo efectuado para a língua Sueca. Os objectivos para o estudo foram: (1) aquisição de informação bi e tridimensional, através de imagens RM, relativa ao tracto vocal de um

falante da língua Sueca, para um conjunto significativo de sons e (2) obter alguns parâmetros articulatórios e geometria do tracto vocal de forma a permitir a modulação do mesmo.

Foram utilizados 2 corpora distintos para aquisição de informação 2D (denominado *midsagittal corpus*) e 3D (*3D corpus*), uma vez que, devido ao tempo de aquisição relativamente longo, só foi adquirida informação 3D para um subconjunto de 43 sons do corpus 2D (foram seleccionados os sons considerados mais importantes).

Corpus 2D (ou sagital)

- Constituído por todas as vogais longas do Sueco e por todas as vogais curtas com articulação diferente da versão longa;
- Constituído pelas consoantes nasais, fricativas não vozeadas ou surdas, oclusivas, líquidas e pelos sons retroflexos;
- As vogais foram pronunciadas como se estivessem colocadas em palavras de referência e as consoantes foram produzidas em contexto VCV (vogal-consoante-vogal);
- No total foram obtidas 88 articulações.

Corpus 3D

- Vogais do conjunto sagital que possam ser sustidas durante um período de aquisição mais longo - 10 vogais longas e 3 curtas pronunciadas tal como na palavra de referência indicada;
- As consoantes neste corpus foram as oclusivas não vozeadas /p,t,k/, as líquidas /l,r/, as fricativas /f, s/ os três diferentes sons (sj) do Sueco, todas elas produzidas em contexto VCV (p.e. *apa, ata, aka*) num total de 47 configurações diferentes;

As imagens RM foram obtidas no Centre Hospitalier Regional Universitaire de Grenoble, França em colaboração com o ICP (Institut de la Communication Parlée) segundo protocolo desenvolvido por Badin et al. (1998) para a língua francesa (ver figura 5.2). Todas as imagens foram adquiridas numa única sessão de 3 horas, num equipamento Philips Gyroscan T10-NT de 1 Tesla.

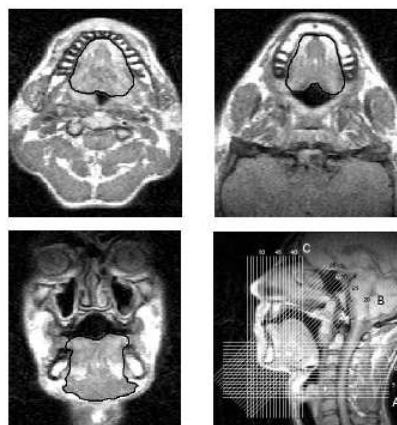


Figura 5. 2 - Contorno da língua em: corte axial (em cima à esq), oblíquo (em cima à dta), corte coronal em (baixo à esq.) e range de cortes projectada numa imagem sagital durante a produção de /ipi/ (em baixo à dta). Fonte: Engwall (2000a).

Foi obtida informação das arcadas dentárias do falante. Foi efectuado um molde das impressões dentárias e do palato duro que foi depois preenchido com água; realizou-se RM do molde com o mesmo equipamento (efectuados 35 cortes coronais para o palato e 41 para a mandíbula e cerca de 48 imagens sagitais)

Não foi efectuada gravação do sinal de voz durante a aquisição, mas imediatamente antes e depois das aquisições dentro do túnel de RM.

2000 - Kröger, B. J., R. Winkler, et al. "Estimation of vocal tract area function from Magnetic Resonance Imaging: Preliminary results."

O estudo de Kröger e colaboradores teve como objectivo a obtenção de informação 3D do tracto vocal utilizando, exclusivamente, imagens RM obtidas no plano sagital. Foram estudadas as seis vogais longas da língua alemã num falante do sexo masculino. Foi utilizado um equipamento de 1 Tesla, tendo sido efectuados 21 cortes sagitais paralelos, não contíguos com espessura de 3.5 mm (com excepção do /u/ que foi obtido com 5 mm). O tempo de aquisição foi de 21s, para cada som tendo sido efectuada apenas uma articulação por cada vogal.

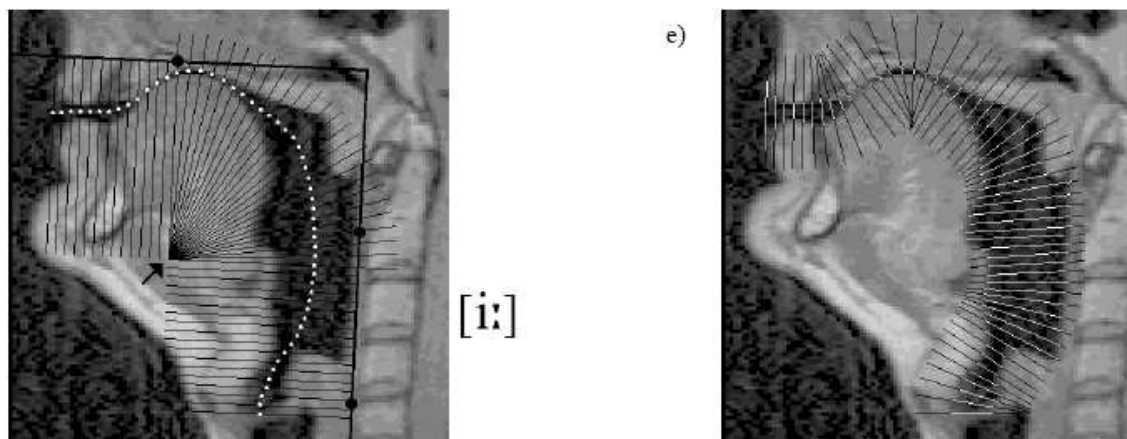


Figura 5.3 - Plano médio sagital para a vogal /i/. À esquerda, grelha inicial para determinação da linha central do tracto vocal (vocal tube centerline) e à direita, grelha final para estimar as áreas transversais do tracto vocal. Fonte: Kröger et al. (2000)

2002 - Gick, B., A. M. Kang, et al. "MRI evidence for commonality in the post-oral articulations of English vowels and liquids."

Estudo efectuado para o Inglês Americano, cujos objectivos foram: (1) Comparar as configurações médio sagitais estáticas do tracto vocal ao nível da úvula e da faringe para as consoantes líquidas /r/ e /l/ com a configuração de cada uma das vogais do inglês (EUA) e (2) testar a hipótese de que, de todas as vogais, a configuração para o (schwa) será mais próxima do /r/ e a do /O/ a mais próxima do /l/.

O corpus contemplou informação de 3 falantes, todos com treino fonético, (um do sexo masculino e dois do sexo feminino) de diferentes zonas dos EUA. As consoantes líquidas /r/ e /l/ foram obtidas para todos os falantes (em posição medial e final de palavra mas nunca em posição inicial). Obtiveram-se imagens de 11 vogais para dois dos falantes e 10 para um dos falantes.

As imagens foram obtidas num equipamento de 1.5 Tesla, com sequência FSE obtendo três cortes sagitais de 5 mm para dois dos falantes, sendo cada configuração repetida cinco vezes para as vogais e quatro para as consoantes. No terceiro falante, foi utilizada uma sequência fast SPGR, tendo sido obtida apenas uma repetição por cada som. Para qualquer dos falantes foi utilizada uma palavra de referência (com 2 sílabas com a vogal em causa, utilizando o sistema de comunicação da sala do magneto). A aquisição começava quando os falantes iniciavam fonação e durante a aquisição (15-16 s) o sujeito mantinha a fonação.

2003 - Demolin, D., V. Delvaux, et al. "Determination of velum opening for French nasal vowels by magnetic resonance imaging."

Este estudo de Demolin e colaboradores teve como objectivo a obtenção do movimento de abertura do velo para as vogais nasais do francês. Participaram quatro falantes, todos falantes nativos do Francês e residentes em Bruxelas. Cada falante pronunciou e susteve cada uma das

vogais nasais e correspondentes vogais orais. Foi fornecida uma palavra de referência com a vogal em causa, uns segundos antes da aquisição de imagens. As imagens foram obtidas num equipamento de 1,5 Tesla utilizando sequência TSE ponderada em DP, 18 cortes axiais com 4 mm de espessura. Tempo de aquisição se 13,8 s. Os cortes axiais foram marcados acima da epiglote e perpendiculares à parede da oro-faringe, numa imagem sagital de referência obtida durante uma articulação igual.

2005b - Serrurier & Badin “Towards a 3D articulatory model of velum based on MRI and CT images”

Foi objectivo deste estudo a obtenção de informação tridimensional relativa à geometria dos articuladores e cavidades envolvidas na nasalidade (velo, cavidades nasais, passagem velofaríngea e seios peri-nasais) com o intuito de desenvolver um modelo articulatorio (3D) do velo com base na informação proveniente de um falante da língua francesa.

Corpus foi constituído por um conjunto de articulações sustentáveis que incluíram as vogais orais e nasais do Francês e as consoantes [p t k f s m n l] em três contextos simétricos [a i u]. O corpus foi complementado com duas articulações específicas, frequentes no discurso: articulação em descanso (em posição de descanso) e articulação pré-fonatória (na fase preparatória antes da fonação). No total foram obtidos 46 fonemas da Língua Francesa.

Foram obtidas imagens com RM e TC, não sendo fornecida informação relativa ao tipo de equipamentos utilizados no estudo. No protocolo de avaliação por RM, foram efectuados 25 cortes no plano sagital, o tempo de aquisição foi de 45 s, durante os quais o falante produziu de forma sustentada cada uma das 46 articulações do corpus. Foi efectuada uma aquisição TC ao crânio, em repouso, para servir de referência; adquiriram-se um conjunto de 149 imagens axiais (desde o pescoço até à porção superior do crânio), com matriz de 512 x 512 e resolução de 20 pixels/cm.

As imagens de TC permitiram a distinção de estruturas ósseas/ tecidos moles /ar e foram usadas como referenciais para localização das estruturas ósseas no processo de segmentação.

2005 - Ericsson - “Articulatory - Acoustic relationships in Swedish vowel sounds”

Estudo efectuado para a língua Sueca, teve como objectivo principal avaliar a *performance* de um método frequentemente utilizado, em estudos de produção de fala, para predição das áreas transversais do tracto vocal a partir das distâncias transversas. Em conjunto com as imagens de RM foram efectuadas, em simultâneo, gravações áudio e vídeo. O estudo incluiu a aquisição de imagens em 2 falantes, um do sexo feminino e outro do sexo masculino sendo o corpus constituído pelas nove vogais longas da língua Sueca e dois alofones curtos. Foi desenvolvido um método de integração das arcadas dentárias nas imagens RM e sistema de gravação adequado ao ambiente ruidoso durante as aquisições de imagem por RM. A aquisição de imagem foi efectuada num equipamento Philips de 1,5 T, recorrendo a sequência TSE ponderada em densidade protónica. A espessura de corte foi de 4 mm e o tempo de aquisição variou entre os 13 e 14 s por som. Foram efectuados cortes sagitais (onze) e 14 ou 15 cortes axiais e coronais

(*imagens multimode*). A partir das imagens adquiridas foram efectuadas medidas articulatórias da laringe, faringe, cavidade oral e estimativas relativas ao *terminus* do tracto vocal. No estudo foram ainda co-registadas imagens provenientes de Raios X (cineradiografia) obtidos nos mesmos falantes em estudos prévios. Deste co-registo foi possível observar diferenças e semelhanças do ponto de vista articulatório, obtidas em condições muito semelhantes.

5.3 Estudos dinâmicos e em tempo real

A aquisição de imagens RM, que permitam observar o movimento do tracto vocal durante a fala, tem sido efectuada por diferentes grupos de investigação, recorrendo a técnicas e estratégias diferentes (Shadle, Mohammad, Carter, & Jackson, 1999). Alguns grupos (Demolin et al., 1997; Demolin, Metens, & Soquet, 2000; Narayanan, Nayak, Lee, & Byrd, 2004) recorreram a técnicas e sequências que permitem aquisições muito rápidas (4, 5 e 9 *frames/s* respectivamente), outros utilizaram métodos de sincronização com um estímulo externo como seja, por exemplo, o batimento cardíaco, ou recorrendo a um *trigger* repetitivo emitido em simultâneo para o falante e para o equipamento (Masaki et al., 1997; Takemoto, Kitamura, Nishimoto, & Honda, 2004). Em ambas as situações, a produção do falante tem que estar sincronizada com a aquisição e a reconstrução é depois efectuada automaticamente. Outros autores (Mathiak et al., 2000; Mohammad, 1999; Shadle, Mohammad, Carter, & Jackson, 1999) utilizam uma técnica em que uma palavra é repetida um número elevado de vezes, durante o qual se adquirem as imagens e se grava o sinal de voz, em simultâneo. As imagens são depois reconstruídas dando origem a uma pseudo variação temporal do tracto vocal, permitindo uma resolução temporal aparente bastante elevada, de cerca de 16 ms, tal como referenciado no estudo de Mohammad (1999).

Convém, no entanto esclarecer que existe alguma confusão e dificuldade em perceber claramente o que nos aparece referenciado como estudo dinâmico ou um estudo em tempo real. Narayanan et al. (2004) referem precisamente essa dificuldade e esclarecem de que forma utilizam, no seu artigo, uma ou outra designação. Assim e citando:

“We use the term *dynamic* to refer to the creation of images from an actively articulating subject, rather than a static postural source. We reserve the term *real-time* MRI to refer specifically to directly capturing or acquiring moving image data in real time. That is, *dynamic* refers to the source, *real time* to its acquisition. The *dynamic* MRI technique provides a way to capture valuable Kinematic information. It relies on gated scanning on numerous repetitions of the same speech sequence to reconstruct the impression of articulatory movement in time”

De seguida, e tal como foi efectuado para os estudos estáticos faz-se um pequeno resumo de alguns estudos dinâmicos e em tempo real. Os estudos seleccionados pretendem ilustrar as diferentes abordagens e estratégias que têm sido seguidas pelos diferentes grupos de investigação na obtenção de informação dinâmica.

1996 - Crary, M. A., M. Gorham, et al. "Dynamic magnetic resonance imaging in the study of vocal tract configuration."

Descrever a utilidade da RM dinâmica no estudo da configuração do tracto vocal

- Foi efectuada a descrição da técnica de aquisição e os autores apresentam 2 estudos preliminares, o primeiro onde foram obtidas medidas relativas a um grupo normal de falantes e um segundo grupo, um grupo de doentes com disфонia espasmódica, antes e após injeção de Botox.

Aquisição de imagem:

- Foi utilizado um Equipamento Magnetom, Siemens de 1Tesla e antena de cervical na aquisição das imagens;
- Foi utilizada uma sequência Eco de Gradiente, ponderada em densidade protónica, com os seguintes parâmetros: TR/TE=4.9 /2 ms, Flip Angle de 10°, 1 NEX;
- Foram obtidos cortes sagitais (na mesma posição) e cortes axiais em diferentes posições, com 8 mm de espessura. O FOV foi de 370 mm e matriz de 64 x 128, resultando numa resolução de ~6 mm na direcção de codificação da fase (direcção Antero-Posterior para os cortes sagitais e Direito-Esquerdo para as imagens axiais) e 3 mm na direcção de codificação de frequência. O tempo de aquisição para cada imagem foi de 327 ms;
- Foram efectuados cortes sagitais e axiais de referência, os cortes sagitais foram adquiridos na mesma posição (linha média da laringe), pretendendo-se avaliar alterações temporais;
- Os cortes axiais foram marcados paralelamente às pregas vocais, começando ao nível da naso-faringe e terminando na porção proximal da traqueia.

1997 - Demolin et al. "Coarticulation and articulatory compensations studied by dynamic MRI."

Objectivos:

- A partir de uma adaptação de uma sequência TSE, os autores esperavam conseguir a monitorização dinâmica e contínua do tracto vocal, com uma resolução temporal de 4 imagens por segundo. Esta técnica permitiria estudar os movimentos relativos dos principais articuladores envolvidos na produção de fala: lábios, língua, laringe, maxilar inferior e velo;

Corpus:

- 4 falantes (2 homens e 2 mulheres);

- Os falantes foram instruídos para pronunciar devagar, e em duas experiências diferentes, uma sequência de pseudo-palavras e segmentos sem sentido: [tagy], [tyga.], [iui], [iai], [ieaou] e [beben]. Pronunciando as palavras devagar foi permitido visualizar os movimentos dos principais articuladores;

Aquisição de imagem:

- Foi utilizado um equipamento Philips Gyroscan ACS NT, de 1.5 Tesla e antena de quadratura de pescoço.
- A sequência utilizada foi uma adaptação de uma sequência TSE denominada LoLo (Local Lock) TSE, com ponderação T1, os parâmetros utilizados foram os seguintes: TR/TE=250/30 ms, FA=60°, ETL= 19 ms, Partial Fourier acquisition de 60%, FOV de 300 x 150 mm com matriz de 32x128 elementos. Foi adquirido um corte sagital, com 6 mm de espessura, de forma contínua durante aproximadamente 20 s.
- Esta sequência utiliza um pulso inicial de excitação de 60°, e pulsos de refocalização subsequentes de 180°. Com esta sequência pode ser utilizado um FOV reduzido sem artefactos de enrolamento permitindo melhorar a resolução temporal ou espacial.

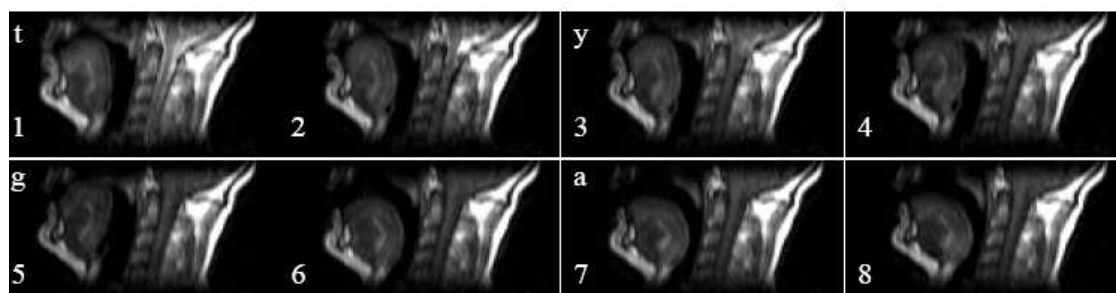


Figura 5. 4 - Produção da palavra [tyga] pelo falante 2. Fonte: Demolin et al. (1997).

Em 2000 (Demolin, Metens & Soquet, 2000), efectuaram estudo semelhante (Real Time MRI and articulatory coordinations in vowels) cujo principal objectivo foi obter descrição espacial detalhada, dos principais articuladores (protusão e abertura labial, abertura e retracção mandibular, posição do dorso da língua, abertura do velo, altura da laringe e avanço da raiz da língua) durante a produção de uma sequência de vogais /ieaou/. Utilizaram sequência com parâmetros idênticos ao estudo efectuado em 1997, mas utilizando gradientes mais rápidos. Assim, foi possível baixar o TE cerca de 9 ms, o que permitiu melhorar a qualidade de imagem (diminuição dos artefactos de susceptibilidade magnética) e permitiu a obtenção de 5 *frames/s*.

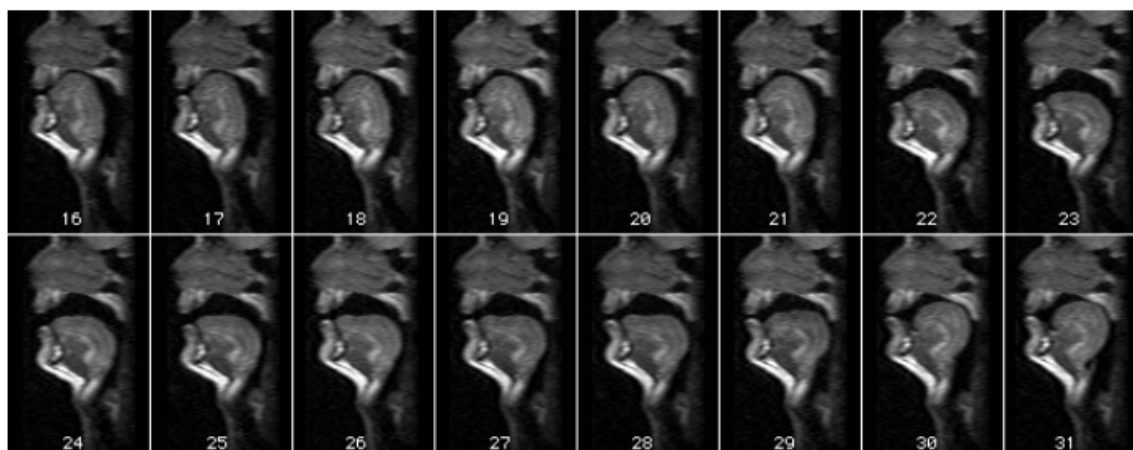


Figura 5. 5 - Aquisição em tempo real da repetição da sequência /ieaou/ produzida pelo falante do sexo masculino. As 16 imagens representam um total de 3.2s, com uma imagem a cada 200 ms. Fonte: Demolin, Metens & Soquet, 2000.

1999 - Shadle, C. H., M. Mohammad, et al. "Multi-planar dynamic magnetic resonance imaging: new tools for speech research."

Estes autores descrevem a obtenção de imagens dinâmicas com RM, utilizando uma técnica em que as imagens obtidas durante a repetição de uma palavra, são reconstruídas utilizando um sinal áudio obtido durante a produção, para formar imagens que apresentam uma pseudo variação temporal do tracto vocal.

Neste caso, a palavra [pasi] foi repetida 360 vezes, tendo sido efectuada aquisição de 3 cortes sagitais. Obteve-se assim, uma sequência de 39 conjuntos de 3 cortes. Para cada *frame* (16 ms) foi reconstruído um volume. Desta maneira, foi possível medir a distância palato-língua para [a, s, i]; estimou-se o volume da língua e calculou-se a função de área do tracto vocal usando, só o corte médio-sagital e depois o conjunto dos três cortes.

- Foi utilizado equipamento GE SIGNA de 0.5 Tesla. FOV= 240mm, foram efectuados 3 cortes sagitais com 5 mm de espessura, 1 marcado na linha média e os outros afastados 11 mm para a direita e para a esquerda, com sequência Eco de Gradiente (Fast RF-spoiled). O TR=16 ms, sendo TA para os 3 cortes cerca de 6 s.
- O falante (cuja língua mãe é o Inglês Britânico) repetiu a palavra [pasi] cerca de 12 vezes por cada aquisição, tendo sido efectuadas 24 aquisições. Foram efectuadas cerca de 360 repetições desta palavra.
- Para comparação foram efectuadas aquisições durante as produções [a,s,i] sustidas durante 6 s.
- O som foi gravado durante a aquisição de imagens, utilizando o intercomunicador, que apesar do ruído permitiu a segmentação necessária para o pós-processamento.

2000 - Mathiak, K., I. Hertrich, et al. "Stroboscopic articulography using fast magnetic resonance imaging."

O corpus para este estudo foi constituído por:

- 3 voluntários (2 mulheres e 1 homem) sem sinais de alterações neurológicas após exame médico;
- Fizeram parte do corpus, 4 sons monosilábicos iniciados pela mesma oclusiva /g/ , seguido de vogal que varia e termina com uma nasal: [gɔŋ], [gaŋ], [giŋ], [gɛŋ]. Dependendo do contexto (vogal) será de esperar alterações no ponto de articulação das consoantes;
- Cada falante produziu 170 repetições de cada uma destas palavras de forma sincronizada com o barulho produzido pelas bobines de gradientes (1.2 produções/s)

O procedimento para aquisição de imagem baseou-se numa estratégia estroboscópica¹³. Cada linha do espaço K, foi adquirida 64 vezes por cada sílaba. Com esta técnica o espaço K é preenchido linha a linha ao longo das 170 repetições da palavra. Assumindo que as repetições produzidas pelo falante são idênticas, as imagens são reconstruídas combinando a diferente informação obtida para cada linha, numa fase específica do movimento. É obtido um corte para cada ponto temporal. Os cortes são depois combinados de maneira a obter um filme com frequência de 120 ciclos/s ou seja TR⁻¹.

- Equipamento (Siemens, Vision, Erlanger, Germany) de 1.5 Tesla e antena de Crânio;
- Sequência FLASH com TR/TE = 8.4 /3.4 ms para minimizar artefactos de susceptibilidade e reduzir o tempo de aquisição, realização de corte no plano médio-sagital;
- FOV de 200x200 mm, matriz 256x256, FA de 10° e espessura de corte de 8 mm. A fase de aquisição das 64 *frames* e o atraso inter-ciclos de 300 ms deram origem a intervalo inter-estímulos de 840 ms. Uma vez que as produções foram sincronizadas com o barulho produzido pelas bobines de gradiente, foi obtido um ritmo de 1.2 sílabas/s.

¹³ Estroboscopia – A ideia subjacente à estroboscopia é obter imagens de um acontecimento, em sincronia, com um flash luminoso. Se as imagens forem adquiridas a uma razão semelhante, mas não igual à frequência da oscilação do objecto, o resultado é uma análise em “câmara lenta”.

2001 - Stone, M., Davis, E., Douglas, A., et al. Modeling the motion of the internal tongue from tagged cine-MRI images.

Stone e colaboradores, utilizando a técnica Cine Tagged MRI (muito utilizada em estudos cardíacos para observar a deformabilidade do miocárdio) pretenderam estudar a língua, articulador que apresenta um complexo e altamente variável processo de deformação durante a produção de fala, causado por uma estimulação diferencial de uma musculatura anatomicamente complexa.

A aquisição das imagens foi efectuada num equipamento Picker de 1,5 Tesla. Foi usada técnica CineTagged, sendo efectuados três cortes no plano sagital. A informação foi adquirida num falante, durante a repetição de sílabas do tipo CV ($k\alpha$). Durante cada repetição da sílaba, foram adquiridas 7 fases temporais a uma razão de 56 ms por fase temporal. A sequência utilizada foi EG (TR/TE=14/4,6 ms) FOV de 24 cm e espessura de corte de 7 mm. Foi utilizada, uma série de 5 pulsos RF (com amplitudes relativas de [+1 -4 +6 -4 +1]), para produzir linhas pretas (*tag lines*) separadas cerca de 11 mm. Este pulso complexo, foi aplicado sequencialmente ao longo dos eixos x e y durante cerca de 30 ms, resultando numa grelha horizontal e vertical de *tag lines*. Sete fases temporais foram adquiridas durante os 392 ms seguintes, acompanhando o movimento da língua para a vogal. Para reconstruir as sete fases temporais, foram necessárias 32 repetições da sílaba, uma vez que, apenas 4 das 128 linhas do espaço K são adquiridas por repetição ($128/4 = 32$). Com esta técnica é possível obter informação acerca da deformação da língua com uma resolução de aproximadamente 2 mm. Quando comparado com outras técnicas dinâmicas, permite a obtenção de informação com menos repetições.

2002 - Mády, K., R. Sader, et al. "Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction."

Estudo realizado em 100 indivíduos com perturbações articatórias pós glossectomia e reconstrução oral, com o objectivo de avaliar os movimentos da língua na produção das consoantes linguais da língua germânica. Fizeram parte do corpus, nove palavras alemãs com estrutura simples do tipo CV, nas quais, todas as vogais e consoantes linguais do Alemão Moderno foram contempladas.

Foi utilizado equipamento Philips ACS NT Gyroscan com as antenas de crânio e de pescoço e foi efectuada sequência EG, ponderada em T1 com SENSE (Fast Gradient Echo) com TR/TE=4 /1.1 ms, FA de 6°, espessura de corte 10 mm permitindo a aquisição de 8 imagens/s.

2003 - Takemoto, H., K. Honda, et al. "Measurement of temporal changes in vocal tract area function during a continuous vowel sequence using a 3D cine-MRI technique."

Estudo efectuado num falante da língua Japonesa, com o objectivo de obter as variações temporais nas funções de área do tracto vocal durante a produção das 5 vogais japonesas de forma contínua e, comparar os espectros obtidos a partir das funções de área, com a fala natural. O corpus foi constituído pela sequência das vogais /a i u e o/ repetida de forma contínua durante a aquisição de imagens.

A aquisição das imagens foi efectuada num equipamento de 1,5 T com técnica RM 3D cine que permitiu obter um volume de informação da região oro-facial durante os movimentos articulatorios. A aquisição de imagem foi efectuada com sincronização da aquisição com a produção do falante utilizando o método de amostragem sincronizada (Synchronized Sampling Method-SSM) desenvolvido por Masaki et al (1997, 1999) para o modo de aquisição 2D cine (ver figura 5.6).

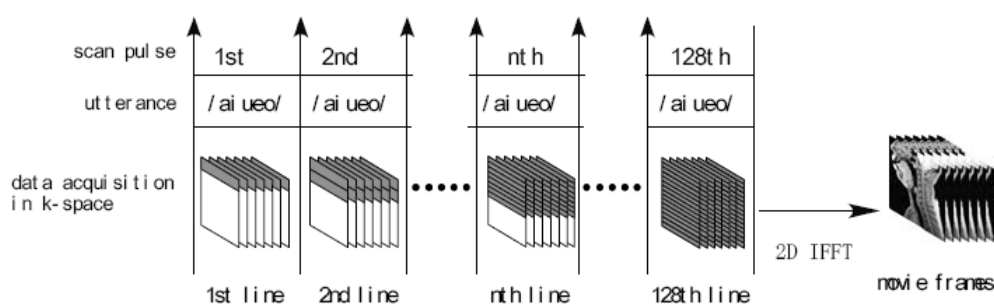


Figura 5. 6 - Procedimento para o método de amostragem sincronizada (SSM). Fonte: Takemoto et al. (2003).

O estímulo auditivo apresentado a partir do *trigger* externo é um conjunto de quatro ruídos, com duração de 100 ms num ciclo de 2000 ms, que auxilia o falante a sincronizar a produção com a aquisição de imagem. Enquanto o falante repete a sequência das 5 vogais (640 vezes com intervalo de 2000 ms) são adquiridas imagens no plano sagital, desde a posição 43 mm à direita até 39 mm à esquerda, obtendo um filme dos 56 volumes a uma razão de 30 *frames/s* (ver Figura 5. 7).

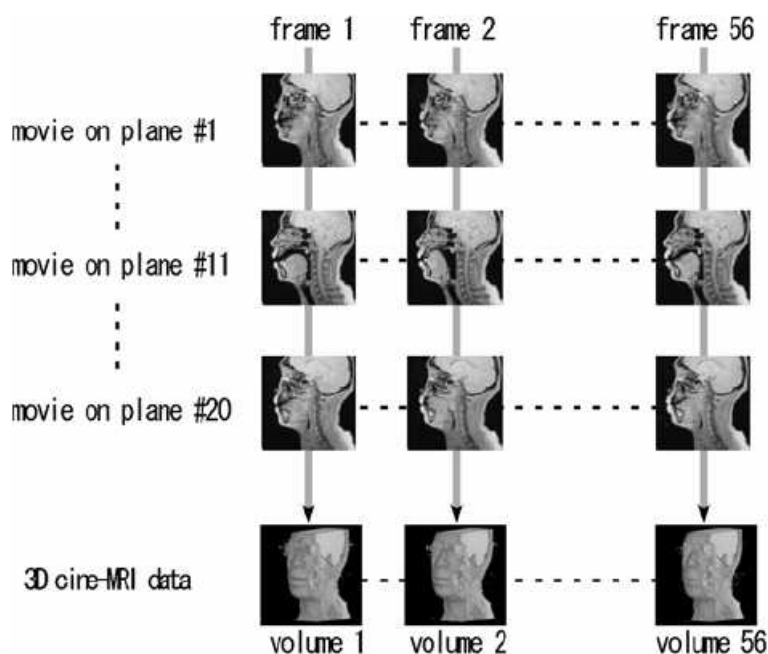


Figura 5. 7 - Processo de reorganização dos dados para obter 56 volumes. Fonte: Takemoto et al. (2003).

2004 - Engwall, O. "From real - time MRI to 3D tongue movements."

Este estudo teve como objectivos, utilizar um método de inversão articulatória para extrapolar os movimentos da língua no plano sagital para movimentos a 3 dimensões e determinar a precisão do modelo da língua obtido a partir de configurações sustidas artificialmente na replicação dos movimentos deste articulador em tempo real.

- Para obtenção das imagens em tempo real, foi adquirida informação produzida pelo mesmo falante que produziu informação para obtenção do modelo 3D da língua e que se encontra descrito com mais detalhe em Engwall (2000a)
- Foi utilizado equipamento Philips de 1.5 T com gradientes rápidos (CompactPlus, Power Track 6000, 20 mT m^{-1} e 100 mT $m^{-1}ms^{-1}$ de amplitude máxima e *slew rate*, respectivamente) e antena de quadratura de pescoço.
- Foi efectuada sequência TSE Zoom, com ponderação T1 no plano médio sagital, espessura de corte de 6 mm e matriz de 128x128 permitindo obter 9 *frames/s*.

2004 - Narayanan, S., K. Nayak, et al. "An approach to real-time magnetic resonance imaging for speech production."

Narayanan e colaboradores pretenderam com este estudo avaliar a dinâmica do tracto vocal durante a produção de fala, utilizando sequência de pulsos que permite a aquisição de imagem em tempo real. Foram efectuados 2 estudos, um (estudo 1) para o Inglês (EUA) e outro (estudo 2) para Tamil. No primeiro estudo participaram 2 falantes (SN e KN), e no estudo 2 apenas 1 falante (SN). No estudo 1 foi utilizada uma frase de suporte " Say _____again" para a produção de palavras onde muda a posição da sílaba de /n, r, l/. No estudo 2 foram produzidas frases em Tamil

que incluíam as 5 líquidas. Foi possível obter imagens do tracto vocal em toda a sua extensão e observar os movimentos dos lábios, língua e velo em tempo real.

O equipamento utilizado tinha intensidade de campo de 1.5 Tesla e estava equipado com gradientes de alta *performance*. Foi utilizada uma sequência Eco de Gradiente rápida com modo de aquisição espiral (Fast spiral gradient echo pulse sequence - ver Figura 5.8) com espessura de corte de 5 mm (seria possível com a mesma sequência baixar a espessura para 3 mm à custa de degradação da relação S/R). Nesta sequência, um pulso de excitação de 1.4 ms (640 μ s rf) é seguido por uma espiral de leitura de 2.4 ms e por um *spoiler*. São efectuadas 20 interpolações (*interleaves*) para obter 112 pixel ao longo de um FOV de 20 cm. São adquiridas imagens em cada 110 ms (cerca de 8/9 *frames* por segundo) sendo reconstruídas a cerca de 24 *frames/s* utilizando janelas deslizantes (*sliding window*).

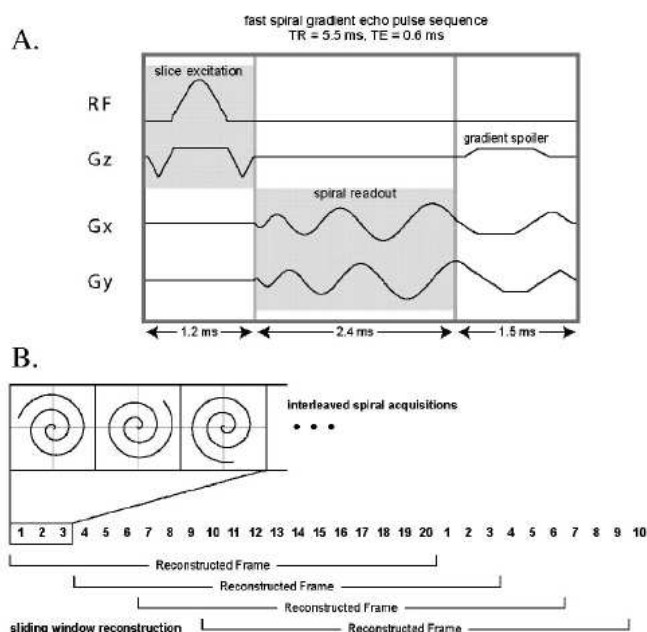


Figura 5. 8 - (a) Sequência de pulso utilizada na aquisição de imagens em tempo real e (b) Esquema de aquisição temporal. Fonte: Narayanan et al. (2004).

2006 - Fabiana Gregio “Configuração do tracto vocal supraglótico durante a produção das vogais do Português Brasileiro: dados de ressonância magnética”.

Este estudo foi efectuado para o Português Brasileiro (PB), para um falante, com o objectivo de observar a configuração do tracto vocal supra-glótico durante a produção das sete vogais orais e cinco vogais nasais do PB. Foi utilizado equipamento GE de 1,5 Tesla, e sequência real-time permitindo obter cerca de 5,8 frames/s. Apesar da aquisição ser em tempo real, o falante susteve a produção da vogal durante cerca de 17s de aquisição (obtenção de 100 imagens) por cada série

de quatro vogais. A autora verificou que as vogais nasais do PB são produzidas em três fases distintas (caracterizadas por mudanças no posicionamento dos articuladores durante a emissão) ao contrário das orais que apresentam uma única fase.

5.4 Sumário

Como se verificou, em relação aos estudos estáticos houve alguma evolução a nível técnico conseguindo-se uma significativa redução do tempo de aquisição, quando comparado com os primeiros estudos, o que é vantajoso uma vez que o falante sustém a produção durante um período significativamente mais curto. Este facto é particularmente relevante quando se pretende obter informação 3D uma vez que, como é necessário um número elevado de cortes o tempo de aquisição é geralmente maior. A par com esta diminuição do tempo de aquisição tem sido também possível obter uma melhor resolução espacial, diminuindo a espessura de corte ou aumentando a matriz, ou ainda, noutras situações aumentando a cobertura anatómica. Globalmente a qualidade de imagem tem vindo a melhorar, devido a menos artefactos relacionados com o movimento.

No que diz respeito, aos processos dinâmicos envolvidos na produção de fala, verifica-se que já é possível obter alguma dessa informação utilizando RM, recorrendo a estudos dinâmicos ou recorrendo a aquisição em tempo real. A resolução temporal conseguida com as técnicas mais avançadas ainda não permite uma resolução temporal equivalente à obtida com a cineradiografia mas os progressos verificados nos últimos anos, neste campo, auguram um futuro promissor em termos de rapidez de aquisição. É fundamental também o desenvolvimento de protocolos, otimizando as sequências em função dos objectivos a alcançar em cada estudo. Um compromisso entre resolução temporal e espacial terá que ser encontrado para cada caso, a minimização de artefactos, a utilização de técnicas de sincronização e gravação adequadas terá que ser aperfeiçoada.

Capítulo 6

Metodologias

Ao longo do capítulo, descreve-se a constituição do corpus e as razões conducentes à sua escolha. Apresenta-se o protocolo RM utilizado para a aquisição do corpus seleccionado para o estudo, faz-se uma descrição detalhada de todos os procedimentos seguidos durante a aquisição das imagens, bem como das sequências e parâmetros utilizados. Faz-se um enquadramento relativo às técnicas de segmentação e descrevem-se, de forma resumida, as técnicas de segmentação de imagem e programas a que se recorreu para a realização deste estudo.

6.1 Corpus

6.1.1 Definição do corpus¹⁴

A definição do corpus a estudar é uma parte importante de qualquer estudo na área da fala, e deverá ter em atenção, não só os objectivos propostos, mas também o tempo disponível para a conclusão do mesmo. Assim, foi definida uma primeira versão do corpus, que foi reavaliada e corrigida, após uma primeira sessão experimental de aquisição de imagens e processamento das mesmas. As limitações temporais e técnicas foram determinantes na selecção do corpus final, nomeadamente no número de sons e no número de informantes a incluir no estudo. O corpus para este trabalho, além dos objectivos já enunciados no primeiro capítulo desta dissertação, foi elaborado com o intuito de poder dar algum contributo no esclarecimento de algumas questões relacionadas com a produção do PE, como por exemplo:

- Será a descrição articulatória das vogais orais demasiado simplificada?
- Serão as vogais nasais do PE bem representadas na sua transcrição fonética, ou seja serão as configurações orais da língua correspondentes ao símbolo base utilizado?
- Qual a posição do velo na produção de consoantes, consoantes nasais e vogais nasais?
- De que maneira se influenciam os segmentos adjacentes numa sequência de sons de fala (coarticulação)?

Para começar a abordagem a estas questões (responder ou resolver na sua totalidade estas questões é uma tarefa que está fora do âmbito deste trabalho, dada a sua complexidade) e porque cada uma delas encerra questões e problemas específicos, foram elaborados, três sub-corpus, com objectivos diferentes, e que identificámos como: Corpus 2D sagital, Corpus 3D, e Corpus Tempo Real. Alguns dos elementos escolhidos foram seleccionados de outros trabalhos já

¹⁴ Corpus é uma base de dados de sinais da fala, utilizado na investigação e desenvolvimento das aplicações em processamento de fala (Ribeiro, 2003).

efectuados para o estudo do PE, nomeadamente no estudo das nasais (Teixeira, Ferreira, Lisboa, Coimbra, & Moutinho, 2003a; Teixeira, Ferreira, Moutinho, Coimbra, & Lisboa, 2004b), fricativas (Jesus, 2001) e, para além destes, os restantes fones foram escolhidos de forma a contemplar a maior parte dos sons do PE. A informação proveniente do corpus 2D e 3D é estática (derivada da produção de sons de forma sustida ou manutenção artificial de uma articulação durante o período de aquisição) e a do corpus obtido em tempo real é dinâmica.

6.1.2 Constituição do Corpus

Corpus 2D

Do corpus 2D, como se pode observar na Tabela 6.1, fizeram parte todas as vogais orais do PE, as vogais e consoantes nasais, as fricativas [f, s, S, v, z, Z] em início de palavra e em contexto VCV, as oclusivas vozeadas [b, d, g], as oclusivas não vozeadas [p, t, k], também em contexto VCV e as laterais [l, L], num total de 64 produções. A descrição detalhada dos fones e palavras de referência utilizadas encontra-se no Anexo 2.

FONES		OBSERVAÇÃO
Vogal	Oral	[i] [e] [E] [@] [6] [a] [u] [o] [O]
	Nasal	[i~] [e~] [6~] [o~] [u~]
Consoante	Nasais	[m] [n] [J]
	Fricativas	[f] [s] [S] [v] [z] [Z]
	Oclusivas	[p] [t] [k] [b] [d] [g]
	Laterais	[l] [L]
		1 produção de cada item
		1 produção de cada item
		1 produção de cada item
		Em início de palavra, 1 produção
		Em contexto VCV, V=[i,u,a]; 3 produções
		Em contexto VCV, V=[i,u,a]; 3 produções
		1 produção de cada item

Tabela 6. 1 - Descrição resumida do corpus 2D Sagital, para o falante PAA.

Corpus 3D (Aquisição em volume)

Do corpus 3D (Tabela 6.2) fizeram parte as vogais orais com excepção do [E] e do [@], todas as vogais nasais do PE, sendo que a vogal [6~] foi produzida duas vezes, uma delas tendo como palavra de referência uma palavra em que a vogal nasal foi precedida de uma consoante nasal [m] e outra em que foi precedida por uma consoante não nasal [k]. Adquiriu-se também informação, relativa às consoantes nasais, fricativas em contexto VCV com as vogais [i a u], e as laterais [l] e [L].

FONES		OBSERVAÇÕES
Vogal	Oral	[i] [e] [6] [a] [u] [o] [O]
	Nasal	[i~] [e~] [6~] [o~] [u~]
Consoante	Nasais	[m] [n] [J]
	Fricativas	[f] [s] [S]
	Laterais	[l] [L]
		1 produção de cada item
		1 produção de cada item e 2 [6~]
		Em contexto VCV, V=[i,a,u]; 3 produções
		1 produção de cada item

Tabela 6. 2 - Descrição resumida do corpus 3D, para o falante PAA.

Corpus 3D (Aquisição clássica)

Foi decidido efectuar uma aquisição para obtenção de informação 3D, semelhante à que vulgarmente tem sido descrita na literatura (obtida a partir de cortes 2D axiais, coronais e oblíquos) para comparação com a informação por nós adquirida com sequência 3D. Assim, foram seleccionados apenas alguns dos fones do corpus 3D, nomeadamente as três vogais cardinais [i], [a] e [u], a nasal [ɐ~], a consoante nasal [m] a fricativa [f] e as fricativas [s] e [ʃ] e as laterais [l] e [ʎ].

Corpus Tempo real

Do corpus em tempo real fizeram parte, tal como se pode observar na Tabela 6.3, duas sequências de vogais, uma oral e outra nasal e pares de palavras como [matu] e [mɐ~tu] , [karu] e [kaRu]. Foram também adquiridas algumas imagens estáticas (2D) para permitir a identificação das frames correspondentes, obtidas em tempo real.

PALAVRAS	PLANO	CODIFICAÇÃO
Sequência [a E i O u]	Sagital	rts_aeiou
Sequência [ɐ~e~i~o~u~]	Sagital	rts_ɐ~e~i~o~u~
Manto [mɐ~tu]	Sagital	rt_manto_sag
Manto	Oblíquo	rt_manto_obl
Mato [matu]	Sagital	rt_mato_sag
Mato	Oblíquo	rt_mato_obl
Minto	Sagital	rt_minto_sag
Minto	Oblíquo	rt_minto_obl
Mito	Sagital	rt_mito_sag
Mito	Oblíquo	rt_mito_obl
Canto [Kɐ~tu]	Sagital	rt_canto_sag
Canto	Oblíquo	rt_canto_obl
Cato [katu]	Sagital	rt_cato_sag
Cato	Oblíquo	rt_cato_obl
Carro	Sagital	rt_carro
Caro	Sagital	rt_caro
Quatro	Sagital	rt_quatro
Coalho	Sagital	rt_coalho
Água	Sagital	rt_agua
Cume	Sagital	rt_cume

Tabela 6. 3 - Corpus obtido em tempo real, para a falante MTC.

FONES	PALAVRA REFERÊNCIA	CODIFICAÇÃO
[a]	pato	tse t1 TC - pato
[E]	leva	tse t1 TC-leva
[i]	pipo	tse t1 TC-pipo
[o]	tôpo	tse t1 TC-topo
[O]	pote	tse t1 TC-pote
[u]	buda	tse t1 TC-buda
[6~]	manto	tse t1 TC-manto
[e~]	penete	tse t1 TC-pente
[i~]	pinta	tse t1 TC-pinta
[o~]	ponte	tse t1 TC-ponte
[u~]	punto	tse t1 TC-punto
[m]	cama	tse t1 TC-cama
[l]	pála	tse t1 TC-pala
[t]	pote	tse t1 TC-pote
[k]	maca	tse t1 TC-maca
[r]	caro	tse t1 TC-caro
[R]	carro	tse t1 TC-carro

Tabela 6. 4 - Fones adquiridos (estáticos) 2D sagital, para identificação das imagens em tempo real, obtidas com a falante MTC.

6.1.3 Informantes¹⁵

Os informantes participantes neste estudo foram dois, um do sexo masculino e outro do sexo feminino, com as seguintes características:

1. PAA, indivíduo do sexo masculino com 25 anos de idade, altura 180 cm, peso 70 Kg, Licenciado em Terapia da Fala, proveniente do Norte do País, língua mãe - Português Europeu, com treino fonético e de canto.
2. MTC, informante do sexo feminino com 41 anos de idade, altura 165 cm, peso 78 Kg, Licenciada em Radiologia, natural do distrito de Coimbra, língua mãe - Português Europeu, sem treino fonético.

Nenhum dos informantes apresentava, quando da colheita de imagens, qualquer tipo de perturbação da fala e/ou da linguagem. O informante PAA fez a aquisição do corpus 2D Sagital e 3D (aquisição em volume e aquisição clássica) e com a informante MTC adquiriu-se o corpus em Tempo real. Quando do planeamento deste estudo, pensou efectuar-se a aquisição dos diferentes corpora com um só falante, contudo, após uma primeira sessão experimental (que descreveremos mais adiante), verificou-se que seria difícil obter toda esta informação num só falante, pois dificilmente permaneceria tanto tempo imobilizado e confortável (cerca de 3 a 4 horas) na mesma

¹⁵ Nota: neste texto utilizamos os termos informante ou falante, de forma indiscriminada, para nos referirmos aos indivíduos participantes no estudo. Os termos em inglês são geralmente: "speaker" ou "subject".

posição. Por outro lado, depois de testar algumas sequências com o informante PAA, observámos que com a sequência que se pretendia utilizar para efectuar a aquisição em tempo real (TrueFISP) os artefactos de susceptibilidade magnética (ver ponto 4.6.1), originados por tratamentos dentários, degradavam a qualidade de imagem. Ainda assim, decidi manter-se o informante PAA para a aquisição do corpus 2D e 3D, pois considerou-se que o seu treino fonético e conhecimentos na área da produção de fala seriam uma mais valia para o estudo.

6.2 Recolha de dados

6.2.1 Local e período

A aquisição das imagens decorreu, na sua totalidade, no Sector de Ressonância Magnética do Serviço de Imagiologia dos Hospitais da Universidade de Coimbra, depois de superiormente autorizado e com o apoio técnico que os responsáveis do sector acharam necessário. As imagens foram obtidas durante o mês de Julho de 2006, durante duas sessões, a primeira das quais teve carácter experimental. Ambas as sessões decorreram após o horário normal de funcionamento da RM, de forma a não perturbar os exames programados. As imagens adquiridas foram gravadas em CD e posteriormente em DVD, tendo o processamento de imagem sido efectuado no Instituto de Electrónica e Telemática da Universidade de Aveiro (IEETA), no âmbito do trabalho de Mestrado de Carbone (2007) e do Projecto HERON.

6.2.2 Equipamento e Material

As imagens foram adquiridas num equipamento de 1,5 Tesla (Magnetom Symphony, Maestro Class, Siemens, Erlanger, Alemanha) equipado com gradientes Quantum (máxima amplitude - 30 mT/m; *rise time* - 0,24 ms; *Slew rate* - 125 T/m/s; FOV - 50 cm). Foram utilizadas as antenas de crânio e de pescoço, em simultâneo, em todas as aquisições.

6.2.3 Segurança

Os informantes, antes de efectuarem qualquer sessão de aquisição de imagem, foram sujeitos a rastreio, efectuado verbalmente por uma técnica de radiologia (PM) e autora deste trabalho, seguindo as normas e o inquérito utilizado no sector de RM onde foram adquiridas as imagens. Houve também a preocupação de informar todos os presentes (cuja formação não é da área da radiologia) das normas a seguir quando da entrada na sala do magneto, evitando assim algum incidente. Durante a aquisição das imagens, não foram ultrapassados os limites referentes ao SAR, nem excitação periférica. Os informantes usaram auscultadores durante a aquisição das imagens, respeitando também as recomendações relativas à exposição ao ruído.

6.2.4 Sessão experimental

A realização da sessão experimental, revelou-se fundamental para o sucesso da sessão de aquisição propriamente dita e que efectivamente permitiu a recolha de dados para o nosso trabalho. Foi possível testar algumas sequências e procedimentos que nos permitiram organizar e rentabilizar a sessão de aquisição. Antes mesmo desta sessão, houve algum tempo de preparação em que se fez uma revisão bibliográfica acerca das técnicas já utilizadas (no âmbito dos estudos de fala) para a aquisição de imagens RM, e uma adequação das mesmas ao equipamento disponível e aos objectivos que nos propusemos atingir. Assim, resumidamente testou-se:

- Os próprios informantes - só efectuando algumas imagens, com as sequências a utilizar no protocolo de aquisição foi possível verificar, por exemplo, se o material dos tratamentos dentários degradava ou não a qualidade de imagem;
- O posicionamento do informante - tendo em atenção que estaria imobilizado durante algumas horas, tentou-se posicionamento (com recurso a alguns apoios) em que ficasse o mais confortável possível, para que se prevenisse eventual rotação e/ou inclinação do pescoço e cabeça;
- Comunicação com o informante - testou-se também a forma como se comunicaria com o informante relativamente ao som que teria que produzir;
- Os informantes treinaram suste a produção, de forma artificial, durante o tempo de aquisição, para as várias sequências;
- Efectuou-se gravação do sinal através do intercomunicador, para cada uma das sequências testadas, ficando-se imediatamente com a ideia do que se poderia esperar em termos de qualidade;
- Planeou-se uma forma de codificar os vários elementos do corpus, que nos permitisse rapidamente gravar e identificar cada som, de forma a não existir qualquer tipo de dúvida na sua posterior identificação;
- Fizeram-se ensaios com diferentes sequências e optimizaram-se, tendo em atenção os requisitos previamente estabelecidos para cada uma das aquisições.

Corpus 2D (Testes)

Para aquisição do corpus 2D os principais objectivos eram; (1) obtenção de uma imagem médio-sagital, com boa resolução espacial, (2) boa S/R e (3) tempo de aquisição inferior a 10s (tendo como referência a espessura de corte e matrizes semelhantes às descritas em estudos semelhantes a que já se fez referência, com algum detalhe, no capítulo anterior). Assim, do menu de sequências frequentemente utilizadas para estudo da laringe, escolheu-se uma sequência TSE ponderada em T1, que foi optimizada para que o tempo de aquisição fosse o mais curto possível, mantendo os pressupostos de qualidade estabelecidos para esta aquisição (o TR para o número de cortes poderia ser mais baixo mas interferiria com o contraste na imagem; foi utilizado também

um ETL um pouco mais elevado do que o normal para ponderação T1). Foi ensaiada também uma sequência EG ponderada em T1 (FLASH 2D), mas optou-se por manter a TSE, dada a melhor qualidade de imagem.

Corpus 3D (Testes)

Para o corpus 3D e, uma vez que se pretendia obter modelação tridimensional do tracto vocal, efectuada, usualmente às custas de imagens 2D, ensaiou-se uma aquisição 3D. Os objectivos para esta aquisição eram os seguintes: (1) obtenção de um volume razoavelmente grande de informação (desde o palato até à glote) (2) idealmente com um tempo de aquisição abaixo dos 30s, e (3) uma SNR e resolução espacial que permitisse a obtenção da informação desejada (por exemplo, as áreas transversais do tracto vocal) através de segmentação. Assim, foram testadas as sequências 3D MP-RAGE e FLASH 3D VIBE, estabelecendo compromissos até obter o tempo de aquisição definido como meta. Foi escolhida a sequência FLASH 3D VIBE, uma vez que nos permitiu obter o tempo de aquisição mais curto. Obteve-se um tempo de aquisição de 18s o que nos parece muito razoável, tendo em atenção a extensão da área a incluir.

Podemos considerar que este método de aquisição será validado se for conseguida a obtenção de informação equivalente à obtida com a aquisição clássica (obtenção das funções de área do tracto vocal), com menor tempo de aquisição.

Corpus tempo real (testes)

A sequência a utilizar para este corpus foi de longe a mais difícil de otimizar. O objectivo estabelecido para este corpus era a obtenção de cinco frames/s com possibilidade de estudar os movimentos da língua e de abertura e fecho do velo. Existindo já estudos na bibliografia consultada que fazem referência a cerca de 9 frames/s, não parece um objectivo muito ambicioso. No entanto, no equipamento em que foi efectuada a aquisição das imagens, não foi possível efectuar modificações a nível da programação das sequências, o que nos impôs algumas condicionantes. Assim, foram testadas algumas das sequências disponíveis no equipamento e que geralmente se utilizam para estudo cardíaco em tempo real. Começou-se com 1 frame/s e optimizou-se a sequência até se obterem 5 frames/s. A sequência escolhida foi um real time TrueFISP (balanced -SSFP). Ainda foi testada a utilização de aquisição paralela, mas foi mal sucedida, uma vez que a antena de crânio que foi utilizada não é uma antena matricial, dando origem a um imenso artefacto de *aliasing*. Como a aquisição foi efectuada no plano sagital, seria necessário mais um canal no sentido Antero-Posterior, para se utilizarem com sucesso as técnicas PAT. A inexistência de sequência do tipo TSE Zoom ou técnicas não cartesianas de preenchimento do espaço K limitou a nossa escolha ao leque de sequências EG ultra-rápidas, com conseqüente aparecimento de artefactos de susceptibilidade magnética. As aquisições para as arcadas dentárias e tracto nasal do falante PAA não foram testadas, fazendo-se apenas uma discussão acerca das possíveis sequências a utilizar.

Toda a informação da sessão experimental foi gravada em CD, para posterior análise e escolha das sequências a utilizar. Todas as sequências testadas e optimizadas foram guardadas e devidamente identificadas num directório dentro do computador da sala de RM para posterior utilização.

Foi efectuada uma reunião de preparação para análise da informação recolhida na sessão experimental de que resultou um guião da recolha com os procedimentos a utilizar, a codificação de cada item dentro de cada sub-corpus, e as sequências a utilizar na sessão de aquisição de dados.

Foi definido o corpus final a adquirir, sendo possível planear a duração aproximada da sessão de recolha de dados. Alguns fones foram retirados da primeira versão do corpus. Tendo a noção do tempo de aquisição de cada uma das sequências a utilizar foi possível elucidar os falantes do tempo que teriam que manter a articulação, para cada um dos corpora 2D e 3D. O falante PAA teve acesso à descrição total do corpus no dia anterior à sua aquisição. A falante MTC só teve conhecimento do corpus total que iria adquirir um pouco antes do início da aquisição, mas foram-lhe explicados os procedimentos e teve a oportunidade de treinar a produção de alguns sons.

6.2.5 Aquisição Final

A sessão de aquisição de dados decorreu no dia 20 de Julho de 2006 e durou cerca de 3 horas. O primeiro conjunto de dados a ser adquirido foi o corpus em tempo real com a informante do sexo feminino (MTC), e que demorou cerca de 50 minutos. Durante este período, foram ainda adquiridas algumas imagens relativas a produções estáticas, utilizando a mesma sequência e procedimentos que descreveremos para a aquisição do corpus 2D, no informante PAA. A finalidade destas imagens será a identificação, por comparação, das *frames* relativas a cada fone obtidas durante a produção e adquirida em tempo real.

A informante recebeu algum treino prévio relativamente à forma como deveria sustentar a articulação para cada fone, durante a aquisição das produções estáticas, e foi informada que deveria repetir cada palavra do corpus em tempo real de forma e ritmo normais.

As imagens foram adquiridas com a informante deitada, em supinação, com apoio sob os joelhos para maior conforto e com a cabeça posicionada e imobilizada de forma a não existir rotação nem inclinação da mesma. Foram utilizadas as bobines de crânio e de pescoço de forma a prevenir perda de sinal nas extremidades do FOV. Foram colocados os auscultadores e a centragem foi efectuada na linha média, aproximadamente a meio da área abrangida pelas duas antenas. Após

identificação da informante, iniciou-se a aquisição das imagens. Através do intercomunicador, foi identificado cada item a repetir, da seguinte forma. MTC repita _____, por favor.

De um modo geral, a informante percebeu todas as instruções transmitidas pelo intercomunicador, só havendo necessidade de repetir a sequência das vogais nasais, por não se ter a certeza se a informante a teria pronunciado de forma correcta. Esta sequência foi repetida no final da aquisição e todos os outros elementos foram adquiridos pela ordem prevista (ver Tabelas 6.3 e 6.4) e à primeira tentativa.

Para recolha do corpus tempo real foi otimizada uma sequência eco gradiente rápida (Balanced SSFP - Siemens TrueFISP) do menu de sequências usualmente utilizadas para estudo cardíaco. Tendo como referência imagens no plano axial e coronal, foi marcado um corte sagital na linha média. Depois de indicar à informante a palavra a repetir, a sequência foi lançada quando, através do intercomunicador, se ouvia a primeira produção da falante. Foram efectuadas 100 medidas (measurements) a uma razão de cinco *frames* por segundo, o que resultou num tempo total de aquisição de 20 s. A resolução temporal conseguida foi da ordem dos 200 ms. Durante os 20 s de aquisição, a palavra foi repetida cerca de 10 a 12 vezes. De seguida, para cada item, e tendo como referência uma imagem sagital, foi marcado um corte com orientação paralela à do velo (coronal oblíquo), utilizando a mesma sequência e procedimentos agora descritos.

Na Tabela 6.5 indicam-se os principais parâmetros da sequência TrueFisp utilizada:

PARÂMETRO	VALOR
TR	3,56 ms
TE	1,5 ms
FA	60°
ETL	1
FoV read	240 mm
FoV phase	100%
Slice Thickness	5 mm
Orientation	Sagital ou Coronal oblíquo
Distance factor	0,1
Base resolution	128
Phase resolution	75%
Phase Partial Fourier	4/8
Bandwidth - BW	543 Hz/pixel
Acquisition Time	0,2 s
Measurements	100
Assymetric echo	Allowed
Gradient Mode	FAST

Tabela 6. 5 - Parâmetros de aquisição da sequência real time TrueFISP utilizada para obtenção do corpus Tempo real.

Cada item foi guardado, utilizando código previamente definido. De seguida, foram adquiridos os fones constantes na Tabela 6.3, cujo objectivo já foi explicado, com a sequência e procedimentos do corpus 2D que iremos descrever de seguida para o falante PAA.

Corpus 2D

Os fones do corpus 2D (num total de 64) foram adquiridos com o informante PAA, posicionado de forma semelhante ao descrito para a falante MTC, utilizando também as antenas de crânio e pescoço. Foi dada especial atenção ao conforto do falante, uma vez que iria proceder-se à aquisição do corpus 2D e 3D, na sua totalidade, neste falante sem reposicionamento. A sequência utilizada e previamente otimizada foi uma sequência TSE ponderada em T1 com os parâmetros que constam da Tabela 6.6. Para cada fone passível de ser sustido, a articulação foi sustida de forma artificial durante os 5,6 s do tempo de aquisição, após ter sido comunicado pelo intercomunicador a palavra de referência para a produção do fone. Por exemplo, diga [a] como na palavra pato [patu].

No caso das oclusivas e fricativas, que foram obtidas em contexto (VCV) com as vogais [i a u], e após indicação dos sons que se seguiriam (por exemplo: vamos iniciar a produção da oclusiva [p] em contexto), o falante fazia a transição VC (vogal - consoante) mantinha a oclusão (nas oclusivas) ou sustinha a produção (nas fricativas) e só nesse momento era lançada a sequência. Depois de concluída a aquisição (quando o informante deixava de ouvir o barulho da sequência) fazia a transição CV. Todo este processo foi monitorizado através do intercomunicador e pelo próprio falante que, quando achou que a produção não tinha sido bem sucedida, nos alertava. Só houve necessidade de repetir um dos sons do corpus. Além do inicialmente previsto, foi adquirido um item adicional, o [l] em final de palavra (velarizado) como na palavra *mal*, para comparação com [l] em início de palavra (*laço*). A aquisição deste corpus demorou 30 minutos, ficando as imagens totalmente codificadas e guardadas.

PARÂMETRO	VALOR
TR	400 ms
TE	8,3 ms
FA	180°
ETL	15
FoV read	200
FoV phase	100%
Slice Thickness	5 mm
Orientation	Sagital
Distance factor	0,1
Base resolution	256
Phase resolution	75%
Phase Partial Fourier	-
Bandwidth - BW	235 Hz/pixel
Acquisition Time	5,6 s
NEX	1
Pixel Spacing	0,7813

Tabela 6.6 - Parâmetros de aquisição da sequência TSE ponderada em T1, utilizada para a aquisição do corpus 2D.

Corpus 3D

Depois da aquisição do corpus 2D, foi adquirido o corpus 3D propriamente dito, utilizando uma sequência 3DFT Eco de gradiente *Spoiled* rápida (FLASH 3D VIBE), que já foi referida mais atrás, e que constituiu uma inovação neste tipo de estudos. A sequência foi marcada no plano transversal desde o nível da glote até um nível ligeiramente acima do palato duro, de forma a permitir eliminar os primeiros e últimos cortes, geralmente degradados por artefactos de *aliasing* característicos da

aquisição 3D. O tempo de aquisição foi de 18s por fone, tempo significativamente inferior ao da aquisição clássica (53 s). Foram adquiridos os fones previstos, à exceção de [i~] na palavra *minto* e [L] em *ralha*. O falante sustentou a produção durante o tempo de aquisição. A aquisição do corpus demorou 50 minutos. Os parâmetros utilizados na sequência são os que se podem observar na Tabela 6.7.

PARÂMETRO	VALOR
TR	4,89 ms
TE	2,44 ms
FA	10°
ETL	1
FoV read	270 mm
FoV phase	81,3%
Slabs	1
Slices per slab	64
Thickness	2 mm
Distance factor	0,20
Base resolution	256
Phase resolution	60%
Phase direction	R-L
Phase oversampling	-
Phase Partial Fourier	6/8
Bandwidth - BW	350Hz/pixel
Acquisition Time	18 s
NEX	1
Pixel Spacing	0,5273

Tabela 6. 7 - Parâmetros de aquisição da sequência 3DFLASH VIBE (axial) utilizada para adquirir o corpus 3D.

Corpus 3D (aquisição clássica)

Depois da aquisição em volume, foram adquiridos apenas alguns fones do corpus 3D, referidos no ponto 6.1.2, utilizando um protocolo semelhante ao referenciado na literatura (Demolin, Metens & Soquet, 1996; Badin, 1998; Engwall & Badin, 1999). Foi utilizada uma sequência TSE com ponderação DP. Foram efectuados um total de 45 cortes distribuídos por 3 ranges, com diferente orientação. Assim, foram marcados sobre uma imagem sagital, 15 cortes axiais, desde a glote até ao nível de C2, 15 cortes oblíquos com a orientação do velo (T/C 40,6°), e 15 cortes coronais, marcados desde os lábios até à extremidade do palato duro. O tempo de aquisição para cada item foi de 53 segundos. O procedimento de comunicação com o falante foi o mesmo que utilizado para

o corpus 2D. O falante foi informado do tempo que teria que sustentar a produção, tentando que fosse o mais estável possível. Dado o tempo de aquisição extremamente longo, o falante sustentou a produção enquanto respirava suavemente. Foram adquiridos 10 fonemas que demorou cerca de 30 minutos.

PARÂMETRO	VALOR
TR	7360 ms
TE	7,3 ms
FA	150°
ETL	33
FoV read	250
FoV phase	85%
Slice Thickness	5 mm
Orientation	Axial Coronal T/C 40,6°
Distance factor	0,3
Base resolution	256
Phase resolution	83%
Phase Partial Fourier	5/8
Bandwidth - BW	270 Hz/pixel
Acquisition Time	53 s
NEX	1
Pixel Spacing	0,4883

Tabela 6. 8 - Parâmetros de aquisição da sequência DP TSE para obtenção do corpus 3D pelo método clássico.

De seguida, e para terminar as aquisições com o falante em posição de supinação, foi realizada uma aquisição 3D de alta resolução a abranger toda a região dos tractos oral e nasal. Durante a aquisição, no plano sagital, o falante manteve o tracto vocal em posição de repouso. Foi seleccionada uma sequência IR, com preparação da magnetização (3D MP-RAGE) ponderada em T1, usualmente utilizada em estudos crânio-encefálicos. Os parâmetros apresentam-se na Tabela 6.9.

PARÂMETRO	VALOR
TR	1900 ms
TE	3.94 ms
TI	1100 ms
FA	15°
FoV read	250 mm
FoV phase	75%
Slice Thickness (eff)	1,5 mm
Orientation	Sagital
Distance factor	-
Base resolution	256
Phase resolution	75%
Phase Partial Fourier	-
Bandwidth - BW	130Hz/pixel
Acquisition Time	8 minutos
NEX	2
Pixel Spacing	0,9766

Tabela 6. 9 - Parâmetros de aquisição da sequência 3D MPRAGE, utilizada para obtenção do tracto oral e nasal em repouso do falante PAA.

Por fim, o falante foi levantado e reposicionado para efectuar uma aquisição para obtenção das suas arcadas dentárias para posterior co-registo com as restantes imagens adquiridas nos corpus 2D e 3D.

As técnicas para obtenção das arcadas dentárias dos falantes têm sido diversas, uma vez que, como sabemos, os dentes são estruturas calcificadas, que surgem nas imagens RM com vazios de sinal, em qualquer das ponderações utilizadas. Nos estudos consultados, foram descritas várias técnicas para obter e co-registar essa informação. O mais comum é obter essa informação mediante a utilização de TC (Serrurier & Badin, 2005ab), EBCT (Story, 1995; Story, Titze & Hoffman, 1996; Tom, et al., 1999) ou ainda com RM, efectuando-se, por exemplo, moldes dentários dos falantes, que são imersos em água sendo posteriormente efectuada uma RM do molde (Narayanan, Alwan, Haker, & 1995, 1997, Badin et al., 1998; Engwall & Badin, 1999). Também surge referida (Takemoto, Kitamura, Nishimoto, & Honda, (2004) a utilização de sumo de mirtilos dentro da boca, durante uma aquisição 3D com ponderação T1, com o falante em pronação. A utilização de sumos de fruta, nomeadamente o sumo de mirtilos, tem sido referido como tendo potencial de utilização como meio de contraste oral, em estudos do tracto abdominal superior. A possibilidade de utilização como contraste oral positivo e negativo tem sido estudada

(devido ao seu teor de Mn e Fe), no entanto, segundo alguns autores a utilização do sumo de mirtilos é mais eficiente recorrendo a sequências ponderadas em T1, condicionando efeito de realce positivo (Karantanas, Papanikolaou, Kalef-Ezra, Challa et al, 2000).

Ericsson (2005) utilizou um tubo fino preenchido com gadolínio e que foi depois colado às arcadas dentárias. No nosso estudo, apesar de não ser absolutamente fundamental a aquisição das arcadas dentárias, optou-se por (dada a oportunidade) fazer também uma aquisição das arcadas dentárias de um dos falantes (PAA), para posterior co-registo.

Excluindo à partida a hipótese de utilizar TC, uma vez que envolveria a utilização de radiação ionizante, e de não ter sido possível obter mirtilos para sumo, efectuou-se uma sequência TSE 3D (de alta resolução), com ponderação T2 e FA variável, ficando o falante com a boca preenchida com água, durante a aquisição. Como se sabe, os líquidos apresentam hipersinal numa ponderação T2, pelo que será possível individualizar as arcadas dentárias do falante e facilitar a segmentação.

Desta maneira, a aquisição foi efectuada recorrendo apenas à antena de crânio, e com o falante numa posição de pronação, com a cabeça em hiperextensão ligeira, para conseguir manter, durante os três minutos da aquisição, a água dentro cavidade oral. Depois de posicionado, preencheu a boca com água com ajuda de uma palhinha. Na Tabela 6.10 apresentam-se os parâmetros da sequência utilizada.

PARAMETRO	VALOR
TR	2500 ms
TE	355 ms
FA	180°
ETL	193
FoV read	260 mm
FoV phase	100%
Thickness (ef)	3mm
Orientation	Sagital
Base resolution	386
Phase resolution	100%
Phase Partial Fourier	-
Bandwidth - BW	850 Hz/pixel
Acquisition Time	3 minutos
NEX	1
Pixel Spacing	0,5469

Tabela 6. 10 - Parâmetros de aquisição da sequência 3D TSE, ponderada em T2 , para obtenção das arcadas dentárias do falante PAA.

6.2.6 Gravação do sinal acústico durante e após a aquisição das imagens

A gravação do sinal acústico, antes ou depois da RM, e mesmo durante a realização de RM, é referida com frequência na bibliografia consultada, servindo diferentes objectivos, em função do estudo realizado. Por vezes, é efectuada uma gravação de alta qualidade, fora da sala do magneto, antes ou após a RM, com o falante a produzir configuração estática igual à que produziu durante a RM (Story, 1995; Story, Titze, & Hoffman, 1996). A finalidade destas gravações é comparar a fala natural do indivíduo com as simulações conseguidas a partir das áreas calculadas. As gravações são efectuadas, tentando recriar as condições da RM, nomeadamente com o falante a usar auscultadores e em posição de supinação. Noutras situações, a gravação é efectuada durante a aquisição das imagens sendo o principal objectivo a obtenção do sinal de voz para efectuar sincronização com as imagens obtidas (Demolin, Metens, & Soquet, 2000; Mohammad, 1999).

No entanto, na opinião de alguns autores, quando se pretende efectuar sincronização, a obtenção de sinal acústico antes ou após a aquisição das imagens não serve os objectivos, pois dificilmente um ser humano consegue repetir exactamente da mesma maneira um determinada produção (NessAiver et al., 2006), devendo a mesma ser efectuada durante a aquisição de imagem. A gravação durante a aquisição de imagens levanta alguns problemas técnicos, uma vez que o índice de ruído, dentro da sala do magneto, durante a aquisição das sequências é significativo e, dada a presença de um campo magnético, nem todo o material poderá entrar de forma segura¹⁶ dentro da sala ou ser compatível (não causar artefactos nas imagens).

Várias têm sido as técnicas e estratégias utilizadas para efectuar as gravações durante a aquisição de imagens:

- (1) Colocação do microfone dentro da sala do magneto (Baer, Gore, Gracco, & Nye, 1991), mas no caso referenciado, foi utilizado um magneto de baixo campo (0,15T) e sequência Eco de *Spin* (menos ruído quando comparada com as sequências eco de gradiente). Mohammad (1999) também começou por utilizar microfone dentro da sala, mas abandonou esta técnica uma vez que o microfone causou interferência (50 Hz) com consequentes artefactos nas imagens (neste estudo foi utilizado um magneto de 0,5 T);
- (2) Gravação do som proveniente do intercomunicador (Demolin, Metens, & Soquet, 2000; Mohammad, 1999), que apesar de não permitir a obtenção de sinais áudio de qualidade, permite a identificação e segmentação do sinal de voz, com algumas sequências;

¹⁶ Material seguro (*MR Safe*) é aquele que pode entrar dentro da sala do magnete sem colocar em risco a segurança do utente ou do *staff*, podendo no entanto dar origem a artefactos. Material compatível (*MR compatible*) é aquele que é seguro e ao mesmo tempo não causa artefactos que degradem a qualidade de imagem (Shellock, 2002).

- (3) Recurso a microfone óptico para gravações de alta qualidade durante a aquisição de imagens RM (Ericsson, 2005; NessAiver, Stone, Parthasarathy, Kahana, & Paritsky, 2006). Com esta tecnologia é possível a gravação de som de elevada qualidade (ausência de componentes metálicos e cancelamento de ruído) mesmo durante a aquisição de sequências ultra-rápidas, e, em consequência, altamente ruidosas;

Neste estudo, e na impossibilidade de se efectuar gravação com microfone óptico, efectuou-se a gravação do sinal de voz durante a aquisição das imagens (ainda que com qualidade inferior) utilizando o intercomunicador de comunicação com a sala do magneto. Desta forma, foi possível gravar toda a comunicação estabelecida com os informantes durante a aquisição das imagens, monitorizar as produções e efectuar a segmentação do sinal áudio para os elementos do corpus 2D e 3D. Do corpus obtido em tempo real, não se conseguiu extrair qualquer informação áudio, uma vez que os índices de ruído com as sequências utilizadas foram muito elevados.

As gravações foram efectuadas recorrendo ao programa SFS (Speech Filing System) versão 4.6 e sistema Kay CSL 4400, com microfone unidireccional colocado junto ao intercomunicador da sala de RM. Para efectuar a segmentação utilizou-se, também, o programa SFS. Foram aplicados filtros para minimizar o ruído.

Apesar de estar prevista a gravação do sinal de voz para o falante PAA, fora do ambiente ruidoso da sala de RM, mas recriando as condições da aquisição, esta ainda não foi efectuada.

6.3 Processamento de imagem

6.3.1 Enquadramento

A segmentação, e em particular a segmentação de imagens médicas, é um problema que tem sido amplamente estudado (McInerney & Terzopoulos, 1996), mas para o qual ainda não se encontrou uma solução ideal. Existem várias tentativas de implementação de algoritmos semi-automáticos para a detecção de contornos (que seja do nosso conhecimento não existe ainda nenhum algoritmo de segmentação completamente automático). Em muitos casos, continua a ser preferível a utilização de métodos manuais, devido à difícil utilização e resultados pouco satisfatórios obtidos pela utilização dos métodos semi-automáticos desenvolvidos até ao momento.

Para que os parâmetros extraídos das imagens de Ressonância Magnética representem, de forma fidedigna, as imagens recolhidas, é necessário não só que o processo de segmentação exista, mas também que seja o mais exacto possível, sem que se perca demasiado tempo neste procedimento (o que pode até inviabilizar o estudo).

Nos estudos consultados, a segmentação é, na maior parte das vezes efectuada de forma totalmente manual (Arnal et al., 2000; Badin et al., 2000; Demolin, Delvaux, Metens, & Soquet, 2003; Demolin, Metens, & Soquet, 1996; Engwall, 2000a, 2001, 2003a; Greenwood, Goodyear, & Martin, 1992; Mohammad, 1999; Serrurier & Badin, 2005a; Shadle, Mohammad, Carter, & Jackson, 1999; Soquet, Lecuit, Metens, & Demolin, 2002; Yang, 1999) ou recorrendo a métodos semi-automáticos. Neste último caso, as técnicas mais utilizadas são o *thresholding* (onde se seleccionam automaticamente pontos que correspondem a um nível de cinzento pré determinado, sendo depois manualmente seleccionados de forma a constituir o contorno). Este método foi utilizado nos estudos (Apostol, Perrier, Raybaudi, & Segebarth, 1999; Badin et al., 1998; Engwall, 2004; Engwall & Badin, 1999; Kröger, Winkler, Mooshammer, & Pompino-Marschall, 2000) - e métodos que utilizam contornos activos (nestes métodos é escolhido manualmente e de forma pouco exacta um contorno inicial que, de modo automático e através da minimização de funções de energia baseadas em características escolhidas da imagem, se vai adaptando ao contorno pretendido, (Davison, Eviatar, & Somorjai, 2000) - utilizado no estudo (Takemoto, Honda, Masaki, Shimada, & Fujimoto, 2003). Encontrou-se ainda um artigo (Narayanan, Nayak, Lee, & Byrd, 2004) que refere a utilização de segmentação automática. No entanto, os métodos utilizados não são expressamente descritos, sendo referida apenas a utilização de *Kalman Snakes* e *Optical flow*.

6.3.2 Métodos utilizados neste estudo

Dado que a base de dados adquirida neste estudo é muito extensa, eliminou-se a hipótese de efectuar segmentação manual, como rotina. Este tipo de segmentação, embora considerada por alguns autores como sendo das mais fidedignas, apresenta uma grande variabilidade inter e intra sujeitos, sendo também um processo demorado (Silva, 2005).

Assim sendo, optou-se pela utilização de métodos semi-automáticos, largamente difundidos em outras áreas, mas usualmente pouco utilizados para a segmentação do tracto vocal. Estes métodos, comparativamente, apresentam como principal vantagem a menor dependência da intervenção do utilizador do que os métodos manuais.

As imagens 2D foram segmentadas com um método de segmentação denominado *Region Growing* (Adams & Bischof, 1994), e as imagens do corpus 3D com um outro método, o *Live Wire* (Mortensen, Morse, & Barrett, 1992). A escolha de métodos diferentes prendeu-se com a qualidade da imagem. O *Region Growing*, tem a vantagem relativamente ao *Live Wire* de só necessitar, como intervenções do utilizador, a escolha de uma semente e a selecção de um valor de limiar. No entanto, esta técnica é menos robusta ao ruído presente na imagem e foi, por isso, utilizada com sucesso nas imagens 2D (que apresentam uma boa SNR), mas não funcionou tão

bem nas imagens em 3D. Para estas, foi escolhida a técnica *Live Wire* como referido, que embora necessite de mais intervenção do utilizador, conduz a resultados mais satisfatórios nesta situação.

A extracção de informação das imagens em Tempo Real teve uma aproximação bastante diferente. Dada a dificuldade acrescida de ter que utilizar métodos de segmentação que conseguissem colmatar a falta de qualidade (fraca resolução espacial) em cada *frame*, através da utilização de informação das *frames* vizinhas, optou-se por pôr de parte (pelo momento) a segmentação destas imagens. Nas imagens em tempo real, apenas foram marcados, de forma manual, alguns pontos (ponta da língua, dorso e raiz da língua) e, numa fase posterior aplicou-se também a técnica *Live Wire* para extracção do contorno da língua.

De seguida, descrevem-se sucintamente, cada um dos métodos utilizados.

Region Growing

O primeiro passo para a implementação desta técnica é a escolha de um ponto que se denomina a semente. O ponto pode ser escolhido arbitrariamente, desde que seja um de entre os que se encontram no interior da região a delimitar (ver o ponto a amarelo na Figura 6.1).



Figura 6. 1 - Representação da semente numa imagem de RM no plano sagital (falante PAA - corpus 2D).

- Esta semente será o primeiro ponto pertencente à área no interior da região a delimitar;
- De seguida, testa-se cada um dos 8 vizinhos da semente. Na Figura 6. 2, a amarelo está representada a semente e a vermelho os seus 8 vizinhos.

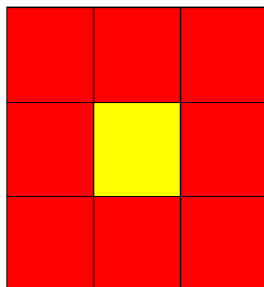


Figura 6. 2 - Noção de vizinhos. Semente a amarelo e vizinhos representados a vermelho.

- O valor de cinzento de cada um dos vizinhos é comparado com o valor de cinzento da semente;
- Se a diferença estiver abaixo de um certo limiar, o *pixel* é classificado como pertencente ao interior da região a delimitar. Caso a diferença seja maior do que o limiar, o *pixel* é classificado como pertencente ao contorno da região a delimitar;
- De seguida, actualiza-se o valor de cinzento com que cada *pixel* será comparado. Este valor é uma média dos valores de cinzento de todos os *pixels* classificados como pertencentes ao interior da região;
- Os vizinhos de cada um dos pontos pertencentes ao interior da região e que ainda não tenham sido considerados, são comparados com a média dos valores de cinzento do interior da região e são classificados de forma semelhante ao descrito acima, ou seja, se a diferença for menor que o limiar eles são acrescentados ao conjunto de pontos do interior da região, caso contrário pertencem ao contorno;
- O algoritmo continua até que todos os pontos pertencentes ao interior da região sejam classificados.

Este algoritmo foi implementado em *Matlab* numa função denominada *RegionGrowing.m*, que tem como parâmetros de entrada a imagem a segmentar, a semente e o valor do limiar. Como parâmetros de saída tem os pontos pertencentes ao contorno e os pontos pertencentes ao interior da região. Procedeu-se ainda a uma avaliação quantitativa do efeito da localização da semente no contorno final. Desta avaliação, concluiu-se que o método é robusto a mudanças na localização da semente (Carbone, 2007).

A aplicação deste algoritmo à imagem e semente representada na figura 6.1, para um valor de limiar igual a 0.09, conduz ao resultado da Figura 6.3.

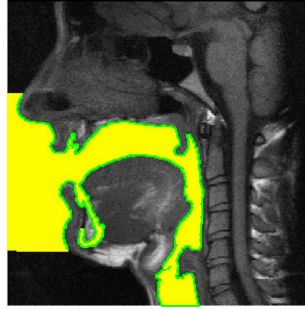


Figura 6.3 - Contorno criado pela técnica *Region Growing*.

Obtenção das distâncias médio-sagitais em 2D

Um dos dados usualmente utilizados para modelação nos sintetizadores de fala (articulatórios 2D) são as distâncias médio-sagitais. A partir destas distâncias, é usual inferir a área da secção a incluir nos modelos dos sintetizadores. Estas distâncias, no nosso estudo, foram obtidas através de uma grelha adaptativa. O processo de criação desta grelha é descrito de seguida:

- 1) Criação do contorno do tracto vocal (*Region Growing*, neste caso);
- 2) Sobreposição de uma grelha fixa ao contorno criado;
- 3) Intersecção das rectas da grelha fixa com o contorno;
- 4) Cálculo do ponto médio entre as intersecções criadas (para cada linha da grelha fixa);
- 5) Criação de um vector com os pontos acima criados e respectiva suavização (linha central).

A linha central está representada a vermelho na Figura 6.4, à esquerda. A grelha adaptativa corresponde ao conjunto das linhas perpendiculares à linha central. A grelha pode conter um número de linhas variável que o operador pode seleccionar. Neste estudo, utilizou-se, geralmente, uma grelha adaptativa constituída por 45 linhas ou cortes.

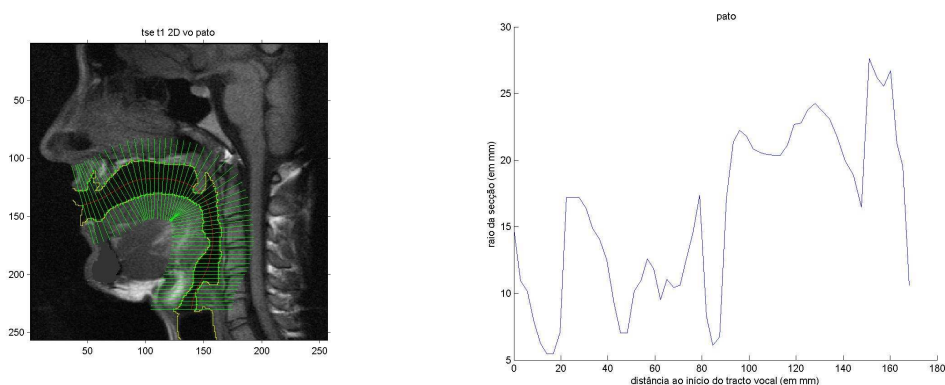


Figura 6.4 - Obtenção das distâncias médio sagitais no corpus 2D: esquerda - contorno (amarelo), linha central (vermelho) e grelha adaptativa (verde); direita - gráfico das distâncias médio sagitais.

As distâncias médio-sagitais correspondem à distância entre as intersecções de cada linha da grelha adaptativa e o contorno criado (Figura 6.4, à direita). No eixo do xx, representa-se o tamanho do tracto vocal. Geralmente, considera-se o início do tracto vocal a nível da glote, terminando nos lábios.

Marcação de pontos para extracção de parâmetros articulatórios (2D)

Para obtenção de alguns parâmetros articulatórios, importantes para o processo de síntese e para a descrição articulatória dos vários fones, marcaram-se em todas as imagens 2D, alguns pontos: TT (ponta da língua), TD (ponto mais alto do dorso da língua), TR (posição da raiz da língua ao nível do espaço C3-C4), JH (altura da mandíbula), UL (lábio superior) e LL (lábio inferior). Podemos observar a marcação dos pontos, na imagem correspondente à produção do fone [6], para o falante PAA (Figura 6.5).

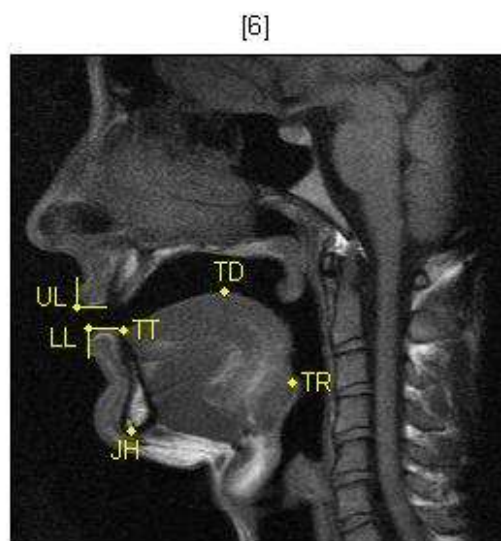


Figura 6.5 – Marcação de pontos referentes a UL, LL, TT, TD, TR e JH, para o falante PAA durante a produção da vogal oral [6].

Live Wire

A técnica de *Live Wire* é uma técnica em que a segmentação é feita passo-a-passo. Desta forma, é uma segmentação feita em “tempo real”. Esta técnica pertence à classe dos contornos activos denominados por *snakes*.

O algoritmo começa por calcular todos os termos de custo¹⁷ local, excepto para a direcção do gradiente, e o utilizador escolhe uma semente. Contrariamente ao *Region Growing*, onde a semente pertencia ao interior da região de interesse, neste caso a semente tem que pertencer ao contorno. De seguida, é calculado o custo para chegar da semente a cada ponto da imagem, através da determinação do caminho que leva ao custo mínimo. A filosofia em que se baseia este algoritmo é que, o caminho de menor custo é provavelmente coincidente com o contorno da região a segmentar, porque as funções de custo serão desenhadas para que às arestas correspondam custos baixos. O utilizador, através do rato, consegue seleccionar diversos pontos na imagem e observar o caminho produzido pelo algoritmo. Desta forma, o utilizador selecciona um novo ponto na imagem, de modo a que o caminho produzido desde a semente até ao novo ponto, seja do seu agrado, ver Figura 6.6. O novo ponto é considerado como uma semente e o processo repete-se até que o utilizador decida terminar a segmentação.

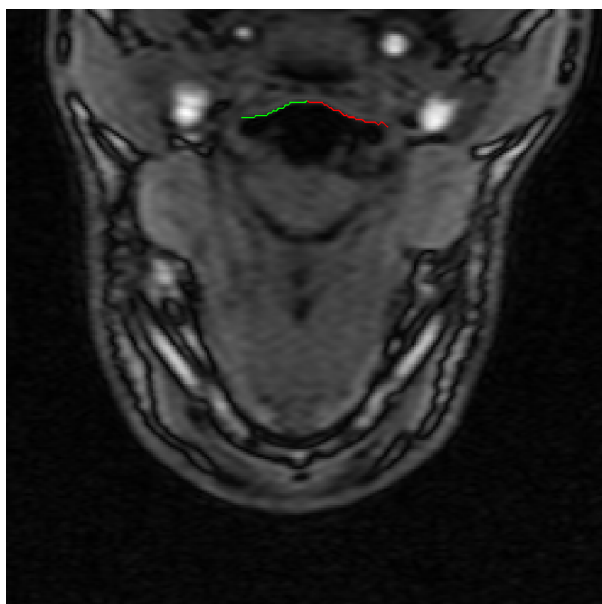


Figura 6. 6 - Exemplo de utilização do algoritmo *Live Wire*: a verde - linha já escolhida pelo utilizador e a vermelho - linha proposta pelo algoritmo.

Esta rotina não foi implementada pela nossa equipa, mas desenvolvida por Chodorowski et al., (2005). É constituída por um ficheiro encriptado, `simplelw.dll`, e um ficheiro do Matlab, `lwcontour.m`.

Após a explicação da técnica utilizada na segmentação, vamos proceder de forma resumida, à explicação de como ela foi aplicada para a segmentação dos volumes obtidos com a RM. Começámos por cortar o volume, de modo a reconstruir uma imagem sagital de referência (no

¹⁷ Neste caso, a função de custo é uma função que depende dos vértices e arestas orientadas (Falcão et al., 1998).

plano médio). O tracto vocal foi segmentado nesta imagem, através do *Live Wire*, procedendo-se depois ao cálculo da sua linha central. Posteriormente, foram calculados 45 planos perpendiculares à linha central do tracto vocal (Figura 6.7 - à esquerda), sendo depois o volume cortado segundo esses 45 planos. Cada um desses planos foi novamente segmentado, recorrendo ao *Live Wire*. A segmentação final observa-se na figura 6.7 (à direita).

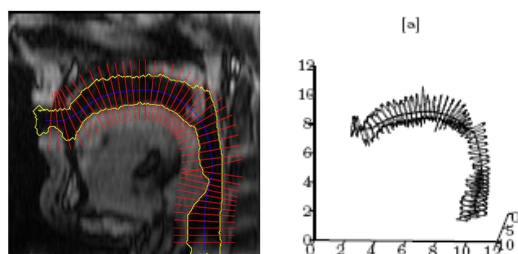


Figura 6.7 - Esquerda: Imagem no plano sagital, respectiva segmentação (amarelo), linha central (azul) e planos perpendiculares (vermelho); Direita: Resultado da segmentação.

6.3.3 Software

O *software* utilizado neste estudo foi, essencialmente, o *Matlab*. Esta ferramenta é bastante versátil e muito difundida pela comunidade científica - foi utilizado em (Apostol, Perrier, Raybaudi, & Segebarth, 1999; Engwall, 2003a, 2004; Engwall & Badin, 1999; Pritchard, 2002) - e existem por isso rotinas de processamento de Imagem já criadas tendo por base esta linguagem e que nos foram bastante úteis. Para o efeito foram criados e implementados alguns códigos (Carbone, 2007).

Uma outra ferramenta utilizada foi o ITKSnap (Yushkevich et al., 2006). Com esta ferramenta foram efectuadas algumas experiências de segmentação e visualização do *Corpus* 3D com a aquisição em volume. No entanto, não foi possível exportar os resultados deste programa para outros programas, de modo a poderem ser realizadas algumas tarefas de pós-processamento (tal como a medição de áreas) e efectuar comparações com outro tipo de dados.

Ambas as ferramentas foram utilizadas no Pré-processamento, Segmentação, Visualização e Pós-processamento das imagens de RM.

6.3.4 Tratamento de dados

Em termos de tratamento de dados, recorreu-se apenas ao cálculo de médias, mediana e desvio-padrão quando adequado (determinação do Coeficiente de Abertura do Velo e dimensão da cavidade faríngea) e à utilização de uma medida conhecida como Índice de Pratt (IP).

Índice de Pratt

Permite efectuar comparações entre contornos, o IP está contido no intervalo pertencente a $[0, 1]$, sendo que dois contornos são iguais se o seu IP for igual a um. Apresenta-se a fórmula de cálculo do IP (equação 6.1). Na fórmula, N representa o número de pontos correspondentes entre os contornos; αd_i^2 representa a distância entre dois pontos e, α está relacionado com a dimensão do contorno a avaliar.

$$PRATT = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + \alpha d_i^2} \quad (6.1)$$

6.4 Sumário

Neste capítulo descreveu-se o corpus, o protocolo de recolha de dados, todos os procedimentos utilizados nas diferentes fases que envolveram o planeamento e execução deste trabalho e a apresentação das técnicas de segmentação utilizadas. Sempre que se considerou adequado ou foi necessário, fez-se referência a técnicas ou estratégias utilizadas por outros autores.

Capítulo 7

Resultados

Neste capítulo, faz-se a apresentação e discussão de alguns dos resultados possíveis de serem obtidos com a base de dados adquirida neste estudo. O capítulo divide-se em 4 secções. Na secção 7.1, depois da avaliação da qualidade de imagem, apresentam-se os resultados (imagens) e análises efectuadas relativas aos dados obtidos com o Corpus 2D, na secção 7.2 são apresentados os resultados (ainda 2D) relativos ao estudo das fricativas e oclusivas em contexto. Na secção 7.3, referente ao corpus 3D, apresenta-se a comparação entre os dois tipos de aquisição efectuada (clássica *versus* volume) e apresentam-se os resultados obtidos relativamente a uma parte dos sons constituintes do corpus 3D (em particular funções de área do tracto vocal para as vogais cardinais do PE, laterais e fricativas em contexto). Para as vogais (orais e nasais) e consoantes nasais apresentam-se ainda, as áreas da abertura nasal, oral e faríngea e o Coeficiente de Abertura do Velo (CAV). Por último, na secção 7.4 são apresentados os primeiros resultados do corpus obtido em tempo real, apontando algumas possibilidades para a obtenção de informação relativa à dinâmica dos articuladores.

7.1 Corpus 2D

Para cada um dos fones do corpus 2D, obtiveram-se imagens no plano sagital, referentes a produções sustentadas artificialmente. Com estas imagens pretendia-se obter informação bidimensional relativa ao tracto vocal. Os resultados apresentados nesta secção dizem respeito ao falante PAA e resumem-se a uma descrição articulatória simples, baseada na observação das imagens, dos seus contornos e sobreposição dos mesmos para as vogais orais [i e E @ 6 a O o u], nasais [i~ e~ 6~ o~ u~] oclusivas surdas [p t k] e sonoras [b d g], consoantes nasais [m n J], fricativas [f s S v z Z] e laterais [l L], do PE. Quando adequado, será efectuada também uma apreciação mais quantitativa baseada na aplicação do Índice de Pratt (IP) e nos valores de alguns parâmetros articulatórios obtidos.

No que diz respeito à qualidade das imagens, pensamos que os objectivos propostos para o corpus 2D foram na sua globalidade atingidos, permitindo obter os parâmetros (traçados dos perfis médio-sagitais, distâncias médio-sagitais, parâmetros articulatórios) previamente definidos como metas para este estudo e com potencial para obtenção de outros que se possam vir a definir. Do ponto de vista da qualidade de imagem, enquadrada nos objectivos inicialmente definidos, podemos referir que: (1) o FOV escolhido revelou-se adequado, na medida em que permitiu a

visualização de todo o tracto vocal (desde a glote aos lábios sem que se observassem artefactos de *aliasing* sobre as imagens) e permitiu a visualização de estruturas que poderão servir como pontos de referência para a obtenção de medidas adicionais (hióide, coluna cervical de C1 até à transição C6-C7, sela turca, *nasion*, como exemplos); (2) a ponderação escolhida foi adequada, uma vez que as imagens ponderadas em T1 permitiram a obtenção de uma boa discriminação anatómica com contraste adequado entre as estruturas, como de resto se pode verificar nas imagens obtidas (Figuras 7.1 a 7.3 por exemplo); (3) todos os articuladores, à excepção dos dentes, que ainda não estão co-registados nestas imagens, são observados com boa definição (mandíbula, lábios, língua, palato duro e velo).

O compromisso entre resolução espacial e S/R permitiu efectuar segmentação semi-automática sem necessidade de aplicação de filtros sobre as imagens, para minimizar o ruído. A utilização das duas antenas em simultâneo (crânio e pescoço) contribui também, por certo, para a obtenção de uma imagem de melhor qualidade, com sinal uniforme ao longo de todo o campo de visão e facilitando a segmentação.

Em termos de tempo de aquisição, este deverá ser o mais curto possível, para que o falante tenha que manter a articulação o menor tempo possível, já que não é, de todo, uma situação natural em termos de produção de fala. O tempo de aquisição foi de 5.6 s, sendo dos mais baixos relativamente aos estudos que consultámos e que já referimos ao longo da dissertação. É certo que só foi efectuado um corte, mas esta foi precisamente a escolha para este corpus (uma só imagem com boa qualidade), uma vez que a informação 3D não será extraída a partir destas imagens. A utilização de uma sequência EG foi pensada (FLASH T1, por exemplo) e poderia minimizar o tempo de aquisição, no entanto o aumento de sensibilidade a diferenças de susceptibilidade magnética poderia comprometer a qualidade da imagem, dado o material dos tratamentos dentários do falante. Nas imagens não são perceptíveis artefactos de movimento, ainda que com o falante em fonação.

7.1.1 Vogais orais

Começamos por apresentar as imagens e os contornos correspondentes às nove vogais orais do PE (Figuras 7.1 a 7.3). Na figura 1 apresentam-se as vogais anteriores, na Figura 2 as médias ou centrais e na figura 7.3 as vogais orais posteriores do PE. Para cada figura, apresenta-se do lado esquerdo, imagem no plano sagital e do lado direito a imagem com o contorno do tracto vocal após segmentação.

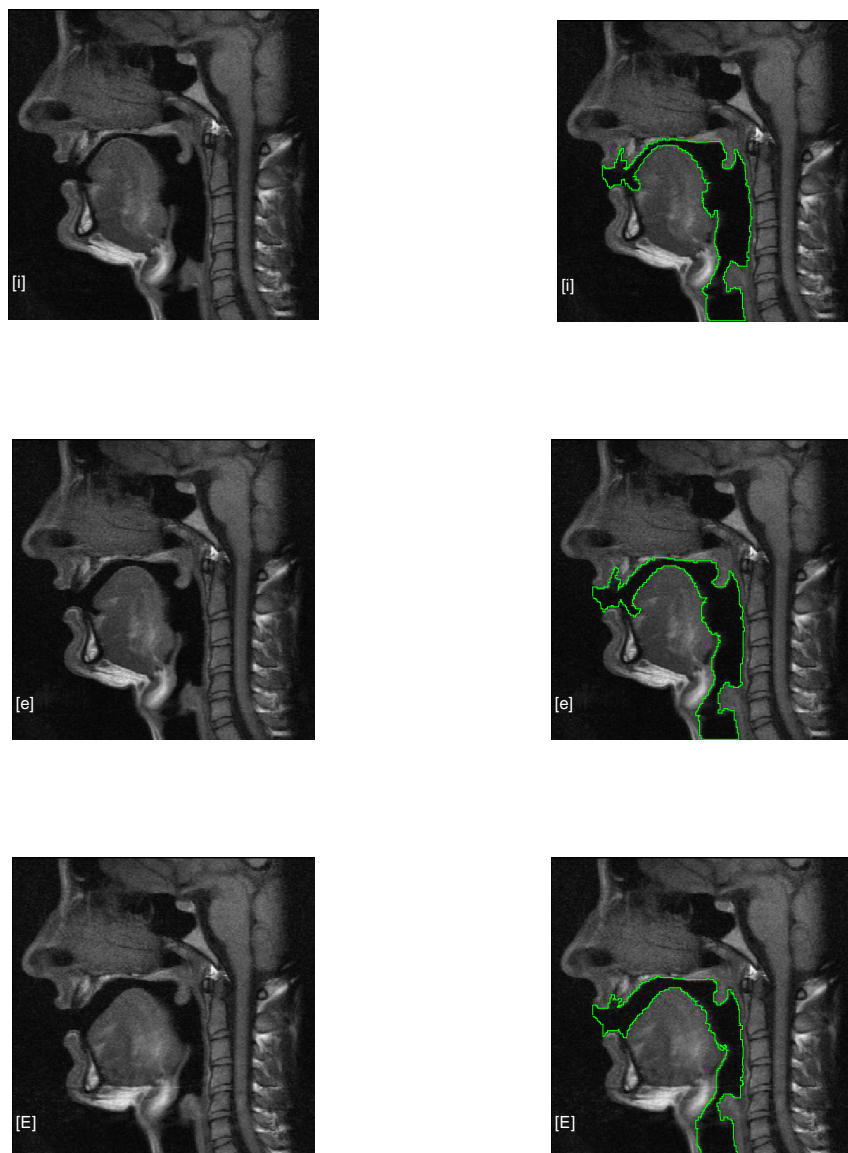


Figura 7. 1 - Vogais orais anteriores do PE [i] em cima, [e] ao centro e [E] em baixo, produzidas pelo falante PAA (imagem no plano sagital à esquerda, com o respectivo contorno à direita).

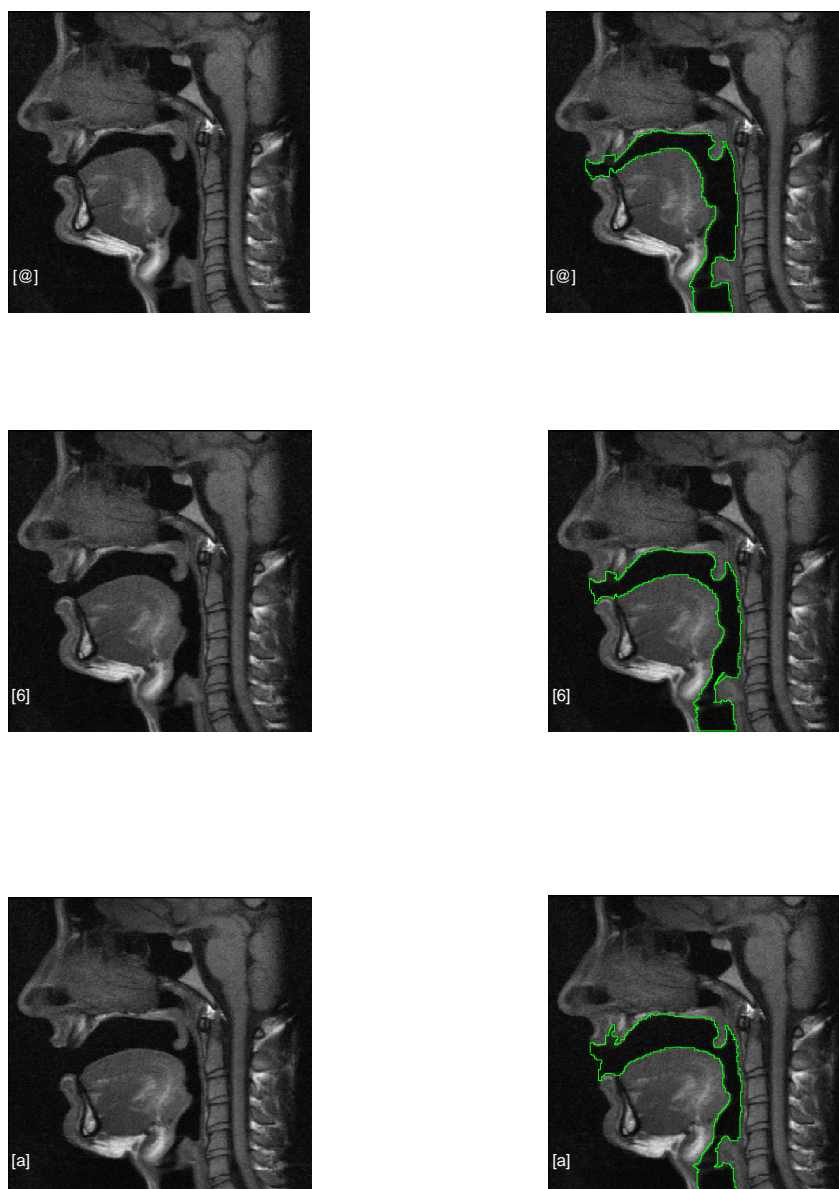


Figura 7. 2 - Vogais orais médias do PE [a] em cima, [6] ao centro e [a] em baixo, produzidas pelo falante PAA (imagem no plano sagital à esquerda e com o respectivo contorno à direita).

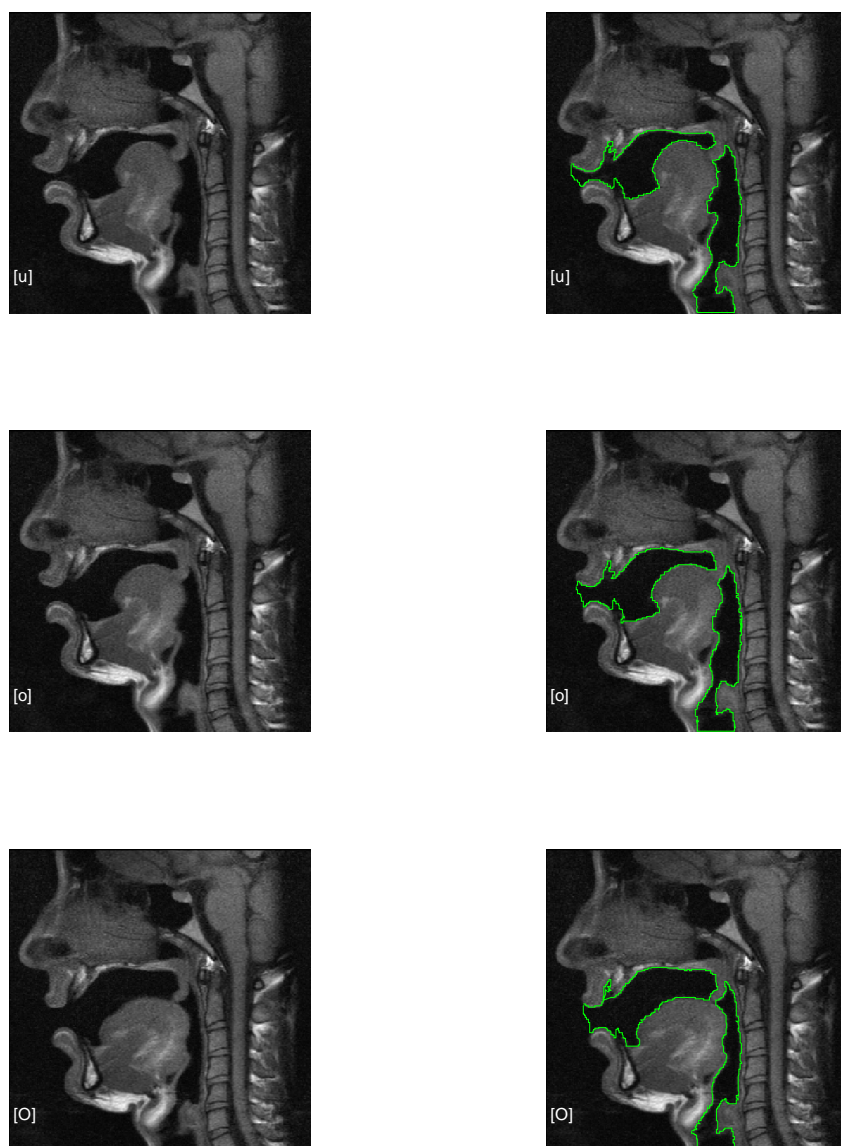
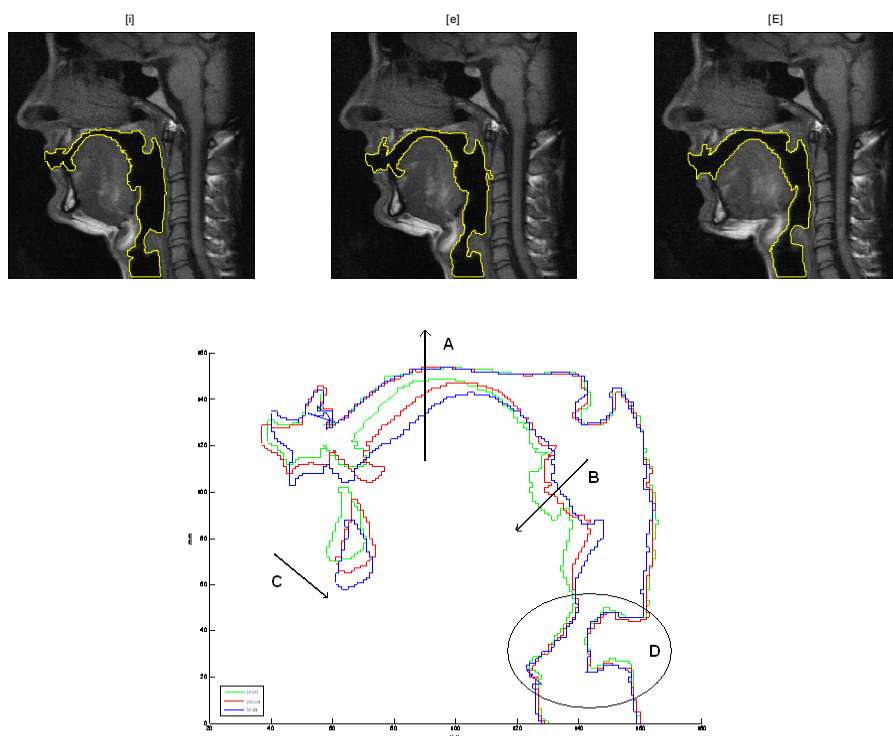


Figura 7. 3 - Vogais orais posteriores do PE [u] em cima, [o] ao centro e [O] em baixo, produzidas pelo falante PAA (imagem no plano sagital com o respectivo contorno à direita).

Apresenta-se de seguida, a sobreposição dos contornos e comparações efectuadas entre séries, nomeadamente entre vogais anteriores, médias e posteriores, entre si.

7.1.1.1 Comparação das vogais anteriores [i e E]



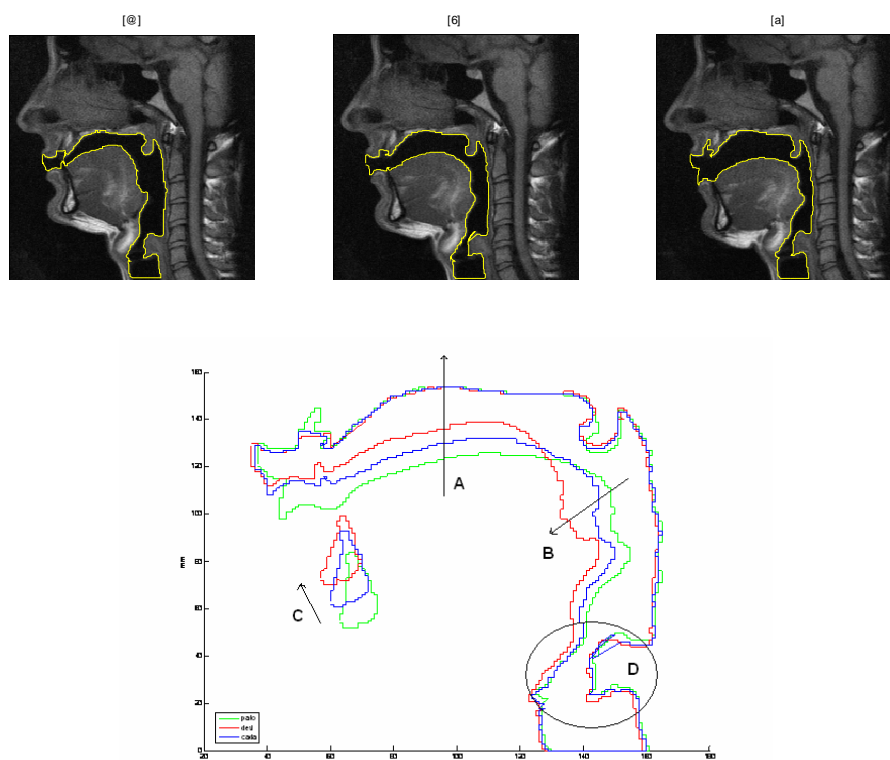
PRATT	[i]	[e]	[E]
[i]	1.000	0.776	0.655
[e]	0.776	1.000	0.699
[E]	0.655	0.699	1.000

Figura 7. 4 - Configurações do tracto para as vogais anteriores [i e E], no plano médio-sagital produzidas pelo falante PAA, tendo como referência as palavras [pipu], pêca [pek6] e leva [lEv6] (em cima). Sobreposição dos contornos referentes às configurações referidas (ao centro) e em baixo os IP relativos aos contornos.

Relativamente às vogais orais anteriores (Figura 7.4), verifica-se que dentro desta série, é a parte anterior do dorso da língua que se vai movimentando ou elevando, na direcção do palato duro **(A)**. Assim, o [E] (contorno azul) é produzido numa posição mais baixa; elevando-se e anteriorizando-se, na mesma direcção, são produzidos o [e] intermédio (a vermelho) e o [i], o mais elevado dentro da série (verde). Observa-se também, que a zona da raiz da língua **(B)** está mais anteriorizada no [i] do que no [E], o que contribui para o aumento da cavidade faríngea e diminuição da cavidade oral quando se passa de [E] a [i], o que está de acordo com o descrito por Barbosa (1994). Quanto ao grau de abertura oral, verifica-se que este é maior para o [E] uma vez que dorso da língua está numa posição mais baixa na cavidade oral (vogal semi-aberta) e menor para o [i] considerada uma vogal fechada, o que corrobora o descrito por Moutinho (2000). A mandíbula **(C)**, está numa

posição mais baixa em [E] e mais elevada e mais anterior em [i]. Para cada uma das três configurações o velo está elevado, não havendo alterações significativas da sua posição entre a produção do [E], [e] e [i]. Na zona da glote não se evidenciam neste plano, e para este falante, alterações significativas entre as três vogais (**D**). A configuração dos lábios e a sua posição também não é muito diferente, nas três vogais anteriores, com excepção de uma ligeira subida do lábio inferior para o [i], quando comparado com [e] e [E]. Finalmente, e em termos de semelhança de contornos, pela análise do IP, o [e] está mais próximo de [i] do que do [E].

7.1.1.2 Comparação das vogais médias [@ 6 a]



PRATT	[@]	[6]	[a]
[@]	1.000	0.761	0.628
[6]	0.761	1.000	0.749
[a]	0.628	0.749	1.000

Figura 7. 5 - Configurações do tracto para as vogais médias [@ 6 a], no plano médio sagital produzidas pelo falante PAA, tendo como referência as palavras devi [d@vi], cada [k6d6] e pato [patu] (em cima). Sobreposição dos contornos referentes às mesmas configurações (ao centro) e em baixo, os IP para os contornos referidos.

Relativamente às vogais centrais, e como é visível na Figura 7.5 e Figura 7.8, a vogal [@] (a vermelho) é produzida com o dorso da língua (**A**) na posição mais alta dentro da série (vogal alta), seguindo-se o [6] (a azul) que é uma vogal média e o [a] (a verde) uma vogal baixa, o que está de

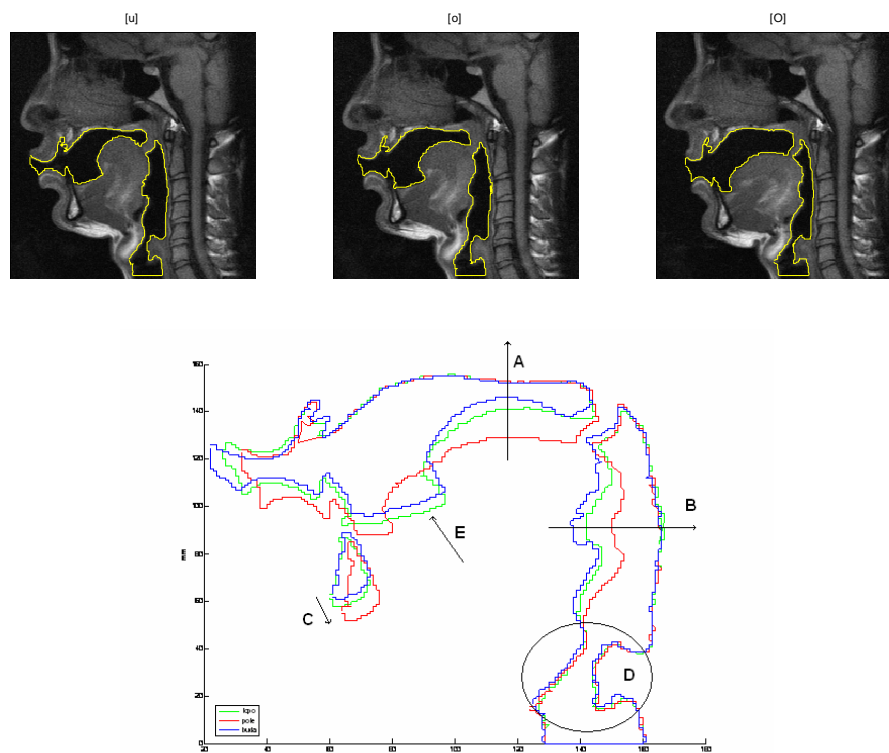
acordo com o descrito para o PE por Barbosa (1994). Se na figura 7.8 observarmos a coordenada (x) para o TD, para as três vogais, verificamos que ela é muito semelhante. Em relação à raiz da língua (**B**), verifica-se que esta se encontra mais anteriorizada durante a produção do [ə] do que do [ɔ] e [a] o que é também claramente visível na Figura 7.8 (Parâmetro TR). A vogal [a] apresenta valor de TR sempre mais posterior do que nas três vogais anteriores. A ponta da língua, acompanha o movimento do dorso sem que sejam aparentes movimentos independentes. O velo apresenta uma posição mais anterior nas vogais [ɔ] e [ə] do que em [a]. Não há diferenças visíveis entre as três vogais a nível da parede posterior da faringe.

Quanto à abertura da cavidade oral, verifica-se que esta é maior em [a] do que em [ɔ] e do que em [ə], o que está de acordo com a classificação descrita por Moutinho (2000) para as vogais: [a] é considerada uma vogal aberta, [ə] uma vogal fechada e [ɔ] é produzido com uma abertura da cavidade oral intermédia, sendo considerada semi-fechada. A mandíbula nesta série está mais baixa e posterior em [a] e mais elevada para o [ə]. Há uma sobreposição dos valores de abertura relativamente às vogais anteriores. Em relação aos lábios, o lábio inferior e superior são semelhantes para o [ɔ] e [ə], mas mais baixo e posterior em [a], o que pode estar relacionado com a posição mais baixa da mandíbula nesta série.

Analisando os IP obtidos, verificamos que há maior semelhança entre os contornos de [ə] com [ɔ] (0,7607) do que de [ə] com [a] (0,6219) e mesmo de [a] com [ɔ] (0,7489). A configuração de [ɔ] (a azul) aproxima-se de um tubo acústico de área aproximadamente uniforme (ressoador), aberto numa extremidade (lábios) e fechado na outra (a glote), tal como descrito em Mateus, et al. (1990).

7.1.1.3 Comparação das vogais posteriores [u o O]

Da comparação efectuada entre as vogais posteriores (Figura 7.6) pode observar-se que: na produção da vogal [u] (azul), o dorso da língua (TD) está na posição mais elevada dentro da série (vogal alta) e que na vogal [O] (vermelho) se encontra na posição mais baixa (**A**). Na região correspondente à parte anterior da língua, a coroa (**E**), verifica-se que esta se encontra numa posição mais posterior na produção de [o] e [u] do que em [O]. A parte posterior e raiz da língua são mais posteriores no [O] e [o] do que em [u]. A parte posterior do dorso dirige-se em direcção ao velo no [u] e [o], enquanto que no [O] se dirige mais em direcção à parede da faringe o que está parcialmente de acordo com o descrito por Barbosa (1994), que refere que na produção do [o] a parte posterior da língua vai na direcção da parede da faringe. A dimensão da cavidade faríngea, no que se pode observar no plano sagital, vai aumentando quando passamos de [O] para [u] com consequente diminuição da cavidade oral. Não se observam alterações a nível da parede posterior da faringe. O velo está numa posição mais baixa na produção do [O] do que nos outras duas vogais posteriores.



PRATT	[u]	[o]	[O]
[u]	1.000	0.812	0.372
[o]	0.812	1.000	0.412
[O]	0.372	0.412	1.000

Figura 7. 6 - Configurações do tracto para as vogais posteriores ou velares [u o O], no plano médio sagital produzidas pelo falante PAA, tendo como referência as palavras buda [bud⁶], tôpo [topu] e pote [pOt@] (em cima). Sobreposição dos contornos referentes às mesmas configurações (ao centro) e em baixo os IP para os contornos referidos.

Finalmente, a abertura da cavidade oral é maior em [O] (semi-aberta), média em [o] e, mais pequena em [u] (fechada). A mandíbula está mais baixa na produção do [O] do que na do [o] e [u]. A nível da glote não se observam diferenças dignas de referência. A protusão labial é maior na produção das vogais [u] e [o] do que em [O].

Através da análise dos IP, conclui-se que a configuração do [o] está mais próxima do [u] do que do [O] uma vez que o valor entre os primeiros é de 0,816 ([o] e [u]) e entre os segundos de 0, 415.

7.1.1.4 Comparação entre todas as vogais orais

Na Figura 7.7, pode observar-se a sobreposição dos contornos relativos a todas as vogais do PE, permitindo efectuar uma análise comparativa entre as séries. Na Figura 7.8 apresentam-se, também, os resultados obtidos relativos aos parâmetros articatórios (TD, TR, TT, JH, LL, UL) obtidos para todas as vogais orais e nasais do PE.

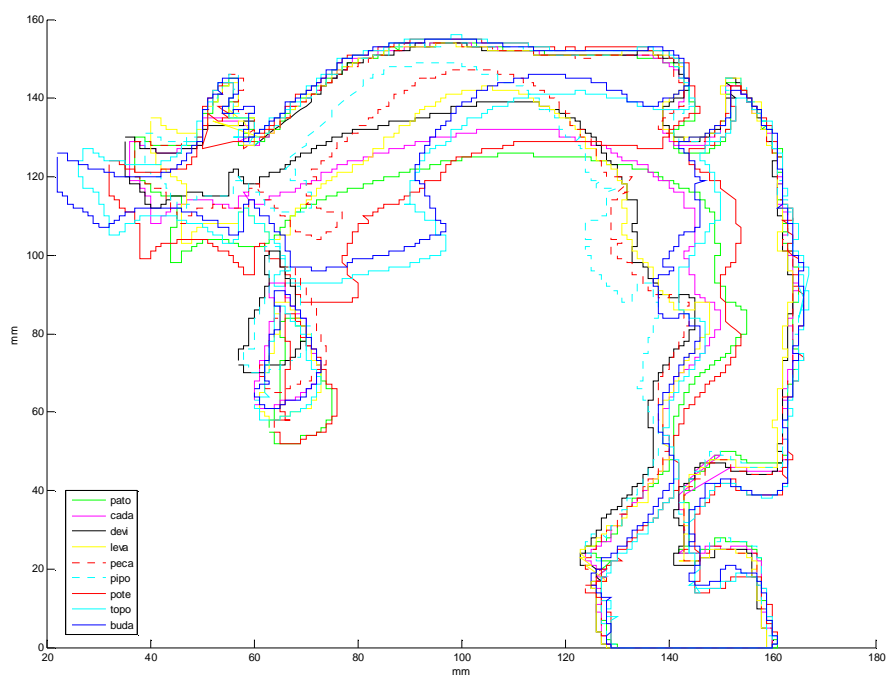


Figura 7.7 - Sobreposição dos contornos obtidos para todas as vogais orais do PE, para o falante PAA.

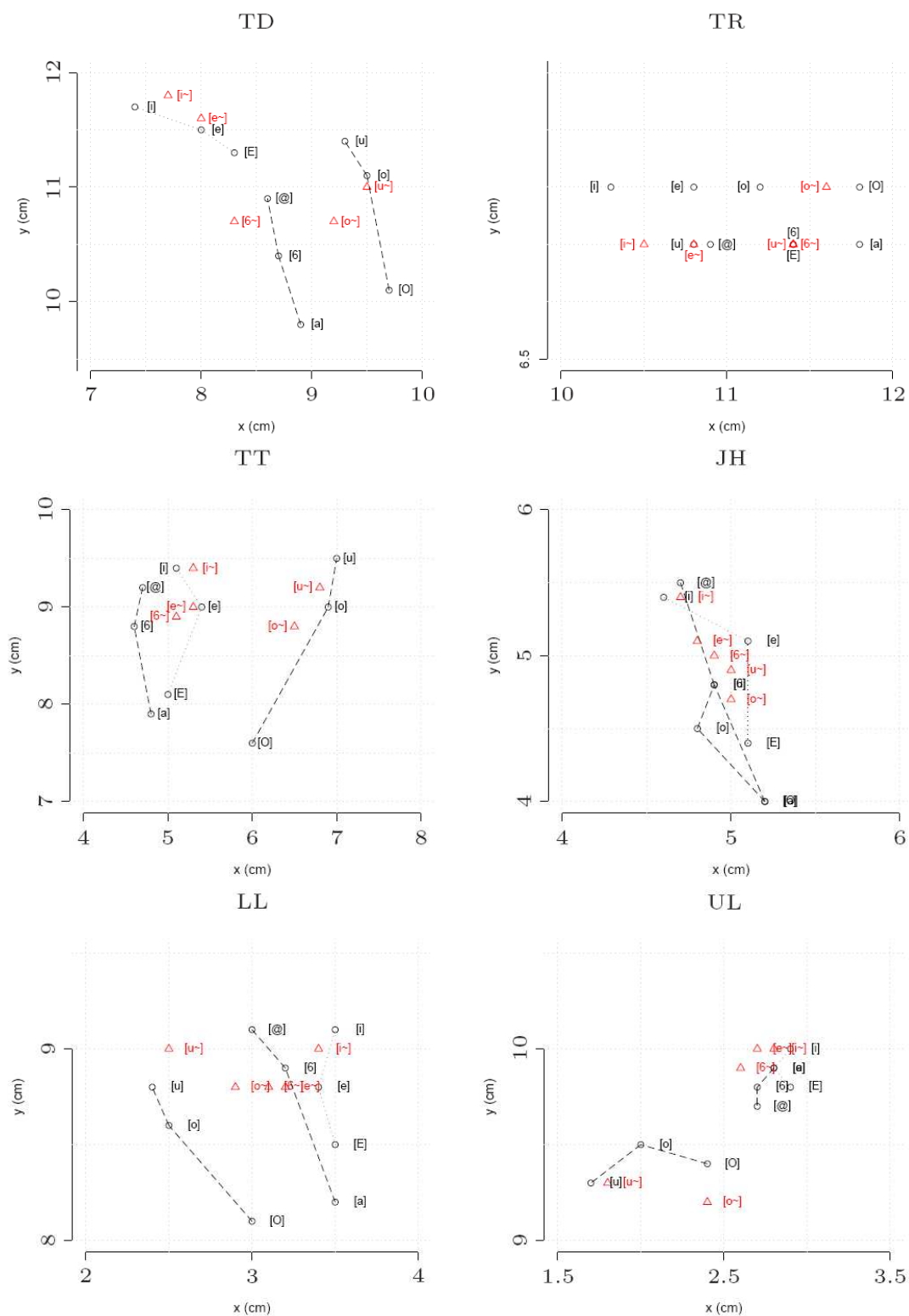


Figura 7. 8 – Parâmetros articulatórios obtidos para todas as vogais do PE. Do canto superior esquerdo: ponto mais alto do dorso da língua (TD), ponto mais recuado da língua, a nível do espaço intervertebral de C3-C4 (TR), ponta da língua (TT), altura da mandíbula (JH) e lábio inferior (LL) e superior (UL).

Comparando agora todas as vogais orais do PE, através da análise das Figuras 7.7 e 7.8, podemos verificar que, para este falante do PE:

- No que diz respeito à **altura do dorso da língua (TD)** e, conseqüentemente à altura da vogal em termos de espaço articulatorio, verifica-se que [ə] é a vogal mais alta das vogais médias, mas não tão alta como [u] e [i]. Em termos de avanço e recuo da língua, considerando uma direcção horizontal dentro da cavidade oral, pode confirmar-se que [E e i] são mais anteriores em relação a [a] e [ɔ], mas não tão claramente em relação a [ə];
- A protusão labial é evidente no [o], [O] e [u] (observável através dos contornos e pelos parâmetros UL, LL), o que está de acordo com as descrições teóricas, pois estas são as únicas vogais arredondadas ou labializadas do PE. Esta protusão poderá determinar um comprimento do tracto vocal maior durante a produção destes sons. De entre as três vogais, a protusão labial é maior no [u] e [o] e menor no [O];
- A dimensão da cavidade faríngea, no plano sagital, é maior na vogal [i] e mais pequena na vogal [O] e [a], o que em grande parte poderá ser explicado pelo posição mais posterior do dorso e raiz da língua durante a produção do [O] e do [a];
- No plano sagital, não são evidentes diferenças dignas de referência, a nível da posição do velo, para qualquer uma das vogais orais, existindo no entanto diferenças na zona da orofaringe, para as diferentes vogais, que se devem muito provavelmente a uma modificação de posição da porção posterior do dorso e da raiz da língua;
- Na produção das vogais do PE, não há qualquer variação ao nível da parede posterior da faringe, como se observa na produção de outras línguas.
- A mandíbula está numa posição mais baixa e posterior na produção do [O] e do [a] e mais alta e anterior na produção do [ə], ver também parâmetro JH;
- A glote está numa posição mais subida na produção do [O] e [a] e mais baixa em [u];
- As vogais, [i], [u] e [a] ocupam os extremos do espaço articulatorio (Figura 7.9), como se esperava (vogais cardinais)

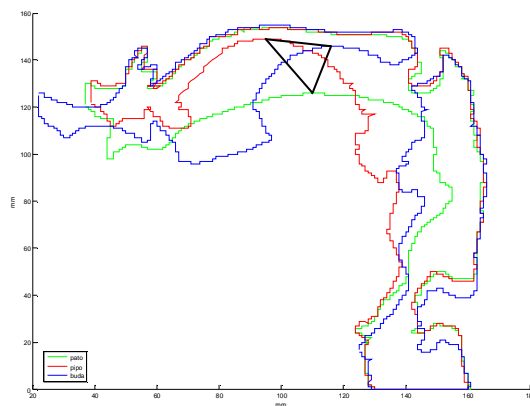


Figura 7. 9 - Ponto mais alto do dorso da língua durante a produção das vogais cardinais do PE: [i] (vermelho), [u] (azul) e [a] (verde).

7.1.1.5 – Acerca do [ə]

Na Figura 7.10, podemos observar a sobreposição dos contornos de [ə], [i] e [u]. Será também útil observar a figura 7.8 relativamente aos parâmetros (TD e TR).

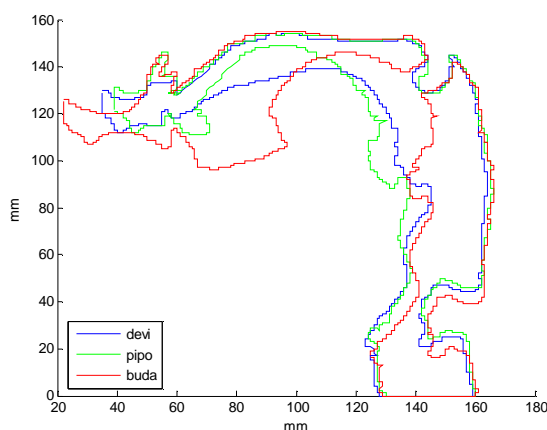


Figura 7.10 - Sobreposição dos contornos de [i] (verde), [u] (vermelho) e [ə] (azul).

Já referimos no capítulo 2, que existe alguma discussão na bibliografia relativamente aos símbolos utilizados para representar alguns dos sons do PE, utilizando os diversos alfabetos fonéticos. Existem também, algumas dúvidas, em torno da existência ou não de um *schwa* para o PE. Assim, e retomando um pouco esta temática, agora com alguns dados articulatórios, observamos o seguinte:

- O alfabeto SAMPA utiliza o símbolo @ para representar a vogal /e/ como nas palavras *devi* [d@vi] ou *de* [d@];
- Viana (1883), Cunha e Cintra (1986) citados por Veloso (1999) e Barbosa (1994) classificam a vogal /e/ como central, fechada, alta e não arredondada. Se esta descrição for adequada e correcta, o símbolo que lhe corresponde no IPA é o [i]. Veloso (1999) e Andrade & Viana (1996) também adoptam este símbolo;
- Moutinho (2000) representa esta vogal por [ə] e, no IPA este símbolo é reservado para representar o schwa;
- Cruz-Ferreira (1999), de acordo com dados acústicos, classifica esta vogal como a contrapartida não arredondada, de um [u] ligeiramente mais avançado e representa-o por [ɯ].

Do ponto de vista articulatório, para o falante deste estudo, e em resultado da observação de contornos (Figuras 7.7 e 7.10) e dos parâmetros articulatórios (Figura 7.8 para os parâmetros TD, JH) é possível verificar o seguinte:

- 1) A vogal [ə], no que diz respeito ao ponto mais alto do dorso da língua (TD), o que se relaciona com a altura da vogal, é a mais alta das vogais centrais, mas não tão alta como o [i] e o [u], para ser considerada uma vogal alta;
- 2) É central em termos do movimento de avanço /recuo da língua dentro da cavidade oral (TD);
- 3) Não apresenta arredondamento, o que pode ser inferido pela ausência de protusão (valores de UL e LL), quando comparado com um [u] ou [o];
- 4) Em termos de grau de abertura, se considerarmos apenas a abertura da cavidade oral, verifica-se que esta é a mais fechada das vogais centrais, mas mais aberta que o [E], tradicionalmente descrita como vogal semi-aberta. Assim, será talvez um pouco “forçado” classificar esta vogal como fechada.
- 5) Para este falante do PE, baseados apenas em parâmetros articulatórios, esta vogal pode ser considerada central, média, não arredondada, e não deve ser considerada fechada (pelo menos à luz do parâmetro TD), sendo o símbolo IPA que melhor lhe corresponde o [ə] ou no SAMPA o [ə]. Ressalvamos no entanto, o facto de esta informação ser relativa apenas a um falante.

7.1.2 Vogais nasais do PE

Como se referiu no capítulo dois desta dissertação, existem ainda algumas dúvidas no que diz respeito à configuração das vogais nasais do PE.

Na figura 7.11, apresentam-se as imagens obtidas e respectivos contornos para as cinco vogais nasais do PE [i~ e~ ɐ~ o~ u~].

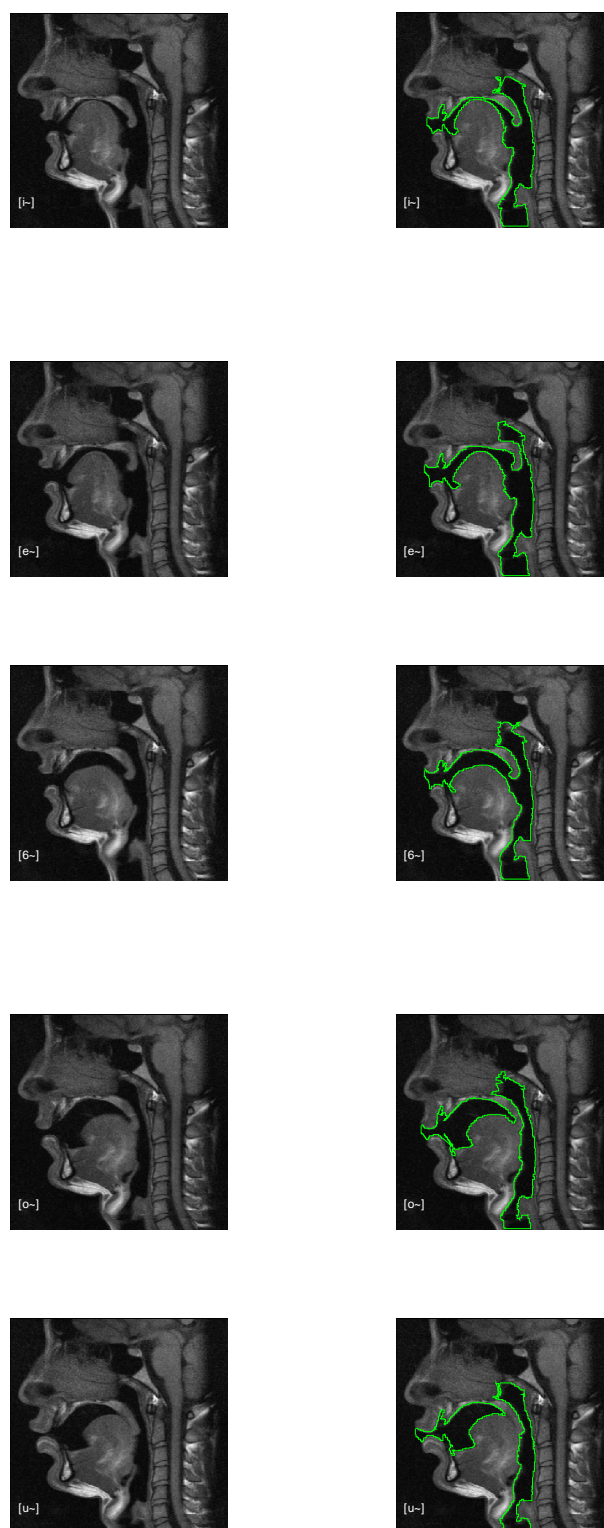


Figura 7. 11 - Vogais nasais do PE: à esquerda a imagem no plano médio sagital e à direita a imagem com o respectivo contorno, para o falante PAA. De cima para baixo, [i~],[e~],[ɛ~], [o~] e [u~].

Na Figura 7.12 apresenta-se a sobreposição dos contornos de todas as vogais nasais do PE.

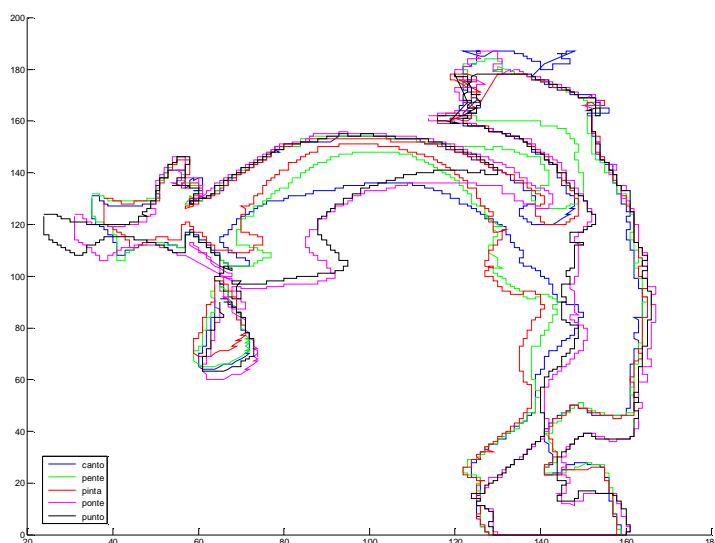


Figura 7.12 - Sobreposição dos contornos referentes a todas as vogais nasais do PE, para o falante PAA.

Analisando os contornos sobrepostos e recorrendo, mais uma vez, aos parâmetros articulatórios da Figura 7.8, podemos referir que, para este falante, se verifica o seguinte:

- De entre todas as nasais, as vogais [i~] e [e~] são produzidas com a língua numa posição mais anterior e elevada (ver TD);
- A vogal [ɐ~] é produzida numa posição central, mais baixa que [e~] e com uma altura muito próxima do [o~] (o [ɐ~] é geralmente descrito como vogal média a par com as vogais [e~] e [o~]);
- As vogais [o~] e [u~] apresentam-se mais recuadas em relação às anteriores e média;
- Em termos de altura da vogal, e para este falante em particular, seria mais correcto considerar estas vogais ([o~] e [u~]), como recuadas e médias; geralmente o [u~] é considerada uma vogal posterior, mas alta;
- As vogais [o~] e [u~] são as mais abertas de todas as nasais. As mais fechadas são as vogais anteriores [i~] e [e~]. O grau de abertura oral, no [ɐ~], é intermédio;
- A protusão labial é maior na produção do [u~] e semelhante à da vogal oral. A posição da mandíbula, ao contrário do que acontece na produção das vogais orais, não varia significativamente, entre as cinco nasais;
- O velo encontra-se aberto em qualquer dos sons nasais, mas a sua altura é um pouco diferente nas cinco vogais. Assim, o velo está numa posição mais elevada na produção da vogal anterior [e~] e, apesar de a sua altura ser semelhante para as vogais [ɐ~], [u~] e [o~], devido à diferente posição da língua (mais recuada nas posteriores), o velo encosta-

se ao dorso da língua na produção das posteriores o que não acontece na produção das anteriores e média [6];

- Não há diferenças a nível da parede posterior da faringe na produção dos cinco vogais nasais;
- As vogais nasais anteriores são produzidas com a zona da glote numa posição superior e mais avançada do que as posteriores.

7.1.3 Comparação entre vogais orais e nasais

Apresentamos nas figuras 7.13 a 7.18, as comparações efectuadas entre as vogais orais e nasais obtidas por sobreposição dos contornos e através dos parâmetros articulatórios (ver Figura 7.8).

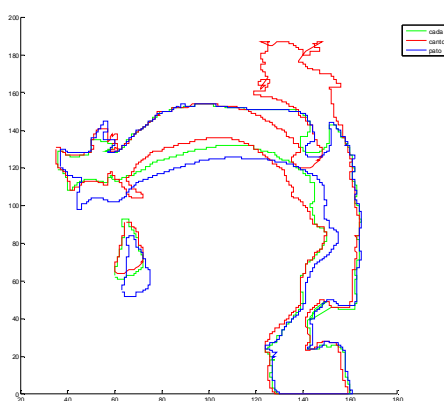


Figura 7. 13 - Sobreposição dos contornos referentes às configurações do tracto para as vogais orais [ɔ] e [a] e da vogal nasal [ɔ̃], produzidas pelo falante PAA.

Podemos verificar na Figura 7.13, por análise dos contornos sobrepostos relativos às vogais orais [ɔ] (verde) e [a] (azul) e da vogal nasal [ɔ̃] (vermelho), que a vogal nasal [ɔ̃] é produzida com a porção anterior do dorso da língua numa posição mais alta do que [ɔ] e que [a]. A porção do dorso, abaixo do velo, está mais anteriorizada em [ɔ̃] relativamente a [ɔ] ou em [a], muito provavelmente devido à posição, também mais baixa do velo na produção da nasal. A abertura da cavidade oral é maior em [a], sendo também, neste som, que a mandíbula se encontra numa posição mais baixa e posterior. O velo encontra-se numa posição mais baixa na produção da vogal nasal [ɔ̃], permitindo uma maior abertura para a cavidade nasal. A dimensão da cavidade faríngea (orofaringe), neste plano e para este falante, é maior na produção da nasal, do que nas vogais orais correspondentes. Não se observam diferenças a nível da parede posterior da faringe, nem da glote. A raiz da língua (TR) está mais recuada na vogal oral [a] e praticamente na mesma posição na produção do [ɔ] e do [ɔ̃]. Na figura 7.14, apresenta-se ainda a sobreposição destes três sons com o [ə].

Podemos verificar que os contornos a vermelho e preto referentes a [a] e a [a~], respectivamente, são muito próximos, sendo a altura de [a~] apenas ligeiramente mais baixa que [a] e mais elevada que [ã]. Este facto pode ser também observado na Figura 8, parâmetro TD.

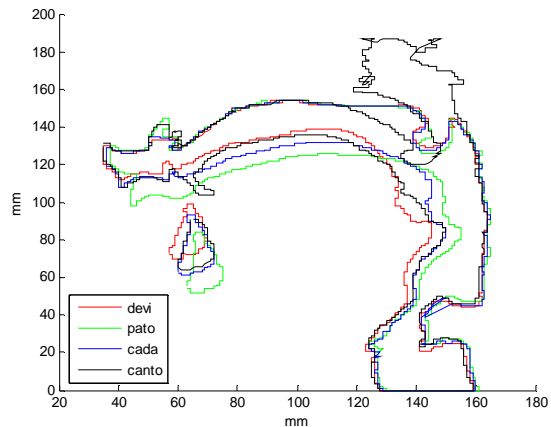


Figura 7.14 - Sobreposição dos contornos referentes às configurações do tracto das vogais orais [a] e da vogal nasal [a~], produzidas pelo falante PAA.

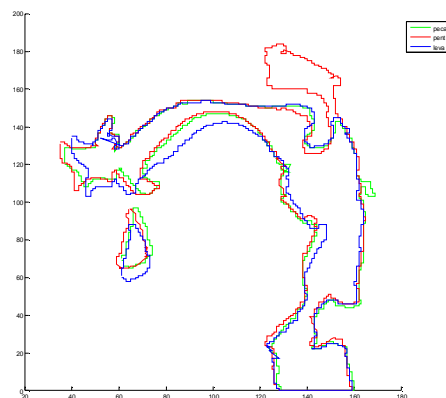


Figura 7.15 - Sobreposição dos contornos referentes às configurações do tracto para as vogais orais [e] e da vogal nasal [e~], produzidas pelo falante PAA.

Pela análise da Figura 7.15, relativa à sobreposição dos contornos das vogais [e] (verde) e [e~] (vermelho) verifica-se que estes são praticamente indistintos, havendo apenas uma muito ligeira anteriorização do velo na vogal nasal, configurando uma pequena abertura para a cavidade nasal. Os IP relativos a estas três vogais confirmam uma maior semelhança entre [e] e [e~] com um valor de 0,7670, do que entre [E] (azul) e [e~] com IP de 0,6013. No entanto, não devemos esquecer que o nosso corpus contempla apenas um falante e que a produção da vogal nasal [e~] poderá ser efectuada por configurações entre o [e] e o [E], em diferentes regiões (Teixeira et al., 2003).

A parte anterior do dorso da língua é mais baixa na vogal [E] do que em [e] e [e~]. Estes resultados diferem, em parte, dos obtidos para as vogais nasais do francês (Delvaux et al, 2002a), uma vez que no francês a vogal oral é produzida com uma maior elevação do dorso da língua quando comparada com a correspondente vogal nasal.

Não se observam diferenças a nível da parede posterior da faringe, nem a nível da glote. No que diz respeito à cavidade orofaríngea, não há diferenças significativas, neste plano entre vogais orais e nasais.

Na figura 7.16, apresentamos a sobreposição dos contornos para a vogal oral [i] e nasal [i~]. Da sua análise podemos referir que:

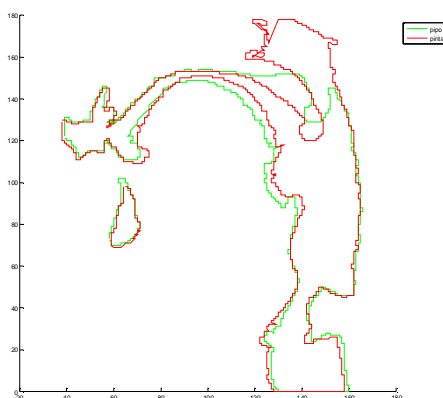


Figura 7. 16 - Sobreposição dos contornos referentes às configurações do tracto para a vogal oral [i] e da vogal nasal [i~], produzidas pelo falante PAA.

A vogal [i~] (a vermelho) é produzida com uma configuração muito semelhante, apenas ligeiramente superior e um pouco mais recuada do que [i] (ver Figura 7.16 e TD na figura 7.8). O velo na produção da nasal está mais abaixado, permitindo o acoplamento com o tracto nasal, o que é de esperar na produção de uma vogal nasal. Não se observam diferenças a nível da orofaringe, parede posterior da faringe e glote na produção destes dois sons. Curiosamente, a posição da raiz da língua e da parte posterior do dorso, não se apresentam mais avançados, o que se poderia supôr devido ao abaixamento do velo.

Na figura 7.17, apresenta-se a sobreposição dos contornos para as vogais orais [o] e [O] e vogal nasal [o~].

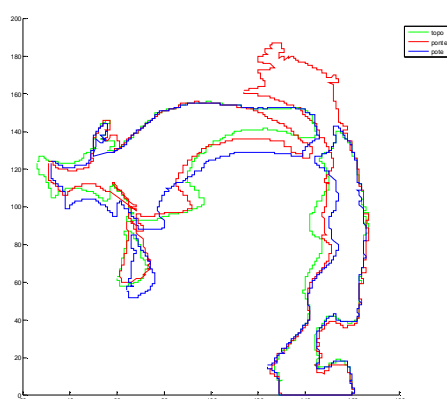


Figura 7.17 - Sobreposição dos contornos referentes às configurações do tracto para as vogais orais [o] e [O] e para a vogal nasal [o~], produzidas pelo falante PAA.

Da análise da Figura 7.17 e recorrendo também à Figura 7.8 (TD, TT, TR), podemos observar que: no que diz respeito à altura do dorso da língua e conseqüentemente da altura da vogal, a vogal nasal [o~] (vermelho) é produzida com uma configuração entre o [o] (verde) e o [O] (azul). Na zona da língua correspondente à ponta e lâmina (TT), a configuração de [o~] é mais próxima de [o] do que do [O]. Na produção do [O], a raiz da língua (TR), está numa posição mais posterior relativamente a [o~] e [o]. Há pequenas diferenças a nível da cavidade faríngea, estando esta mais estreitada na produção da vogal oral [O]. O velo encontra-se mais elevado na produção das orais, no entanto, para qualquer das vogais posteriores (orais e nasais) a úvula toca na língua. Será importante verificar o que se passa no corpus 3D, para estes sons.

Finalmente, na figura 7.18, apresentamos a sobreposição dos contornos, para o par [u] e [u~].

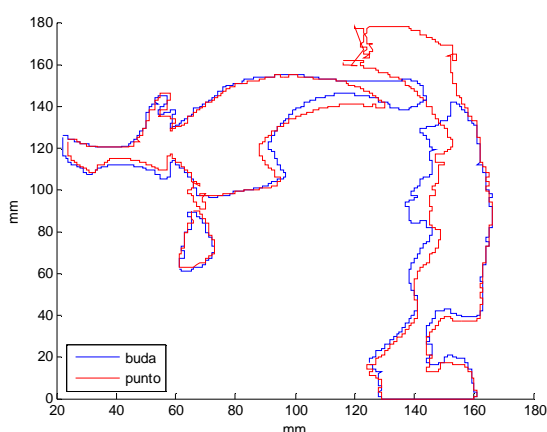


Figura 7.18 - Sobreposição dos contornos referentes às configurações do tracto para a vogal oral [u] e para a vogal nasal [u~], produzidas pelo falante PAA.

Como se pode observar na Figura 7.18, as configurações do [u] (azul) e de [u~] (vermelho), no que diz respeito à cavidade oral, são semelhantes, sendo a grande diferença apenas ao nível do

velo que se encontra mais baixo na vogal nasal com conseqüente abertura da passagem velo-faríngea. A cavidade faríngea está mais estreitada na produção da nasal, do que na oral.

De forma genérica, e no que diz respeito à cavidade faríngea, estes resultados diferem dos obtidos para a língua francesa por Delvaux, Metens & Soquet, (2002a b) e Demolin et al. (2003), uma vez que, na produção das vogais francesas se verifica que a faringe está sempre mais estreitada durante a produção das vogais nasais do que nas suas correspondentes orais, o que nem sempre se verificou neste falante do PE. Descrições mais antigas (Delattre, 1968) referem já uma forma diferente de nasalização para o português, comparativamente ao Francês, referindo o facto de as vogais nasais do PE, se manterem relativamente fechadas, havendo apenas abaixamento do velo e não havendo o ajuste, como acontece para o Francês, da cavidade faríngea ao volume da cavidade vélica.

7.1.4 Consoantes oclusivas

Apresentamos de seguida, nas figuras 7.19 e 7.20, as imagens e contornos referentes às consoantes oclusivas do PE. Apresentamos primeiro as oclusivas surdas, na figura 7.19 e na figura 7.20 as oclusivas sonoras.

7.1.4.1 Oclusivas surdas [p t k]

Na Figura 7.19, apresentam-se as imagens com respectivos contornos durante a produção das oclusivas surdas (ou não vozedas) do PE, produzidas pelo falante PAA em contexto VCV, em que $V = [a]$

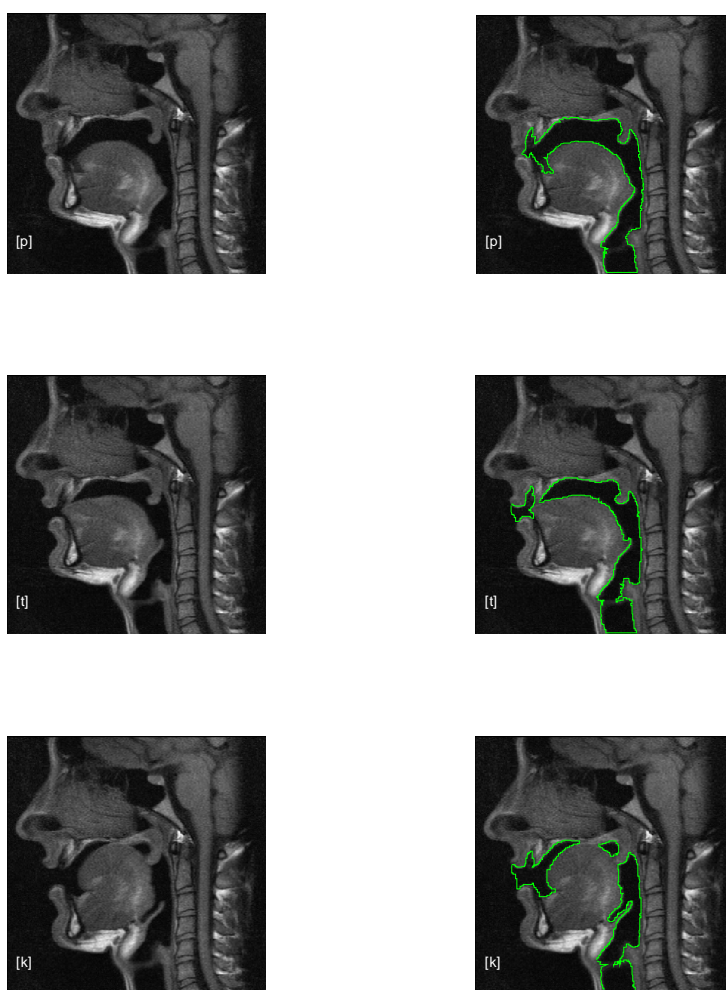


Figura 7. 19 - Imagens correspondentes à produção das oclusivas surdas [p] em cima, [t] ao meio e [k] em baixo, produzidas pelo falante PAA, no contexto V_1CV_2 em que $V_1 = V_2 = [a]$.

Pela análise da Figura 7.19, podemos verificar que na produção do [p] há aproximação e fecho dos lábios, uma vez que é uma oclusiva bilabial, na produção do [t] e apesar de não ser visível o

contorno dos dentes verifica-se uma aproximação da ponta da língua da língua ao sector anterior da arcada dentária superior (SAADS), o que está de acordo com o descrito na literatura para o PE (Moutinho, 2000; Mateus, Falé & Faria, 2005). Na produção do [k], o ponto de articulação não será claramente velar, ficando a ideia de que o ponto de maior constrição será, ligeiramente mais anterior, na zona de transição do palato com o velo. No entanto, uma só imagem no plano sagital poderá não ser suficiente para verificar, com rigor, o ponto onde ocorre a oclusão (o palato ou o velo).

7.1.4.2 Oclusivas sonoras [b d g]

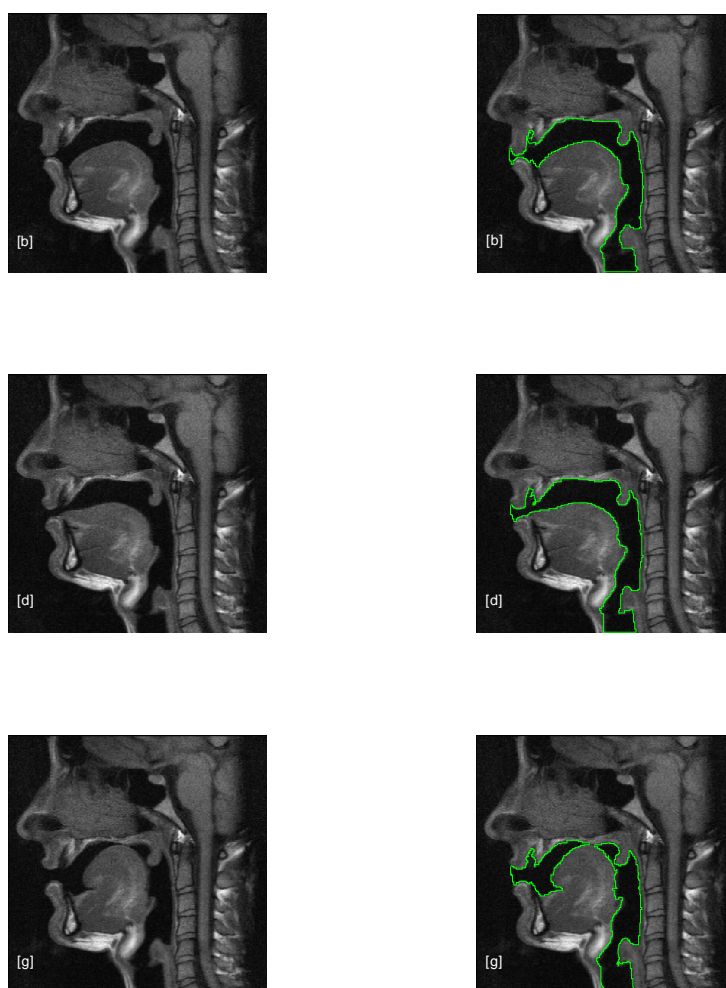


Figura 7. 20- Imagens e respectivos contornos correspondentes à produção das oclusivas sonoras [b] em cima, [d] ao meio e [g] em baixo, produzidas pelo falante PAA, no contexto V_1CV_2 em que $V_1=V_2=[a]$.

Pela análise da Figura 7.20, podemos observar que, à semelhança das oclusivas surdas, também nas sonoras se verifica fecho dos lábios na produção do [b]. O [d] é produzido com aproximação da coroa da língua ao SAADS e que o [g] é também produzido na transição palato-velo.

7.1.4.3 Comparações pelo ponto de articulação

Apresenta-se, de seguida, a comparação efectuada entre as oclusivas surdas e as sonoras tendo em consideração o ponto de articulação. Podemos observar, pela análise da Figura 7.21, e no que diz respeito às oclusivas bilabiais, que na produção do [p] a parte posterior e raiz da língua está mais posterior que em [b], o que poderá dar origem a uma diminuição das distâncias médio sagitais na zona da cavidade faríngea. Nas dentais, o [t] é produzido com a língua numa posição mais elevada do que em [d], com conseqüente diminuição da distâncias médio sagitais na zona da cavidade oral. Para todas as oclusivas sonoras, existe um estreitamento na zona da glote, em relação às correspondentes surdas, o que poderá ser explicado pela necessidade de se efectuarem ajustes musculares, para a obtenção de uma tensão adequada nas pregas vocais para produzir vozeamento.

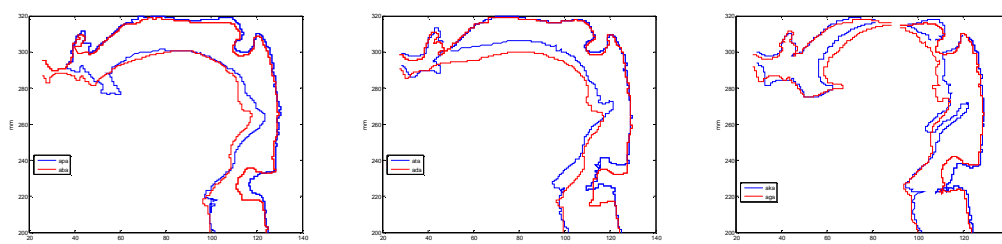


Figura 7. 21 - Sobreposição dos contornos obtidos para as oclusivas bilabiais [p] e [b] à esquerda, para as dentais [t] e [d] ao centro e para as velares [k] e [g] à direita, produzidas pelo falante PAA, em contexto V_1CV_2 em que $V_1= V_2 = [a]$.

7.1.5 Consoantes Nasais

Apresentam-se agora as imagens e contornos relativos às configurações obtidas para as consoantes nasais (Figura 7. 22) do PE: [m], [n] e [J].

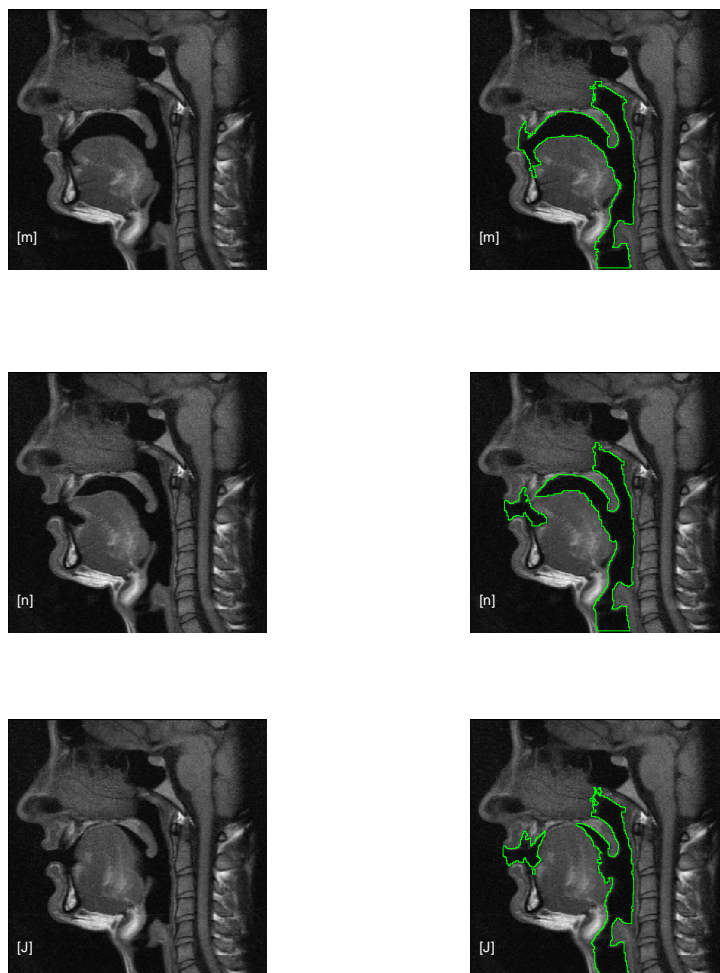


Figura 7. 22 - Imagens e respectivos contornos correspondentes à produção das consoantes nasais: [m] em cima, [n] ao centro e [J] em baixo, produzidas pelo falante PAA, tendo como referências as palavras *cama* [k6m6], *cana* [k6n6] e *canha* [k6J6].

Como podemos observar pelas imagens e contornos, a nasal [m] é produzida com oclusão labial (bilabial), a consoante [n] é produzida com oclusão formada pela região da coroa na zona alveolar superior. Esta descrição está de acordo com Moutinho (2000) e Mateus, Falé & Freitas (2005). Cruz-Ferreira (1999) considera o [n] como dental. No que diz respeito ao [J], o ponto de articulação é claramente palatal, com oclusão formada pelo dorso da língua e o palato. O [J] é produzido (como esperado) com o dorso da língua na posição mais elevada dentro das três

nasais, seguida pelo [n] e [m]. A raiz da língua (TR) é mais posterior no [m], seguida pelo [n]. Na produção de qualquer um destes sons é claramente visível a abertura do velo. Comparando agora, as consoantes nasais com as oclusivas surdas e sonoras, tendo em atenção o ponto de articulação (Figura 7.23) verifica-se o seguinte: os contornos referentes ao [m] e [b] são praticamente sobreponíveis, com excepção da posição do velo, uma vez que um é um som oral [b] e o outro nasal. Comparando com a não vozeada [p], verifica-se que existem diferenças a nível da glote. A raiz da língua está também mais posterior na oclusiva oral não vozeada. Para o grupo do [n], [t], [d] (à direita) o comportamento é semelhante. Informação complementar relativa às consoantes nasais é apresentada mais adiante, na secção relativa aos resultados do corpus 3D.

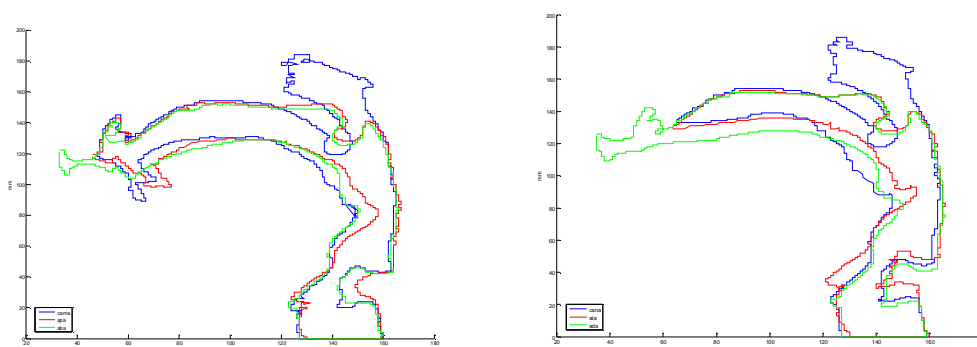


Figura 7. 23 - Comparação das oclusivas orais e nasais pelo ponto de articulação. À esquerda sobreposição dos contornos para as bilabiais [m], [p] e [b], à direita para [n], [t], [d].

7.1.6 Fricativas

De seguida, incluem-se as imagens obtidas para as fricativas surdas (Figura 7.24 e 7.25) e para as fricativas sonoras (Figura 7.26), obtidas tendo como referência palavras com fricativas em início de palavra, por exemplo [f] em [fal6]. À esquerda apresenta-se apenas a imagem e à direita a imagem com respectivo contorno.

7.1.6.1 Fricativas surdas [f s S]

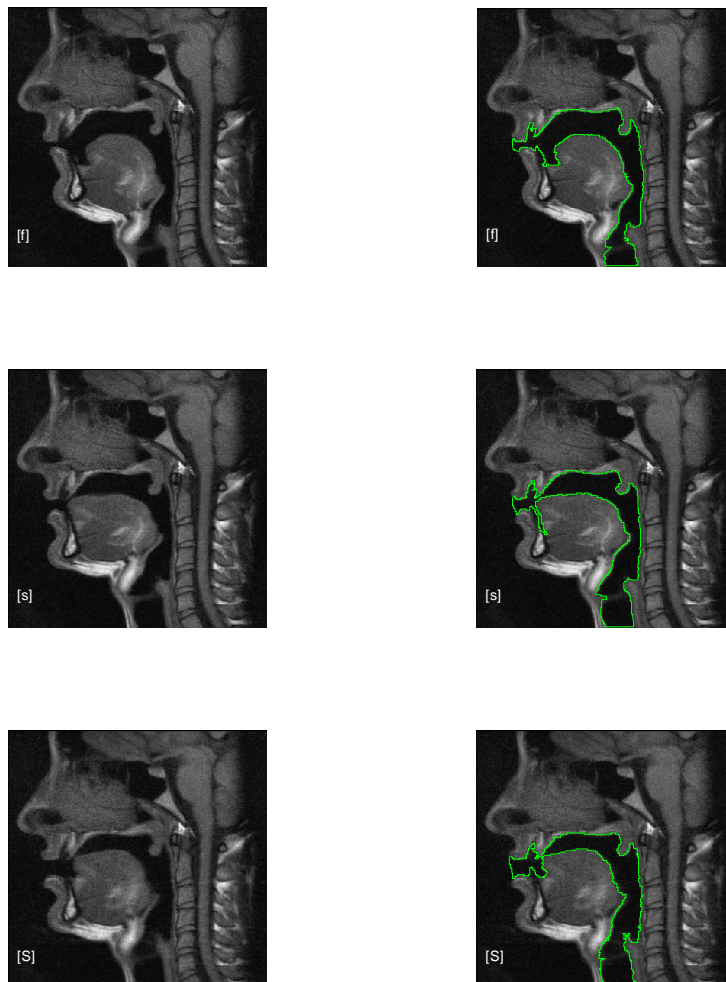


Figura 7. 24 - Imagens e respectivos contornos, correspondentes à produção das fricativas surdas em início de palavra: [f] em cima, [s] ao centro e [S] em baixo produzidas pelo falante PAA, tendo como referência as palavras **f**ala [fal6], **s**aca [sak6] e **ch**á [Sa] respectivamente.

Pela análise das imagens, (Figura 7.24), podemos inferir pela posição dos lábios (os incisivos superiores não estão aqui co-registados) que o [f] é produzido através da aproximação entre o lábio inferior (suportado pelo maxilar inferior) com o sector anterior da arcada dentária superior (fricativa lábio-dental). O [s] produzido com aproximação da ponta da língua/lâmina à região dos alvéolos superiores e o [S] com aproximação da ponta da língua a uma zona ligeiramente mais posterior em relação ao [s] mas que poderá ainda ser alveolar. Através da sobreposição dos contornos (figura 7.25), é possível observar que a ponta da língua está mais recuada e elevada, na produção do [S] relativamente ao [s]. Tradicionalmente o [S] é descrito como sendo pré-palatal

ou palato-alveolar (para outros) produzida com a coroa da língua. As imagens não são suficientemente esclarecedoras, restando dúvidas relativamente ao local exacto do ponto de articulação. Podemos afirmar claramente que, neste falante, o [S] foi produzido com o dorso da língua numa posição mais elevada que o [s].

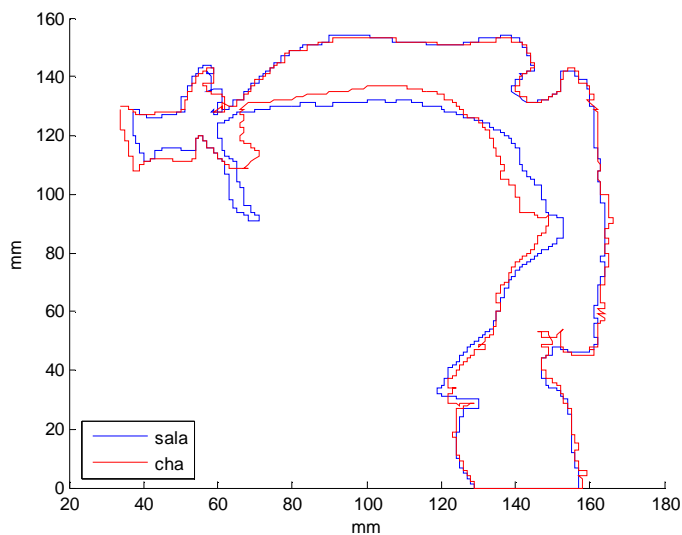


Figura 7. 25 – Sobreposição dos contornos, relativos ao [s] (a azul) e [S] (a vermelho).

7.1.6.2 Fricativas sonoras [v z Z]

As imagens e contornos da figura 7.26 dizem respeito às fricativas sonoras, [v] em cima, [z], ao centro e [Z] em baixo.

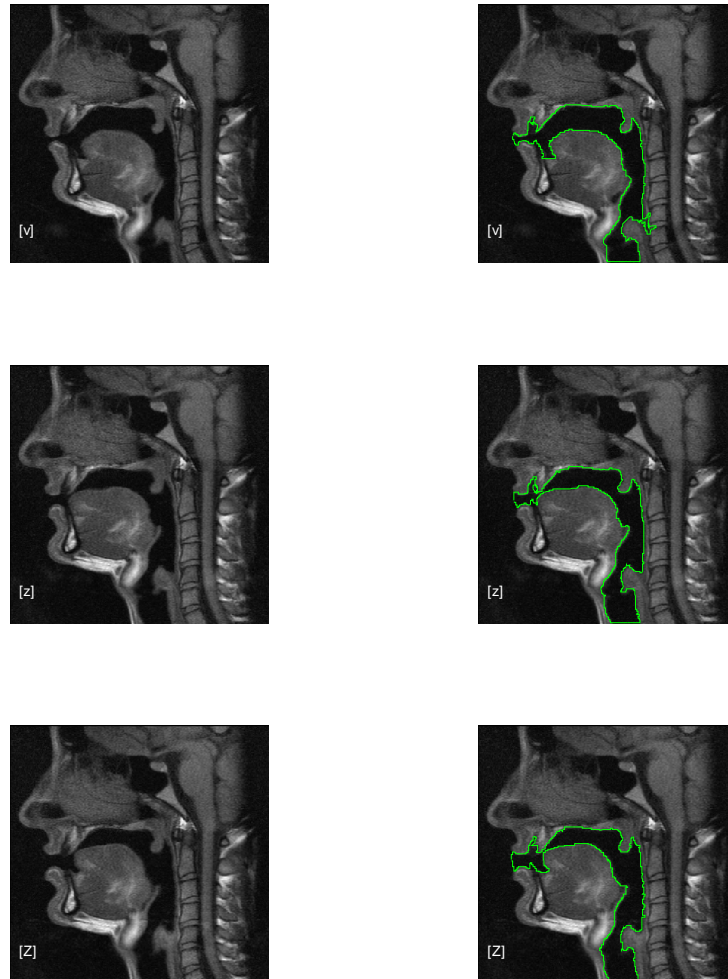


Figura 7. 26 - Imagens e respectivos contornos, correspondentes à produção das fricativas sonoras em início de palavra: [v] em cima, [z] ao centro e [Z] em baixo produzidas pelo falante PAA, tendo como referência as palavras *vaca* [vakô], *zarpa* [zarpô] e *jacto* [Zatu], respectivamente.

Nas fricativas sonoras verifica-se também que o [v] é produzido com a aproximação do lábio inferior ao SAADS (lábio-dental), o [z] com aproximação da ponta da língua/lâmina à zona alveolar, e o [Z], com aproximação da ponta da língua a um ponto mais posterior, mas aparentemente, tal como na correspondente surda, ainda na zona alveolar. Na produção do [Z], a ponta da língua e o dorso estão mais elevados relativamente a [z].

7.1.6.3 Comparação pelo ponto de articulação

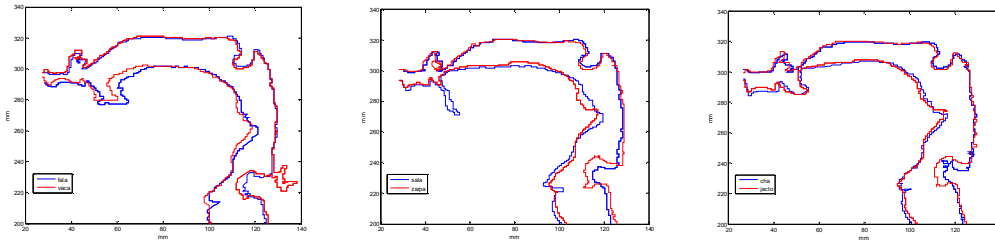


Figura 7. 27 - Comparação dos contornos relativos a pares de fricativas surda /sonora em função do ponto de articulação ([f] e [v] à esquerda, [s] e [z] ao centro e [ʃ] e [ʒ] à direita).

Podemos verificar pela análise dos contornos sobrepostos (Figura 7.27), que as diferenças na configuração, para o mesmo ponto de articulação não são muito acentuadas (pelo menos pelo que se observa no plano médio-sagital) nos pares surda-sonora. Ao nível da glote, há um maior estreitamento nas sonoras, qualquer que seja o ponto de articulação o que também já se verificava na produção das consoantes oclusivas e que como já referimos, estará relacionado com as exigências relativas às diferenças de pressão e ajustes necessários para tensão adequada à vibração das pregas vocais, durante a produção dos sons sonoros.

7.1.7 Laterais

Apresentam-se, de seguida (Figura 7.28 e 7.29) as configurações do tracto vocal, para as laterais: [l] em início da palavra como na palavra *laço* [lasu], o alofone [l~] como na palavra *mal* [mal~], em final de palavra ou sílaba; e o [L] tal como produzido na palavra *falha* [faL6].

Como se pode verificar pelas imagens da figura 7.28, e particularmente pela sobreposição dos contornos relativos aos dois alofones do [l] (Figura 7.29), o ponto de articulação é o mesmo para ambas variantes, ou seja apico-dental ou apico-alveolar, produzido com a ponta da língua na região denti-alveolar. Nas duas variantes observa-se uma elevação da porção posterior do dorso da língua, em direcção ao véu do palato. Os contornos são praticamente idênticos para os dois alofones.

O [L] como se pode observar na figura 7.28 (em baixo), é produzido, com aproximação da lâmina e porção anterior do dorso da língua na região palato-alveolar.

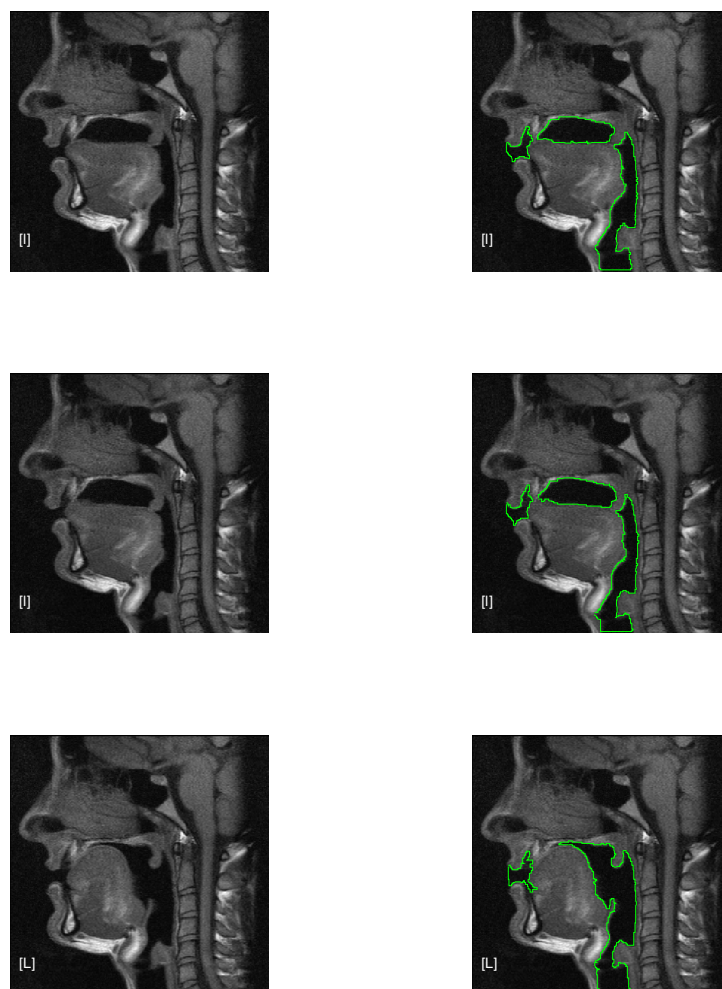


Figura 7. 28 - Imagens e respectivos contornos, correspondentes à produção das laterais, [I] em cima, [I~] ao centro e [L] em baixo, produzidas pelo falante PAA, tendo como referência as palavras laço [lasu], mal [mal~] e falha [faL6], respectivamente.

7.1.7.1 Comparações entre as laterais [l] e [l~] e [l] com [L]

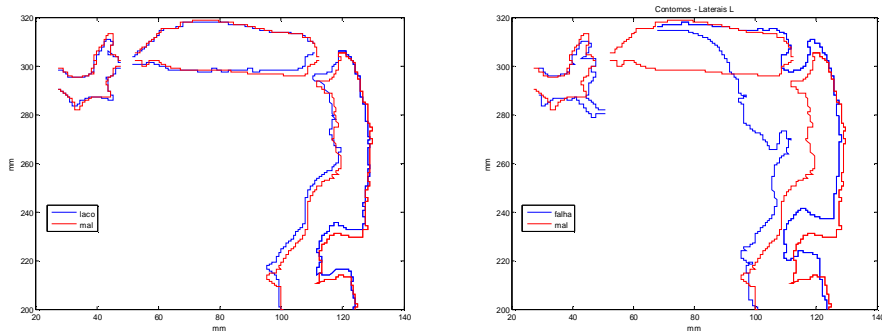


Figura 7. 29 - Sobreposição dos contornos, referentes a [l] e [l~] à esquerda e à direita de [l] com [L].

Em relação às laterais, a posição dominante relativamente a estes sons, nas últimas décadas é de que o fonema /l/ apresenta dois alofones: um que ocorre em início de palavra ou de sílaba e outro que só ocorre em final de sílaba e que se designa por l-velarizado. Este último, além da constricção efectuada pela ponta da língua nos alvéolos (primária), apresenta uma articulação secundária formada pela elevação do dorso da língua, em direcção ao véu do palato (Mateus et al., 1990; Andrade, 1999; Moutinho, 2000). Esta diferença existe também na língua inglesa, sendo atribuído ao [l] velarizado o nome de “dark” e ao [l] não velarizado o nome “light” (Narayanan, Alwan & Haker, 1997). No entanto, algumas descrições mais antigas do PE (Stevens, 1954), apontam para uma velarização do [l], também em início de sílaba. Este facto, é corroborado por Andrade (1997, 1999) que verificou, também, em três falantes da zona de Lisboa, a ocorrência de velarização em início de sílaba ou em “ataque”. Apesar de, não ser possível fazer generalizações, dada a natureza do corpus adquirido, as configurações do tracto vocal neste falante, para as variantes do [l], apontam para a possibilidade de existência de velarização, também em início de sílaba. Outro dado importante, é que, ao contrário dos falantes referidos nos estudos de Andrade (1997, 1999) provenientes de Lisboa, este falante é proveniente da região Norte. A análise da informação 3D, mais adiante, poderá complementar a informação extraída dos contornos, uma vez que permite obter informação que não se restringe à linha média e porque foi introduzido um terceiro contexto (/l/, em posição inicial, mas intervocálico).

A consoante [L] no PE é tradicionalmente descrita como palatal, de forma idêntica ao [J]. No entanto, Sá-Nogueira (1938) para o PE, refere a possibilidade de esta consoante ser produzida numa zona mais anterior, quando comparada com a palatal [J]. Num artigo recente, para o Catalão da zona de Maiorca, Recasens & Espinosa (2006), referem que a lateral palatal (L), não pode ser articulada exclusivamente na área palatal, referindo que as línguas de origem Românica, onde se

inclui o Português, apresentam também um fechamento na região alvéolo-palatal, podendo mesmo ser alveolar. Assim, e no que diz respeito a estas duas palatais [J] e [L], a constricção frontal decresce no sentido [L] > [J]. Neste falante do PE, como já referimos a produção da lateral [L] é bastante anterior, o que está de acordo com estes autores e com descrições mais antigas do PE (Cagliari, 1977 citado por Recasens & Espinosa, 2006).

7.2 Coarticulação

Foi nosso objectivo estudar o efeito do contexto em alguns sons do PE. Os sons estudados foram as oclusivas e fricativas, surdas e sonoras, do PE. Pretendemos observar qual o efeito do contexto neste caso, o efeito da vogal, na produção destes sons. No corpus escolhido, o contexto é sempre simétrico ($V_1=V_2$), do tipo V_1CV_2 , com a consoante entre as vogais cardinais [i a u]. Pensamos que, dada a forma como foi adquirido o corpus, poderemos estudar o efeito da vogal precedente sobre a consoante, mas não o contrário. Também, não foram contemplados contextos assimétricos, nem foram adquiridas configurações do tipo CVC, pelo que não será possível tirar conclusões acerca da influência da consoante, na produção das vogais.

Na análise efectuada, tivemos em atenção, essencialmente, a posição das diferentes porções da língua (ponta, lâmina, dorso, raiz), a mandíbula (coarticulação lingual e mandibular) fazendo referência também à protusão labial uma vez que é a informação que podemos extrair ou inferior, dos contornos sagitais. No final, comparamos estes resultados com os obtidos para outras línguas uma vez que não existe, que seja do nosso conhecimento, nenhum estudo semelhante para o PE.

7.2.1 Coarticulação nas oclusivas

Nas figuras seguintes, apresentam-se as imagens referentes à sobreposição dos contornos relativos a cada uma das oclusivas surdas (Figura 7.30) e sonoras (Figura 7.31), produzidas no contexto das três vogais cardinais do PE [a i u].

Oclusivas surdas [p t k]

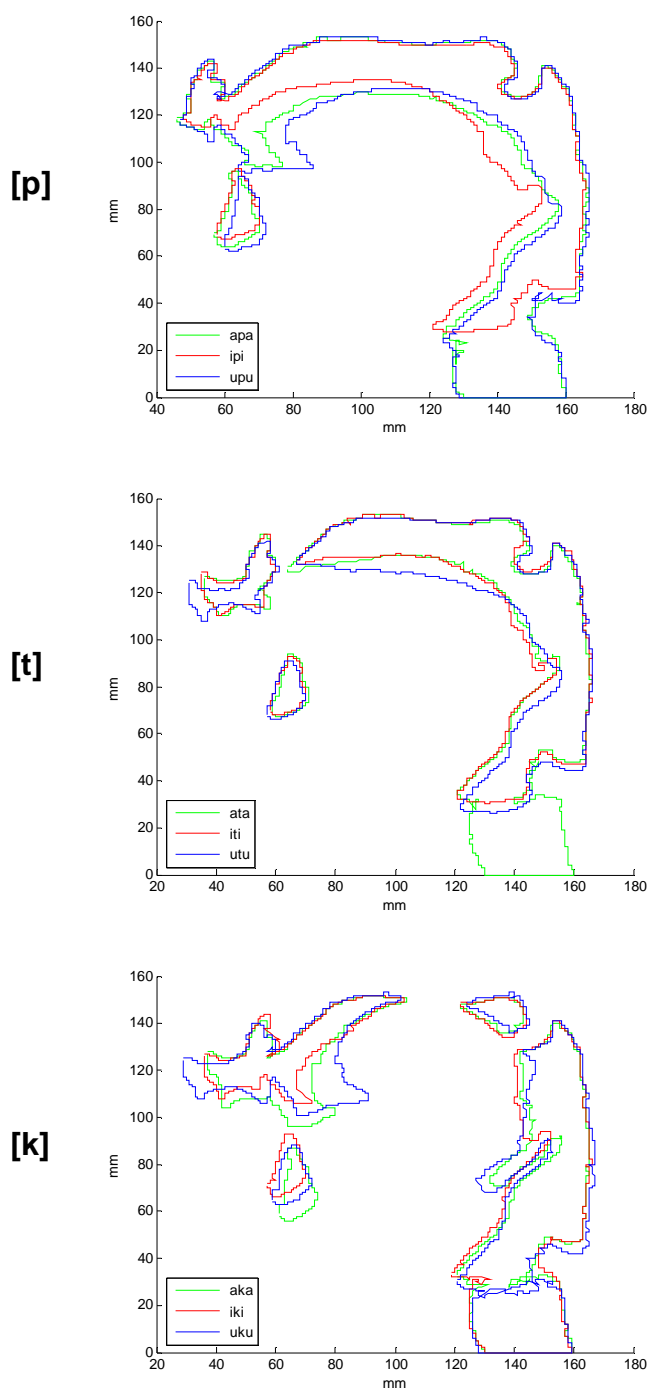


Figura 7. 30 - Sobreposição dos contornos obtidos para as oclusivas surdas [p] em cima, [t] ao centro e [k] em baixo produzidas em contexto simétrico com as vogais [a],[i] e [u], para o falante PAA.

Oclusivas Sonoras [b d g]

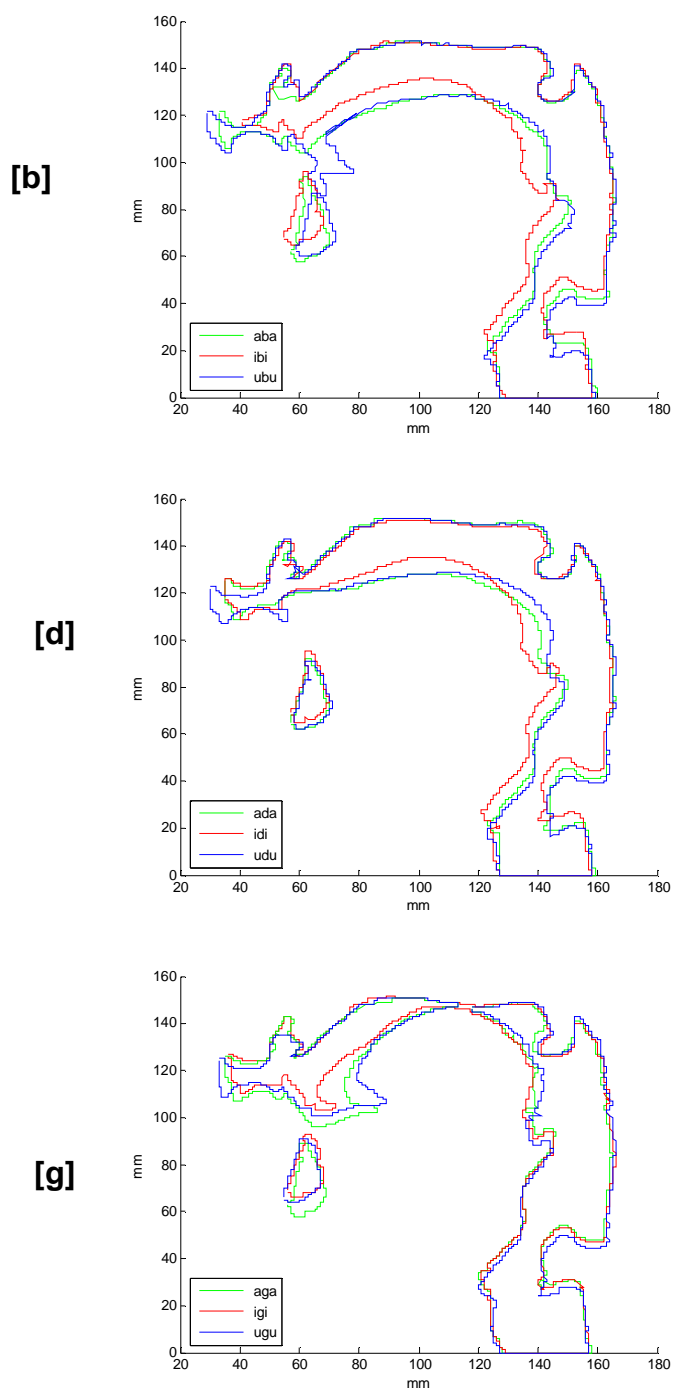


Figura 7. 31 - Sobreposição dos contornos obtidos para as oclusivas surdas [b] em cima, [d] ao centro e [g] em baixo produzidas em contexto simétrico com as vogais [a],[i] e [u], para o falante PAA.

Da análise dos contornos, podemos constatar que o efeito de coarticulação parece ser evidente nas oclusivas. No caso da consoante velar [k], as diferenças são mais significativas na zona da ponta da língua, uma vez que este articulador (ponta da língua) fica mais livre para a produção da

vogal, enquanto que a zona posterior do dorso, está já posicionada para a produção da consoante. Algo similar verifica-se para o [t]: a zona mais estável (a que sofre menos influência) é junto ao ponto de articulação da consoante (neste caso, dental), enquanto que o dorso da língua se posiciona para a produção da respectiva vogal. No caso do [p], toda a língua fica livre para a produção da vogal, já que a articulação do [p] é bilabial (os lábios é que estão envolvidos na oclusão).

O comportamento das oclusivas sonoras, em termos da influência da vogal adjacente, tendo em atenção o ponto de articulação, segue um padrão semelhante ao das surdas, como se pode depreender pela análise da Figura 7.31.

7.2.2 Coarticulação nas fricativas

À semelhança do que já se efectuou para as oclusivas também no caso das fricativas pretendemos observar qual o efeito do contexto (vogal) na produção destes sons. No corpus escolhido o contexto é sempre simétrico ($V_1=V_2$), do tipo $V_1C_fV_2$ com as vogais extremas [i a u]. Na análise a efectuar, teremos essencialmente em atenção, a posição da língua e da mandíbula (coarticulação lingual e mandibular).

Fricativas surdas [f s S]

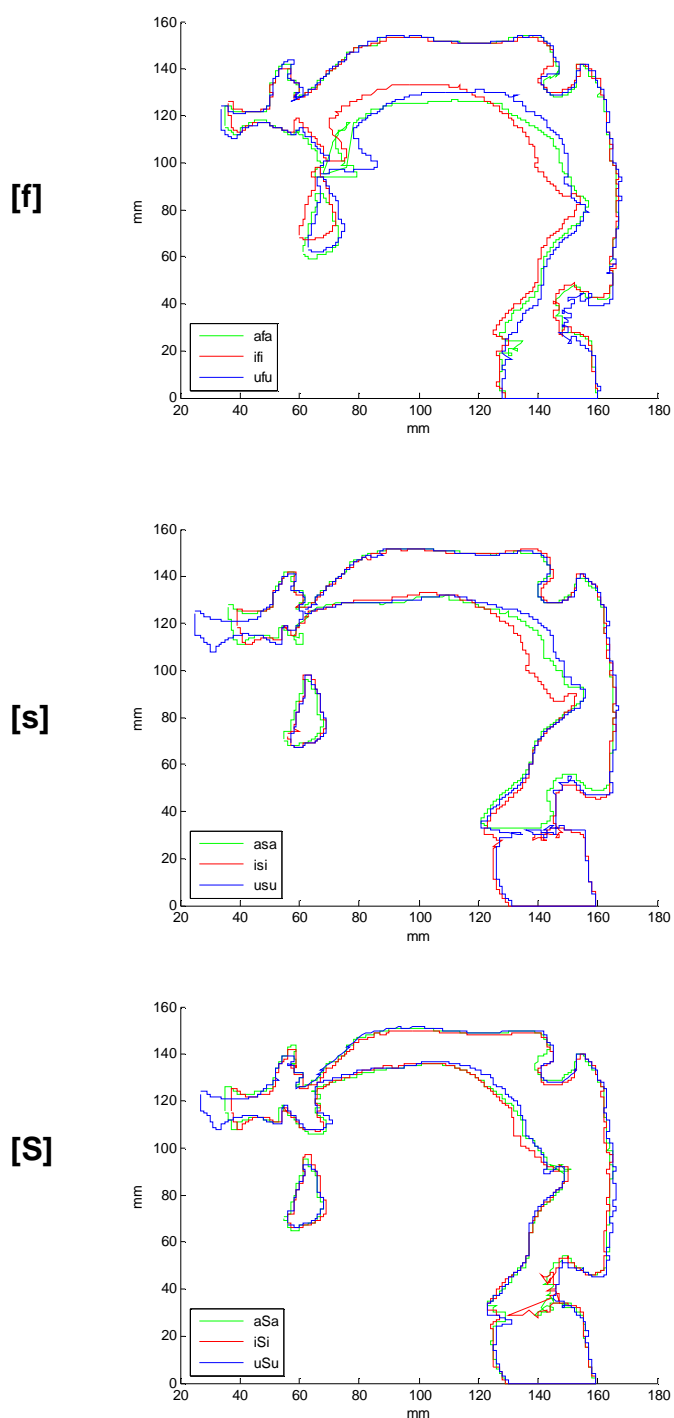


Figura 7. 32 - Sobreposição dos contornos obtidos para as fricativas surdas [f] em cima, [s] ao centro e [S] em baixo, produzidas em contexto simétrico com as vogais [a], [i] e [u], para o falante PAA.

Mais uma vez o efeito de coarticulação ao nível da língua é evidente mas em menor grau do que o observado nas oclusivas, nos contextos estudados. Para o [f], que é uma fricativa lábio-dental verificam-se diferenças quer a nível da coroa, quer do dorso da língua na produção das diferentes vogais. Na realidade, na produção deste som a língua fica relativamente livre para a produção da vogal, pelo que a altura da mesma pode variar em função da altura da vogal (Figura 7.32).

No caso do [s], apenas existem diferenças dignas de menção na porção posterior do dorso da língua, não se verificando o efeito da vogal na parte da coroa da língua nem na porção anterior do dorso, uma vez que esta zona da língua é necessária para a produção da consoante (denti-alveolares).

Relativamente ao [S] o efeito da vogal na língua, é ainda menor, podendo significar, que este som comparativamente aos outros incluídos neste estudo, apresenta uma resistência à coarticulação bastante grande. No que diz respeito à posição dos lábios, verifica-se que há uma ligeira protusão labial, quando a vogal é arredondada (no caso do [u]), nas fricativas [s] e [S] mas que esse efeito coarticulatório não se observa, mesmo no contexto da vogal [u] para a fricativa [f], uma vez que o lábio inferior é o articulador activo não permitindo o efeito de coarticulação.

Fricativas sonoras [v z Z]

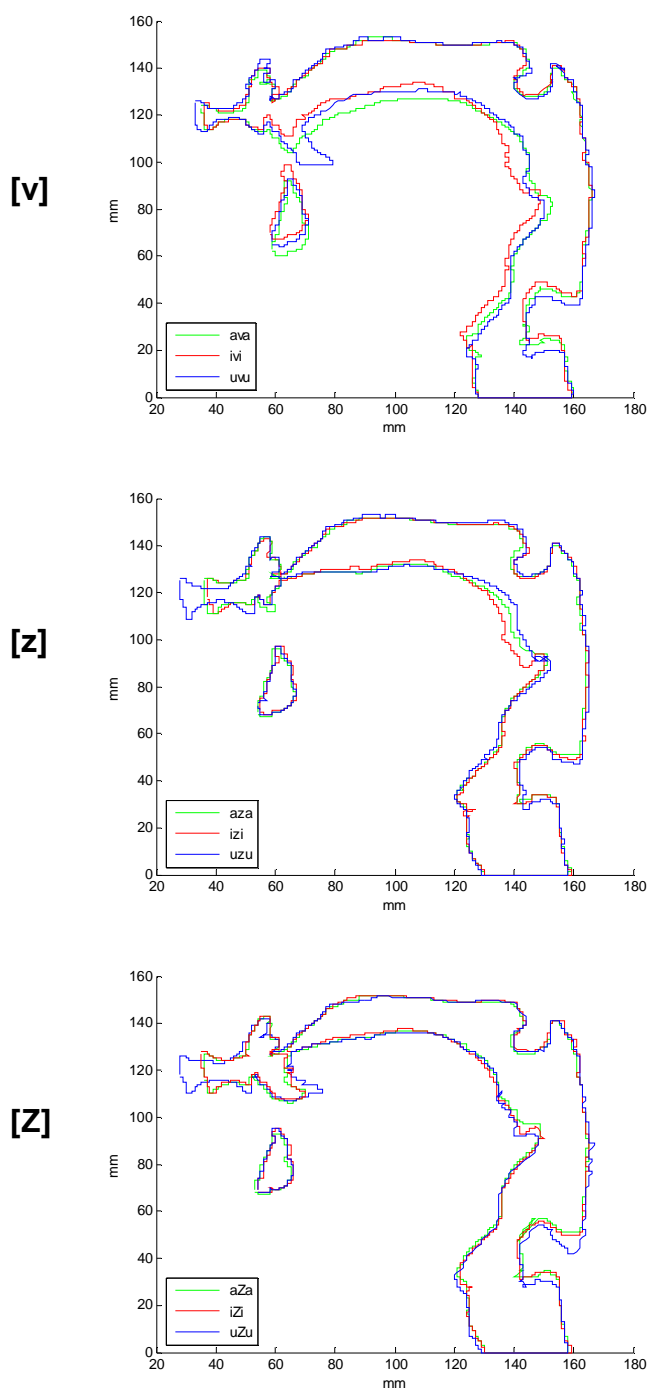


Figura 7. 33 - Sobreposição dos contornos obtidos para as fricativas sonoras [v] em cima, [z] ao centro e [Z] em baixo, produzidas em contexto simétrico com as vogais [a], [i] e [u].

Pela análise da Figura 7.33, podemos observar que para as fricativas vozeadas o padrão de influência da vogal, na produção da fricativa é semelhante ao observado para as não vozeadas, sendo maior na lábio-dental [v], menor na dental ou ou alveolar [z] e, também praticamente imune

ao efeito da vogal, na produção do [Z] no que concerne à coarticulação lingual. No que diz respeito á protusão labial, também aqui, há protusão quando a vogal é arredondada, para o [z] e [Z] mas não para o [v].

7.2.3 Discussão coarticulação

Os resultados obtidos neste estudo, relativos à coarticulação, são bastante encorajadores uma vez que foi possível verificar, para o PE e para este falante, alguns factos já reportados, por outros autores, relativos a efeitos coarticulatórios para outras línguas. Estes resultados não deixam de nos surpreender positivamente, uma vez que, a informação obtida neste estudo refere-se apenas a um informante, apenas com uma repetição de cada item e obtido através de imagens de RM, quando na maior parte dos outros estudos consultados e que fazemos referência, os dados foram obtidos através de EMA, EPG e estudos espectrográficos. Excepção feita para Engwall (2000) que também apresenta um estudo RM, na avaliação de coarticulação em fricativas, para a língua Sueca. Da revisão efectuada relativa a este tema e enquadrando os nossos resultados podemos referir que:

No geral, no nosso estudo verificou-se que as oclusivas se mostraram menos resistentes à coarticulação do que as fricativas. Isto está de acordo com o descrito por Farnetani (1999) e Farnetani & Recasens (1999) que referem que de todos os sons consonânticos os menos resistentes à coarticulação são as oclusivas, seguidas pelas nasais, laterais e as fricativas as mais resistentes;

Recasens (1999) refere que uma determinada zona da língua pode apresentar comportamento articulatorio diferente em função do seu envolvimento na produção de uma determinada configuração. Assim, a **nível do dorso da língua**, em particular, e tendo em atenção o ponto de articulação, as fricativas apresentam menos efeitos de coarticulação ao nível do dorso, quando comparado com as oclusivas e as laterais da mesma classe ou classe semelhante, o que também se verificou no nosso estudo. Comparando a fricativa lábio-dental [f] com a oclusiva bilabial [p] observa-se que o efeito da vogal adjacente é superior na oclusiva do que na fricativa de classe correspondente. No entanto, esta diferença é ainda mais pronunciada quando se compara por exemplo a fricativa denti-alveolar [z] com a oclusiva de classe semelhante, o [t];

Também está descrito (Recasens, 1999), que a lâmina deva ser mais resistente à coarticulação durante a produção de consoantes denti-alveolares [t d s] do que nas velares [k g], uma vez que a lâmina participa activamente na produção das primeiras e não das segundas. No que diz respeito à região da lâmina lingual, no nosso estudo verificou-se que nas oclusivas [t d] e na fricativa [s], nesta zona da língua, o efeito da vogal é negligível sendo perfeitamente evidente uma maior influência (variação) na produção das oclusivas [k] e [g];

O acoplamento entre a ponta e o dorso da língua nas consoantes denti-alveolares não é suficientemente forte para prevenir coarticulação no dorso da língua. Desta maneira, a altura do dorso altera-se em função da vogal adjacente, ou seja, é maior no [i] que é uma vogal alta e menor no [a] que é uma vogal mais baixa (Recasens, 1999 pág 89). No corpus estudado, verificou-se que de facto nestas consoantes [t d s z] apesar de não se verificarem alterações ao nível da lâmina lingual verifica-se essa dependência ao nível do dorso, sendo que para as oclusivas [t] e [d] o efeito da vogal se verifica ao longo de todo o dorso lingual enquanto que nas fricativas [s] e [z] esse efeito é mais localizado na porção posterior do dorso ou seja na porção situada abaixo do velo. Nesta porção da língua verifica-se que esta é mais posterior em [u] que é uma vogal posterior do que em [a] e [i] que são média e anterior, respectivamente;

As consoantes pós-alveolares, palato-alveolares e palatais de que no nosso corpus temos apenas os sons [S] e [Z], são referidas como menos sensíveis à coarticulação na zona do dorso, quando comparados com as denti-alveolares, o que também se verificou no nosso estudo e que está em consonância com o maior efeito de acoplamento com o articulador primário referido em Farnetani & Recasens (1999); Num estudo efectuado com EPG, e referido em Farnetani & Recasens (1999), de todos os sons estudados, o [S] praticamente não variou (a percentagem de contacto língua palato na EPG) no contexto das duas vogais [i a], demonstrando ser mais resistente à coarticulação.

Quando se consideram as zonas da língua que não estão envolvidas directamente na constrição verifica-se que o efeito da vogal na actividade do dorso da língua é maior nas bilabiais [p b] e fricativas lábio-dentais [f v] do que nas consoantes em que a ponta da língua, é o articulador primário uma vez que esta não interfere na produção das primeiras (Recasens, 1999);

De acordo com Kiritani (1983), referenciado por Fletcher & Harrington (1999) quando se considera o sistema língua - mandíbula em conjunto, um exemplo paradigmático deste tipo de coarticulação é a anteriorização das consoantes velares no contexto de vogais altas e anteriores. De facto, se observarmos o [k] e o [g] nas Figuras 7.30 e 7.31, verifica-se que, no contexto da vogal [i] a zona da lâmina da língua está numa posição mais anterior do que para [a] e [u] não sendo esta anteriorização tão evidente a nível da mandíbula. Tuller (1981) e Fletcher & Harrington (1999), referem que: a altura da mandíbula não se altera em contexto VCV, para o [t] e [f] mas sofre alteração em função da vogal no [p] e no [k]. No nosso corpus, verificou-se que para o [t] não há alteração na altura da mandíbula, mas que esta se verifica na produção do [f]. Para o [k] e [p] observa-se efeito de coarticulação tal como descrito por (Tuller, 1981);

De entre todos os sons, que estudámos e não considerando agora nenhum articulador em particular, os que apresentam uma maior resistência a efeitos de coarticulação são os sons [S] e [Z]. Este facto, descrito por Farnetani (1999) e Farnetani & Recasens (1999) poderá estar relacionado com a complexidade envolvida na produção destes sons, já referida por Hardcastle (1976). Também Recasens (1983, 1987) refere que alguns sons apresentam um grau de limitação ou de impedimento articulatorio (DAC) bastante elevado e que poderá estar relacionada com a complexidade envolvida na sua produção.

As diferenças no grau de coarticulação apresentado pelas diferentes consoantes podem ser explicadas por impedimentos específicos impostos pelo corpo da língua na produção das consoantes. As fricativas, por exemplo terão que limitar a posição ao nível do dorso da língua, de forma a assegurar a constricção anterior e a pressão intra-oral necessária à produção do ruído que lhes é característico. A produção de laterais e oclusivas não impõe tantos constrangimentos permitindo uma maior gama de variações coarticulatórias (Farnetani & Recasens, 1999). Neste estudo, como já se referiu, verificou-se também um maior efeito coarticulatório nas oclusivas do que nas fricativas. As laterais não foram estudadas em contexto.

Apesar destes resultados serem próximos dos obtidos por outros autores, não nos permitem uma generalização, uma vez que o nosso corpus inclui apenas apenas um falante e uma só produção de cada som. Foram, no entanto, muito úteis para testar e validar o método de aquisição e a robustez do método de segmentação quando se pretenda estudar coarticulação (Carbone, Martins, Teixeira & Silva, 2007) e que poderá ser utilizado num corpus mais alargado, mas vocacionado para estudar coarticulação.

Obtenção das distâncias médio-sagittais a partir das imagens 2D sagittais para as vogais [a], [i] e [u].

Um dos objectivos relativos ao Corpus 2D, era a obtenção das distâncias médio sagittais para todos os fones estudados, o que foi efectuado. No entanto, no contexto da presente dissertação limitámos a apresentação destes dados às vogais cardinais, do PE.

Assim, apresentam-se de seguida as imagens obtidas no plano sagittal, para as vogais orais [a], [i] e [u], com sobreposição das grelhas adaptativas, que permitiram a obtenção das distâncias médio-sagittais (ver Figura 7.34).

Considerou-se a glote como o ponto zero (início do tracto vocal) sendo as distâncias apresentadas (em mm) a partir deste ponto.

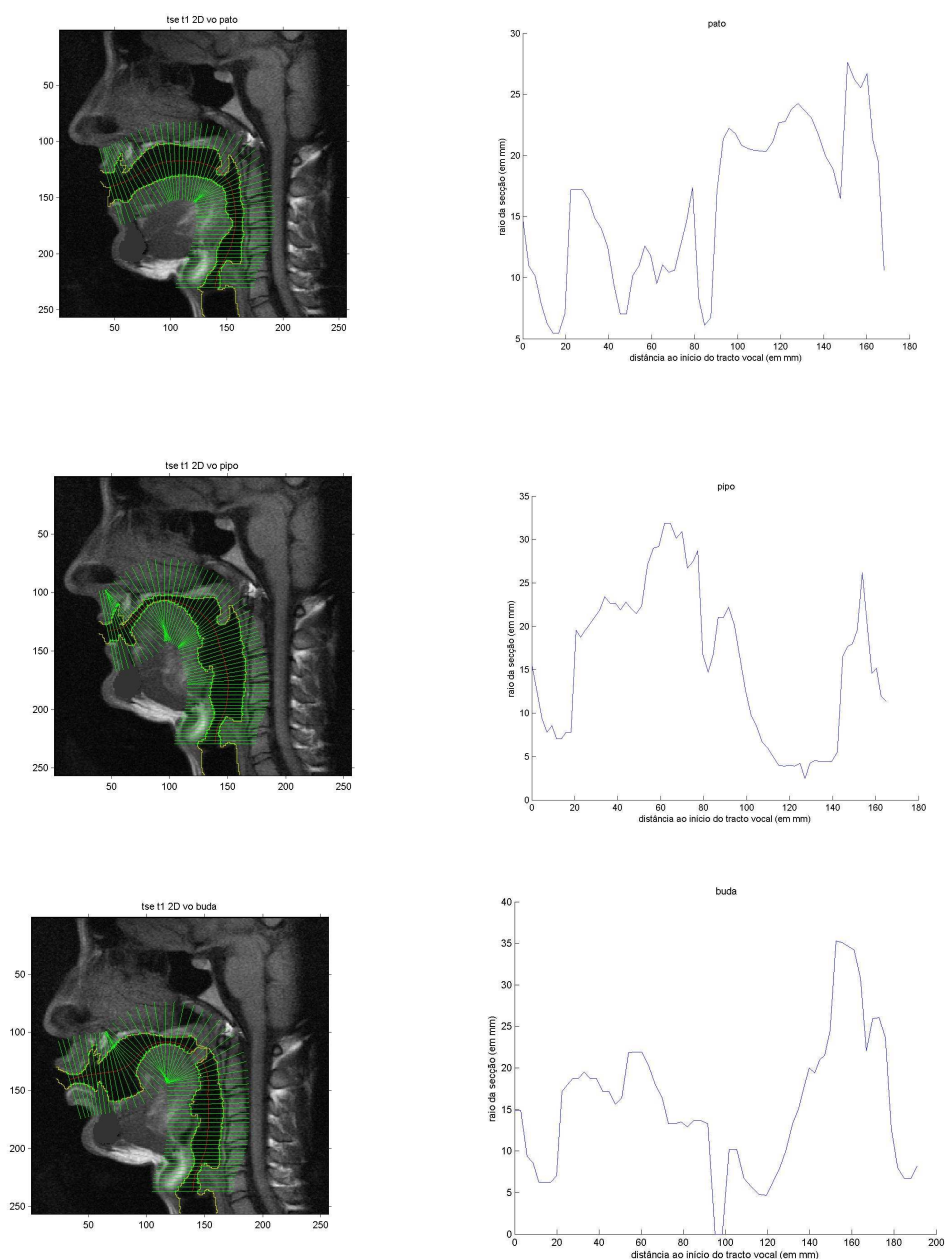


Figura 7.34 - Distâncias médio-sagitais obtidas, a partir da aplicação de grelha adaptativa sobre as imagens, referentes às vogais [a] em cima, [i] ao centro e [u] em baixo.

Como se pode verificar, foi possível obter as distâncias médio sagitais e obter se tal for necessário dados quantitativos relativos à dimensão do tracto vocal a nível da linha média para todos os fones do corpus 2D.

7.3 Corpus 3D

A partir do corpus 3D esperávamos poder obter informação que permitisse a modelação tridimensional do tracto vocal nomeadamente a obtenção das funções de área do tracto vocal, de uma forma directa, para um vasto número de sons do PE. Era também objectivo complementar a informação obtida com o corpus 2D, para algumas classes de sons.

Como foi referido, no capítulo referente às metodologias, propusemos um método de aquisição diferente do que é frequentemente utilizado no âmbito dos estudos de produção de fala, quando o objectivo principal é modelar em 3D. Para se ter uma primeira informação relativa à validação do método, é fundamental que seja possível retirar informação, se não superior, pelo menos equivalente à obtida através do método de referência.

Para o corpus 3D, ao invés de apresentarmos toda a informação, relativa a todos os sons adquiridos, optámos por apresentar apenas algumas das suas potencialidades, dando exemplos do tipo de informação que pode ser extraída da base de dados. Assim, e no que diz respeito às funções de área, apresentamos, como exemplo, os resultados obtidos para as vogais orais [a], [i] e [u], aproveitando para efectuar a comparação entre os dois métodos de aquisição (método clássico e o método de aquisição 3D) e também as funções de área para as laterais e fricativas surdas em contexto.

Como uma das áreas de interesse do Projecto onde este trabalho se enquadra é o estudo das nasais, apresentamos com mais detalhe alguns dos parâmetros extraídos das imagens, após segmentação e que para além das funções de área (não incluídas no contexto desta dissertação), incluem a área da cavidade oral, nasal e faríngea na zona de acoplamento entre o tracto nasal e oral e o coeficiente de abertura do velo (CAV) para os sons nasais e também para as vogais orais para comparação.

7.3.1 Funções de área do tracto vocal

7.3.1.1 Visualização 3D e funções de área do tracto vocal obtidas através do método clássico

Na Figura 7.35, apresentamos visualização 3D (à esquerda) e as funções de área do tracto vocal (à direita), obtidas de uma forma directa, a partir da informação proveniente dos cortes coronais, oblíquos e transversais adquiridos com sequência TSE ponderada em DP, de forma semelhante ao efectuado noutros estudos já referenciados anteriormente neste trabalho. Na porção superior da figura apresenta-se informação relativa ao [i], ao centro [a] e em baixo [u].

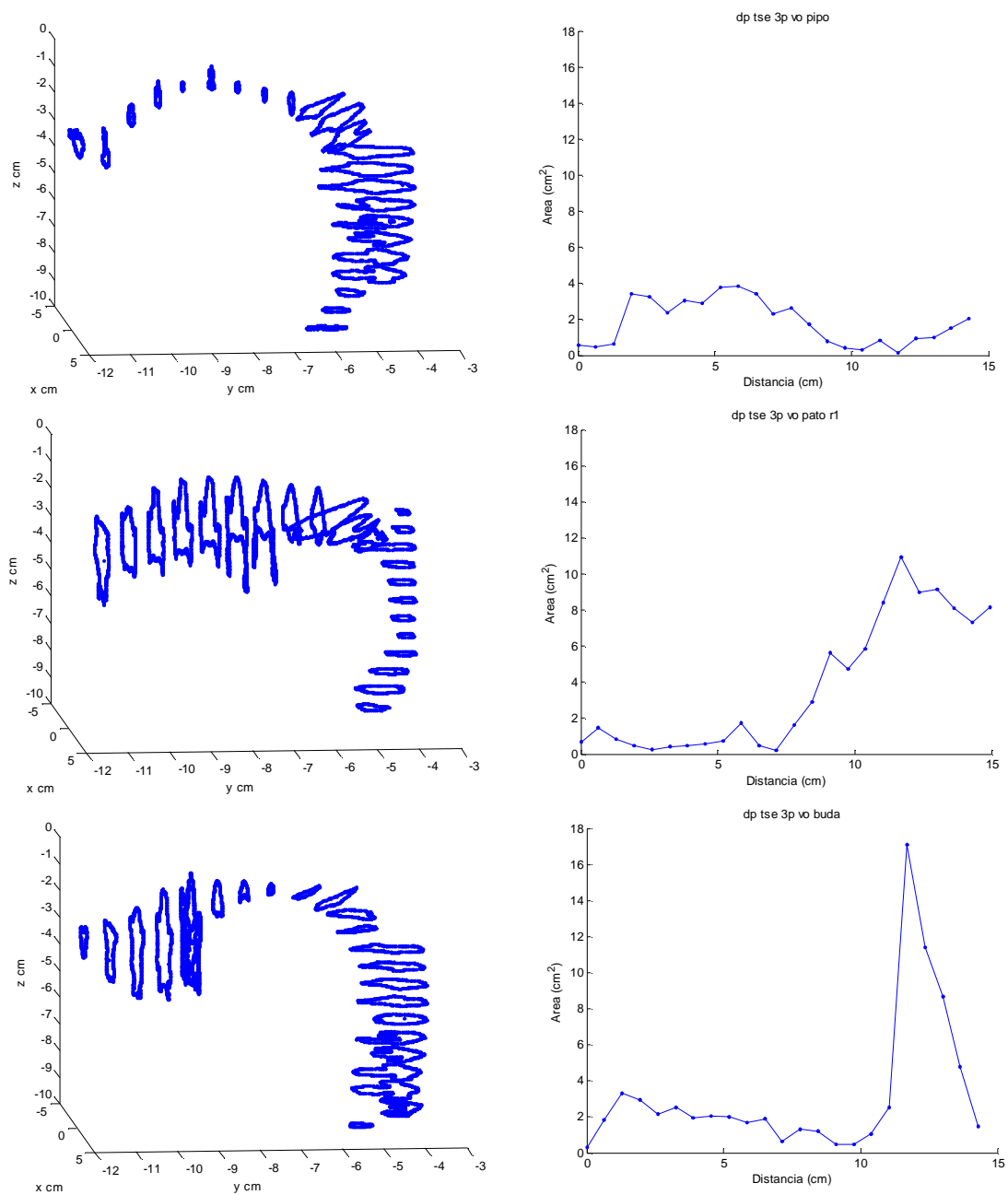


Figura 7. 35 - Funções de área obtidas para as vogais [i] em cima, [a] ao centro e [u] em baixo através da aquisição clássica (áreas transversais relativas ao tracto vocal obtidas a partir de cortes efectuados nos planos coronal, oblíquo e transversal).

7.3.1.2 Corpus 3D (Aquisição em volume)

De forma semelhante, mostram-se na na Figura 7.36, as visualizações 3D (à esquerda) e funções de área do tracto vocal obtidas para o [i] em cima, [a] (ao centro) e, [u] em baixo, obtidas a partir do método por nós proposto neste estudo.

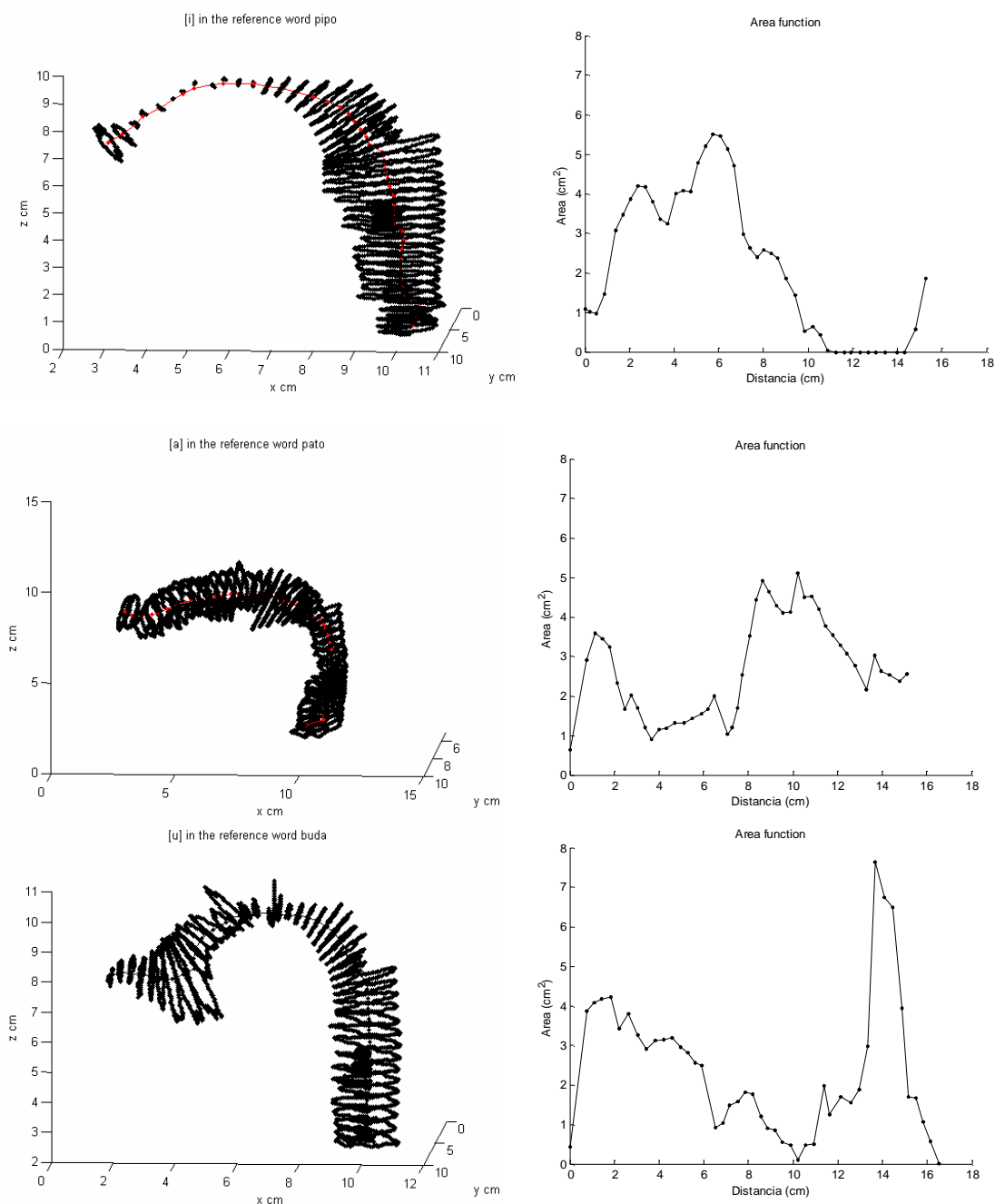


Figura 7. 36 - Funções de área obtidas para as vogais [i] em cima, [a] ao centro e [u] em baixo, através da aquisição volumétrica (método proposto no nosso estudo).

7.3.1.3 Comparação dos dois métodos: Clássico versus 3D

Na Figura 7.37, pode-se observar a sobreposição das curvas relativas às funções de área calculadas a partir dos dois métodos de aquisição para o [a] (à esquerda), o [i] (ao centro) e [u] (à direita). O zero corresponde à região da glote. As duas curvas foram alinhadas tendo como referência os seus máximos.

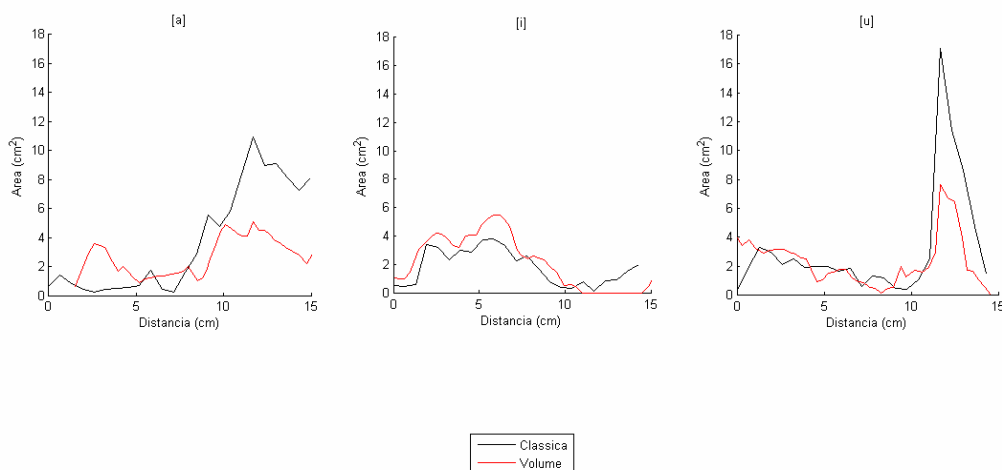


Figura 7. 37 - Comparação das áreas obtidas através dos dois métodos para as vogais cardinais do PE, [a] à direita, [i] ao centro e [u] à esquerda, por aquisição volumétrica (a vermelho) e clássica (a preto).

Verifica-se que, para o [i] não existem diferenças relevantes obtidas através dos dois métodos. No entanto, para o [a] e [u] existem algumas diferenças, particularmente a nível da cavidade oral, onde as áreas são significativamente superiores com a aquisição clássica, do que as obtidas com a aquisição em volume. Independentemente do método, a cavidade oral é uma zona de difícil segmentação, uma vez que, não são claros os seus limites, particularmente na zonas laterais e frontais, pela não visualização das arcadas dentárias. A análise da informação sugere uma modelação de espaçamento mais uniforme ao longo do tracto vocal, na aquisição volumétrica, o que está relacionado coma possibilidade de reconstruir cortes mais finos e mais próximos, do que o conseguido com o método clássico (neste caso a espessura de corte foi 5 mm).

7.3.1.4 Funções de área para as Vogais Orais [i], [u], [a]

Em relação às áreas obtidas, e reportando-nos agora à análise da Figura 7.36, podemos observar o seguinte:

As áreas na cavidade faríngea são superiores para a vogal [i] quando comparadas com o [a] e [u], atingindo um máximo de 6 cm^2 , a cerca de 7cm da glote, na zona da orofaringe. Há depois uma diminuição progressiva das áreas ao longo da cavidade oral até atingir um mínimo, na parte anterior desta cavidade, uma vez que há uma elevação e anteriorização da língua para a produção do [i], uma vogal alta e anterior. Após este ponto de maior aproximação da língua com o palato as áreas aumentam novamente até à zona dos lábios.

Para o [a], podemos observar que as áreas são mais uniformes ao longo do tracto vocal, existindo um aumento da área na cavidade faríngea (atingindo cerca de 4 cm^2) até à parte superior da epiglote, diminuindo depois e mantendo-se aproximadamente constante na zona da orofaringe. Há depois um aumento progressivo na zona da cavidade oral até aos 10 cm (a contar da glote) e posteriormente uma diminuição progressiva até à zona dos lábios. Comparativamente ao [i] e [u] a língua está mais baixa e central no [a], o que se enquadra nas classificações articulatórias e acústicas para esta vogal. As áreas, na cavidade oral, são maiores relativamente ao [i] e [u] sendo maior a abertura da cavidade oral nesta vogal (considerada uma vogal aberta).

Finalmente, para o [u] verifica-se que a área a nível da cavidade faríngea atinge um valor de aproximadamente 4 cm^2 , logo acima da glote, diminuindo progressivamente até uma zona localizada a 10 cm da glote (área de maior constrição). Depois desse ponto de maior constrição a área aumenta ligeiramente, atingindo depois um pico estreito (8 cm^2), relacionado com o recuo e abaixamento da ponta da língua, na produção desta vogal. O [u] é uma vogal posterior e alta, como de resto se pode observar pela posição da língua na visualização 3D. Depois do aumento da área na zona da ponta da língua, há uma diminuição brusca até aos lábios. Podemos observar que o comprimento do tracto vocal é maior (16,5 cm) para o [u], do que para os outros sons, pois é uma vogal produzida com protusão labial (arredondada).

7.3.1.5 Funções de área para as Laterais

Apresentamos agora as funções de área do tracto vocal para as laterais do PE (Figuras 7.38 e 7.39). No corpus 3D, obteve-se informação relativa a três contextos diferentes para o [l]. Em cima: à esquerda na Figura 7.38 temos o [l] em início de sílaba e de palavra como em [lasu], ao centro o alofone [l~], em final de sílaba como em [mal~] e, finalmente à direita, o [l] em início de sílaba, mas em posição intervocálica. Em baixo, apresenta-se a sobreposição das três funções de área do tracto vocal.

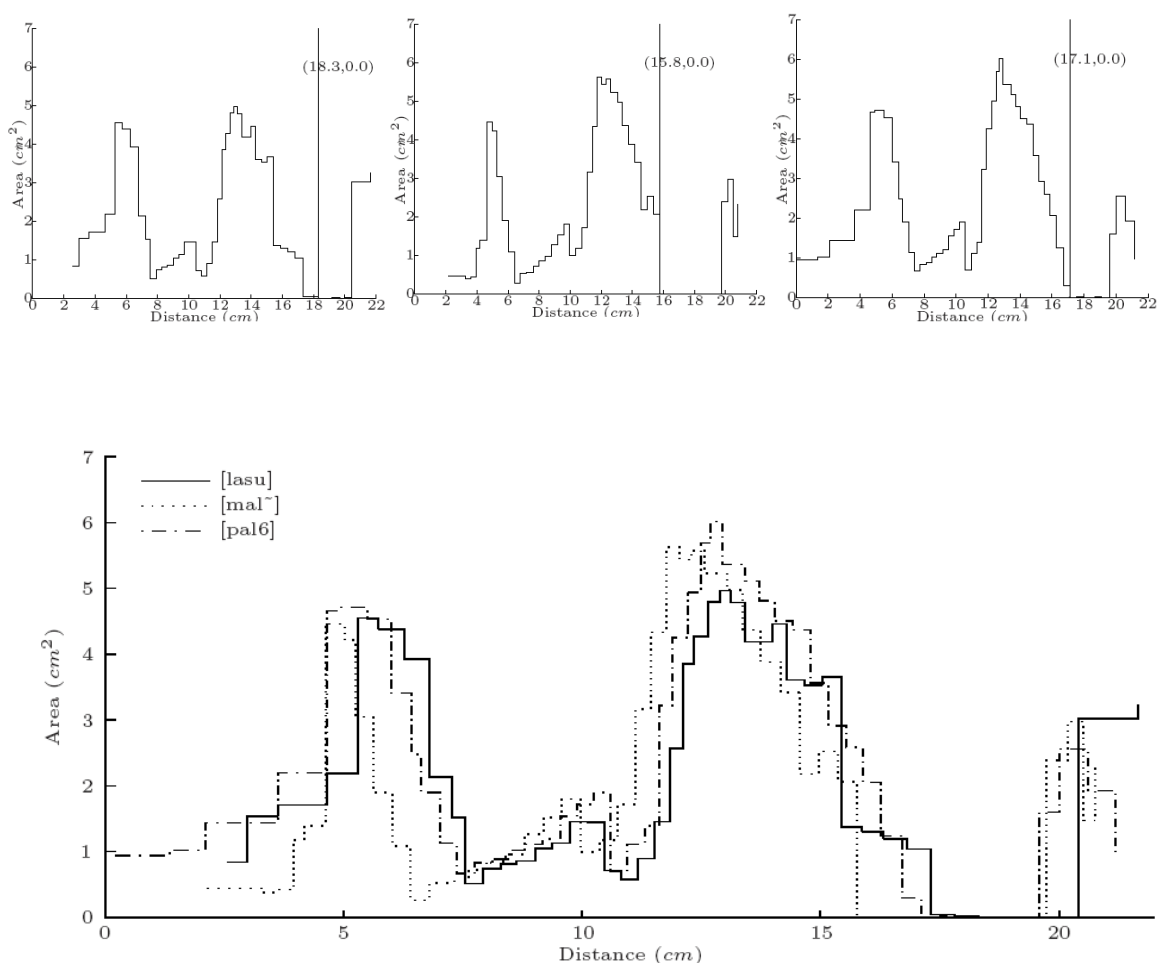


Figura 7.38 - Funções de área do tracto vocal para três contextos do [l] do PE (em cima): à esquerda em início de sílaba e palavra [lasu], ao centro em final de sílaba [mal~], e à direita em início de sílaba, mas intervocálico [pal6]. Em baixo, a sobreposição das três funções de área.

Analisando a Figura 7.38, podemos confirmar e complementar a informação obtida para estes sons no corpus 2D. Observando em particular a sobreposição das três funções de área, podemos verificar que o seu padrão de variação é muito próximo. Assim, depois de um aumento das áreas

na cavidade faríngea (com um máximo a 5-6 cm da glote) existe uma diminuição das áreas, para qualquer uma das diferentes posições do [l] na sílaba, a cerca de 10 cm da glote e que corresponderá à zona do velo. Depois as áreas aumentam até se atingir o máximo de 6 cm², a 13 cm da glote. As áreas decrescem novamente até um valor mínimo (a cerca de 17 cm da glote) que corresponderá ao ponto de articulação na zona alveolar. As áreas aumentam depois até aos lábios. Ou seja, tal como já se tinha verificado nas imagens sagitais, pode existir velarização também em início de sílaba, uma vez que há uma segunda constrição na zona do velo para qualquer dos sons. Estes resultados corroboram os de outros autores como Andrade (1997, 1999) e descrições antigas relativas ao PE (Strevens, 1954).

Na figura 7.39 surge a função de área relativa à lateral palatal [L]. Podemos observar que na cavidade faríngea as áreas são elevadas, cerca de 7 cm² (aos 10-11 cm) e que começam a decrescer progressivamente na cavidade oral, até à zona da constrição, bastante anterior (aos 15 cm) e que se mantém por uma extensão relativamente grande (3,5 cm) na zona anterior do palato, transição palato-alveolar e zona alveolar.

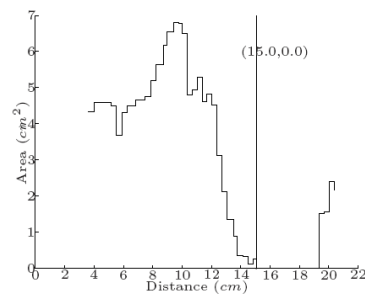


Figura 7. 39– Função de área para a lateral [L].

7.3.1.6 – Funções de área obtidas para as fricativas surdas, em contexto

Apresentam-se de seguida os resultados obtidos para as funções de área das fricativas surdas, obtidas em contexto VCV no corpus 3D.

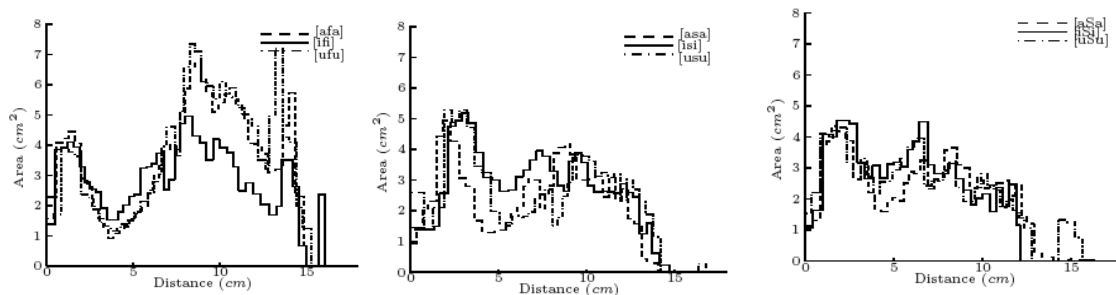


Figura 7. 40 – Funções de área para as fricativas surdas [f] à esquerda, [s] ao centro e [ʃ] à direita obtidas em contexto com as 3 vogais [a], [i] e [u], no corpus 3D.

Como se depreende da análise da Figura 7.40, verificamos que os efeitos coarticulatórios observados no corpus 2D, se mantém a nível da informação 3D, nas funções de área. Assim, na produção do [f], é evidente o efeito coarticulatório, essencialmente a nível da cavidade oral na zona de todo o dorso da língua, em particular na porção mais anterior. Dos três sons este é o que permite maior efeito coarticulatório e verifica-se que as áreas são menores no contexto do [i], pois na produção desta vogal o dorso da língua está mais elevado, condicionando uma menor área. Nesta porção anterior da cavidade oral, as áreas para o [a] e [u] são superiores pois a língua está numa posição mais baixa, na produção destas vogais. Na zona da constrição o efeito coarticulatório não é visível. Para o [s], o efeito é no geral, menos evidente, sendo as diferenças agora mais visíveis, na zona da faringe, onde as áreas são maiores no contexto do [i], pois a língua avança na produção deste som. A fricativa palato-alveolar [ʃ], é a que apresenta mais resistência à coarticulação, não se observando diferenças com significado nas áreas obtidas para os três contextos, confirmando os resultados obtidos com as imagens do corpus 2D.

7.3.2 Consoantes nasais e vogais do PE

Como um dos objectivos deste estudo era a obtenção de dados relativos à produção de sons nasais, e dada a riqueza desses sons para o PE, foram calculados alguns parâmetros que se consideram importantes na modelação de sons nasais, como por exemplo, a área de abertura nasal, a proporção entre abertura nasal e oral (o coeficiente de abertura do velo - CAV ou em Inglês VPOQ) e a área da cavidade faríngea, em particular as áreas na zona da oro-faringe.

Assim, na Tabela 7.1, podemos observar as áreas relativas ao grau de abertura nasal e abertura oral, para as vogais nasais, consoantes nasais e também as respectivas áreas para as vogais orais. Como se referiu no capítulo dois da presente dissertação, pelo facto de haver abertura ou acoplamento nasal não quer dizer que o som seja percebido como nasal. Um factor importante para a distinção nasal/oral é também, a proporção entre as duas áreas. Para estabelecer paralelo e permitir comparações foram incluídas também as vogais orais, como já referido.

Nas Figuras 7.41 a 7.43 podemos observar os cortes com orientação perpendicular ao tracto vocal, após aplicação de grelha adaptativa, que foram segmentados pela técnica de *Live-Wire*, para obtenção das áreas nasais e orais. O corte segmentado é o primeiro corte onde se observam duas passagens, uma para a cavidade oral, outra para a cavidade nasal, procedimento referido por Demolin et al.(1998), Demolin et al. (2003) e Engwall et al. (2006) e efectuado pelos autores para as vogais do Francês. Além da área para a abertura nasal é também observável a proporção nasal/oral que é diferente entre as diversas classes de sons e dentro da mesma classe, entre os diferentes sons.

7.3.2.1 Áreas da abertura nasal e oral

	FONE	ABE. ORAL (cm²)	ABE. NASAL (cm²)
<i>Vogal nasal</i>	i~ [pi~tu]	2.35	1.22
	e~ [pe~t@]	2.60	1.60
	6~ [ka~tu]	1.29	1.32
	o~ [po~t@]	3.74	4.16
	u~ [pu~tu]	2.94	2.47
<i>Vogal oral</i>	i [pipu]	3.89	0.71
	e [pek6]	2.33	0.72
	a [patu]	1.61	0.70
	6 [k6d6]	1.79	0.61
	o [topu]	1.22	0.63
	O [pot@]	1.57	0.57
	u [bud6]	1.43	1.03
<i>Consoante nasal</i>	m [k6m6]	2.57	2.35
	n [k6n6]	2.77	2.43
	J [k6J6]	2.57	1.65

Tabela 7. 1 - Áreas da abertura nasal e oral para vogais nasais, vogais orais e consoantes nasais.

Pode verificar-se pela análise da Tabela 7.1, que a abertura nasal é sempre superior nas vogais nasais do que nas correspondentes vogais orais. Apesar da abertura nasal ser mais pequena nas vogais orais ela nunca é zero, ou seja o velo nunca se encontra completamente fechado na produção das vogais orais. A vogal oral que apresenta uma maior abertura nasal é o [u], os mais baixos o [O] e [o]. As vogais anteriores e médias apresentam valores intermédios e muito próximos (aprox. 0,70). De entre as vogais nasais, as que apresentam valores maiores de abertura nasal são as vogais posteriores [u~] e [o~] sendo o [o~] o que apresenta a maior abertura nasal (4,16 cm²). Para as vogais nasais verifica-se a sequência [o~] > [u~] > [e~] > [6~] > [i~]. Podemos dizer que as vogais posteriores apresentam valores superiores aos das vogais médias e anteriores. O [6~] aqui considerado, foi o som sustido no contexto da palavra [k6~tu], já que o mesmo fone mas quando produzido tendo por referência a palavra [m6~tu] apresenta um valor mais elevado e bastante próximo dos valores obtidos para as consoante nasais [m] e [n]. Este facto, poderá ser justificado pelo facto de o [6~] neste caso ser precedido pela consoante nasal [m]. As consoantes nasais, apresentam também valores elevados, de um modo geral superiores aos das vogais nasais.

Analisando agora as imagens (Figuras 7.41 a 7.43) podemos ainda verificar que na produção das vogais nasais posteriores [o~] e [u~], a língua está mais subida, praticamente encosta ao velo, no entanto, a área de abertura oral ainda, é grande. Relativamente às vogais orais, verifica-se para o [O] e principalmente para o [o], que a língua se encosta ao velo (principalmente à esquerda), não

obstruindo o tracto, mas fazendo diminuir a área da abertura oral, na zona de acoplamento. Estes resultados corroboram o que se observou nas imagens 2D na linha média, mas que só podemos confirmar agora, com os cortes axiais.

Vogais Nasais

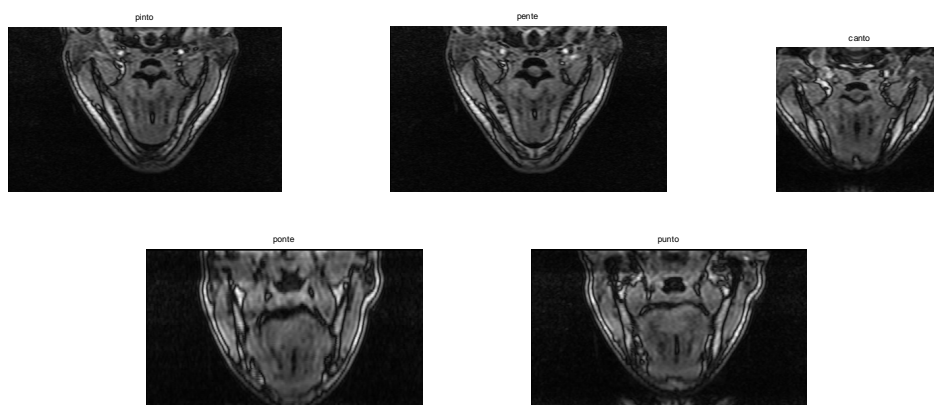


Figura 7. 41 - Corte perpendicular ao tracto vocal, na zona de acoplamento dos tractos oral e nasal, para as vogais nasais do PE (falante PAA).

Vogais Orais

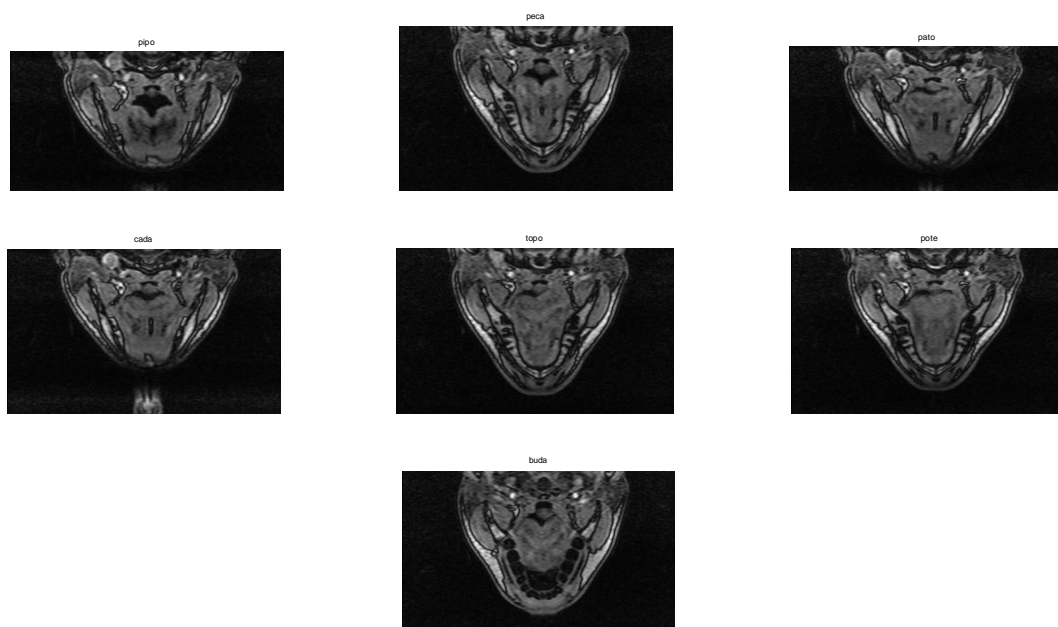


Figura 7. 42 - Corte perpendicular ao tracto vocal, na zona de acoplamento dos tractos oral e nasal, para a vogais orais do PE adquiridas no corpus 3D (falante PAA).

Consoantes Nasais

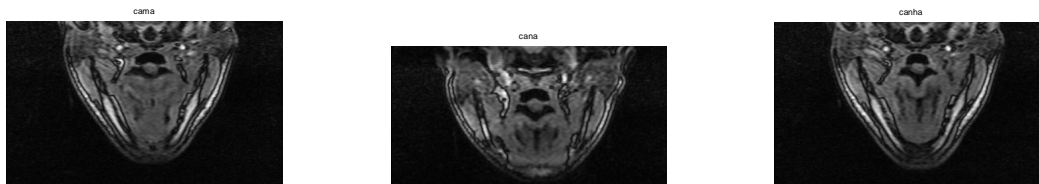


Figura 7. 43 - Corte perpendicular ao tracto vocal, na zona de acoplamento dos tractos oral e nasal, para as consoantes [m] à esquerda, [n] ao centro e [ɲ] à direita.

7.3.2.2 VPOQ médio

O VPOQ médio, foi calculado fazendo a média dos primeiros 5 cortes, em que se observaram duas aberturas:

$$VPOQ_{médio} = \left(\frac{\text{área}_{nasal}}{\text{área}_{oral}} \right)_{médio}$$

Na Figura 7.44, apresenta-se o gráfico relativo às áreas nasais e orais médias e respectivo coeficiente para vogais nasais, orais e consoantes nasais. Na Figura 7.45, apresentam-se os *boxplots* relativos ao VPOQ, intervenientes na média.

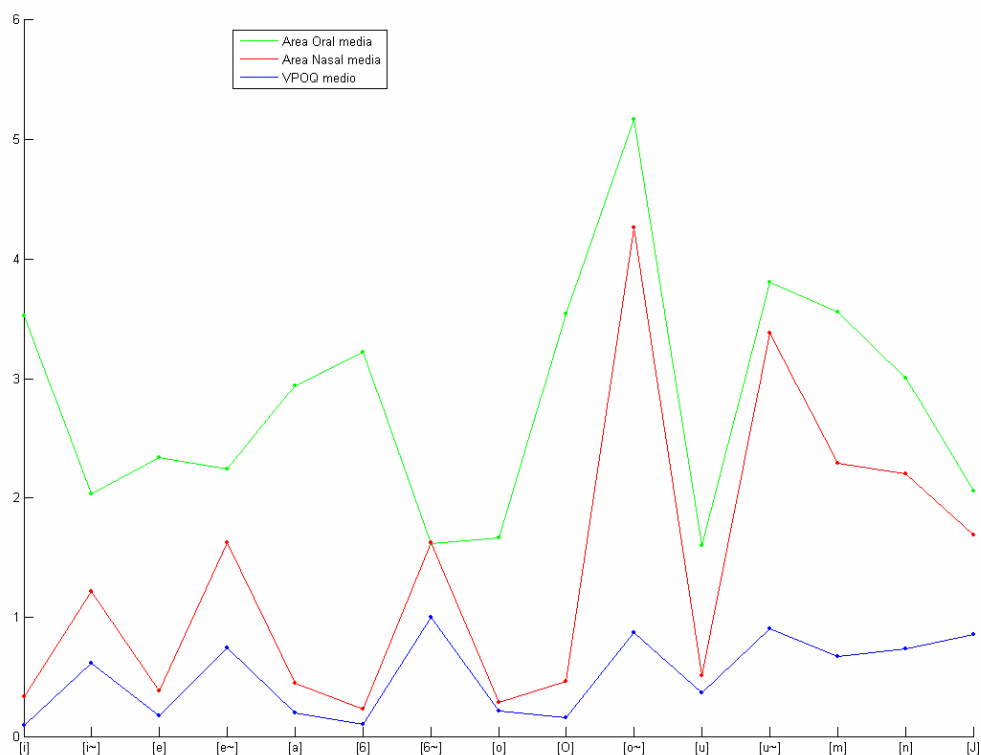


Figura 7.44 - Valores médios relativos às áreas nasais e orais e respectivo VPOQ ou CAV (azul) para vogais nasais, orais e consoantes nasais.

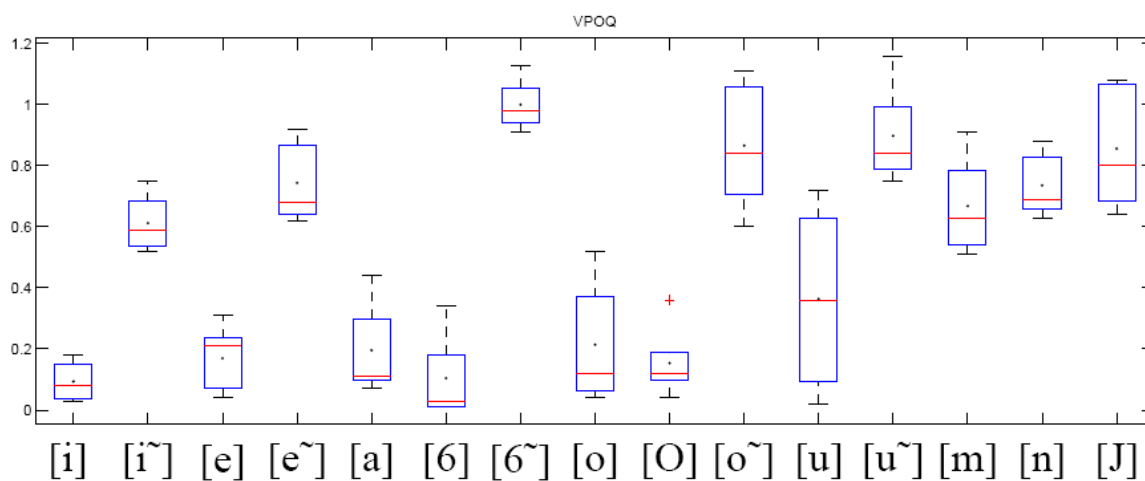


Figura 7.45 - Boxplots que representam a mediana (traço a vermelho), desvio-padrão e médias (ponto) relativas ao VPOQ (CAV) para vogais nasais, orais e consoantes nasais.

Da análise das figuras 7.44 e 7.45 podemos referir que:

- Para este falante do PE, verifica-se que o VPOQ médio é sempre superior nas vogais nasais do que nas correspondentes vogais orais.
- De todas as vogais nasais, a que apresenta o valor mais elevado de VPOQ é a vogal [6~] com valor de 1, seguida do [u~], [o~], [e~] e [i~] sendo esta a que apresenta valor mais baixo (0,61).
- Para as vogais orais verifica-se que o [u] é a vogal oral que apresenta VPOQ mais elevado, seguido de [o], [a], [e], [O], [6] e finalmente o [i] com o valor mais pequeno (0,009).
- A maior diferença entre VPOQ médio, para cada par oral/nasal foi observada para o [6] e [6~] com um valor de 0,9 e a menor diferença entre [u] e [u~] com 0,54.
- As consoantes nasais na generalidade, apresentam valores intermédios entre as vogais nasais e as vogais orais, sendo que, a que apresenta um valor de VPOQ superior é o [J], seguido de [n] e de [m].

Apesar de o valor do VPOQ ser mais pequeno nas orais este valor nunca é zero o que significa que, ainda que pequena, existe sempre passagem para a cavidade nasal mesmo nos sons não nasais o que corrobora o facto de que, nem sempre o facto de haver abertura para a cavidade nasal é, só por si, suficiente para se ter um som nasal.

À excepção do [6~] que é uma vogal central e que apresenta o maior VPOQ, podemos dizer que nas nasais se verifica que as vogais posteriores apresentam um VPOQ mais elevado que as anteriores. Será importante analisar até que ponto um VPOQ mais elevado é obtido à custa de um aumento da passagem nasal ou a uma diminuição da passagem oral.

Comparando com Engwall et al. (2006) para os sons do francês verificamos que: (1) o VPOQ médio segue, em termos gerais, um padrão semelhante ao observado para PE, isto é, superior nas nasais do que nas correspondentes orais e (2) os valores de VPOQ para o francês são significativamente mais elevados do que os obtidos para o PE, particularmente para as vogais nasais. Para o francês existe mesmo uma das vogais que apresenta um valor de VPOQ de aproximadamente 14, para um dos falantes, mas é também elevado para os outros 3 falantes. Segundo os autores (este valor tão elevado para esta vogal) poderá resultar não de um aumento significativo da cavidade nasal mas sim de uma muito pequena passagem oral. Para o PE, e particularmente para este falante tais valores, não se verificaram. Estes autores

encontraram também algumas diferenças entre os sujeitos que no nosso estudo, só com um falante não permite, inferir.

7.3.2.3 Área da cavidade faríngea

Apresentam-se de seguida os resultados obtidos para a área da cavidade faríngea na zona da oro-faringe para os mesmos fones (os valores apresentados representam a média de quatro cortes obtidos abaixo da área de acoplamento).

	FONE	CAVIDADE FARÍNGEA (M) (cm ²)
Vogal nasal	i~ [pi~tu]	4,86
	e~ [pe~t@]	4,27
	6~ [ka~tu]	2,41
	o~ [po~t@]	2,98
	u~ [pu~tu]	1,43
Vogal oral	i [pipu]	5,15
	e [pek6]	3,70
	a [patu]	1,77
	6 [k6d6]	2,26
	o [topu]	1,79
	O [pot@]	1,19
	u [bud6]	2,51
Consoante nasal	m [k6m6]	3,98
	n [k6n6]	4,52
	J [k6J6]	4,07

Tabela 7. 2 - Áreas (médias) da cavidade faríngea (orofaringe) para vogais nasais e orais e consoantes nasais do PE.

Pela análise da Tabela 7.2 podemos observar que, para este falante do PE, as áreas do tracto vocal na zona da orofaringe são maiores, no geral, para as consoantes nasais e para as vogais nasais anteriores [i~] e [e~] e também para a vogal oral [i]. Comparando os valores de vogais orais e nasais correspondentes verificamos que:

- A área da cavidade faríngea é maior, ainda que próxima da vogal nasal, para a vogal oral [i];
- Para o par [e] e [e~] verifica-se que área da cavidade faríngea é maior na vogal nasal do que na oral;
- Para o [6] [6~] e [a] verifica-se que o valor é muito próximo entre [6] e [6~], mas é ligeiramente superior na nasal, quer comparado com o [6] ou com o [a];
- No caso da vogal posterior [o~] a área a nível da oro-faringe é maior na nasal do que na oral;
- Finalmente para o [u] e [u~] verifica-se que a área é menor na produção da nasal.

Estes resultados obtidos com este falante, estão um pouco discordantes do descrito para o Francês relativamente às dimensões da cavidade faríngea na produção das vogais orais e nasais, uma vez que, geralmente, durante a produção das vogais nasais do Francês a faringe está sempre mais constrangida que a sua correspondente oral (Demolin et al. 1998, Demolin et al. 2003; Engwall, Delveaux & Metens, 2006) Estas diferenças estão provavelmente relacionadas com o diferente tipo de nasalidade apresentada pelas duas línguas.

Também, quando comparamos estes resultados com os contornos que se obtiveram em 2D, verificamos que nalguns sons há diferenças a nível da cavidade faríngea entre os resultados 2D e 3D.

7.3.3 Qualidade das imagens do corpus 3D

De forma genérica, podemos referir que a qualidade obtida com a aquisição volumétrica, proposta neste estudo, foi boa, e que o método poderá ser validado, uma vez que foi possível obter, por segmentação, as funções de área do tracto vocal, parâmetros como o CAV e obter visualizações tridimensionais. As maiores dificuldades observadas na segmentação das imagens foram ao nível da cavidade oral, onde a resolução no plano coronal foi inferior. Recordemos que neste caso, apesar de se adquirirem as imagens com uma sequência em volume, o que, em teoria permitiria uma muito boa resolução espacial nas reconstruções em qualquer dos planos, tal não se verificou pois optámos por uma aquisição muito rápida, não tendo sido possível ter voxel isotrópico. Apesar de, ter sido efectuada interpolação sobre as imagens, antes da segmentação o problema não ficou totalmente solucionado. O facto de a informação relativa às arcadas dentárias não estar ainda co-registada também não contribuiu para facilitar a tarefa de segmentação.

O tempo de aquisição (18s), tendo em atenção o volume de informação obtido (desde uma zona acima do palato até ao nível de C5-C6), foi bastante aceitável e muito inferior ao conseguido com a aquisição clássica por nós efectuada (53 s), e com os tempos de aquisição referenciados noutros estudos. Em termos de produção da fala, quanto menor for o tempo de aquisição melhor (dada a dificuldade em manter a produção estável e minimizar a hiperarticulação associada às produções estáticas) representando claramente uma vantagem.

O facto de se obter informação relativa a um volume permite-nos reconstruir a informação de forma diferente, as vezes que forem necessárias, tornando-a uma opção mais versátil. A sequência escolhida, 3D VIBE, geralmente utilizada a nível abdominal, em particular, para estudos dinâmicos após contraste, mostrou-se também válida neste contexto e com potencial para utilização em estudos noutras áreas do corpo nomeadamente a nível cerebral. Aliás, esta possibilidade foi referida recentemente (Dale, Schultze-Haakh & Priatna, 2006), uma vez que o VIBE, permite obter imagens de alta resolução, com ponderação T1, com tempos de aquisição curtos e boa cobertura anatómica. Foi também possível, obter através da utilização de software específico, visualização 3D do tracto vocal, de uma forma extremamente rápida. Como exemplo,

desta possibilidade apresenta-se visualização 3D, extraída com recurso ao ITK Snap, para a vogal nasal [6~] e correspondente oral [6], obtidas em cerca de um minuto (por fone). Esta forma de apresentação da informação tem aplicabilidade na construção de modelos 3D, mas não directamente na extracção de parâmetros articulatórios.

No entanto, em nossa opinião, a grande vantagem na utilização da aquisição volumétrica é permitir a obtenção de áreas a partir de cortes que foram obtidos perpendicularmente a cada ponto do tracto vocal (após aplicação de grelha adaptativa, ver capítulo 6). Esta possibilidade veio permitir contornar uma das limitações que têm sido referenciadas noutros estudos relativamente à RM. Um dos problemas com a aquisição refere-se às imagens dos extremos do volume, degradadas por artefactos de aliasing 3D (ver capítulo 4, ponto 4.6.2) característicos da aquisição volumétrica, o que não facilitou a segmentação na zona da glote.

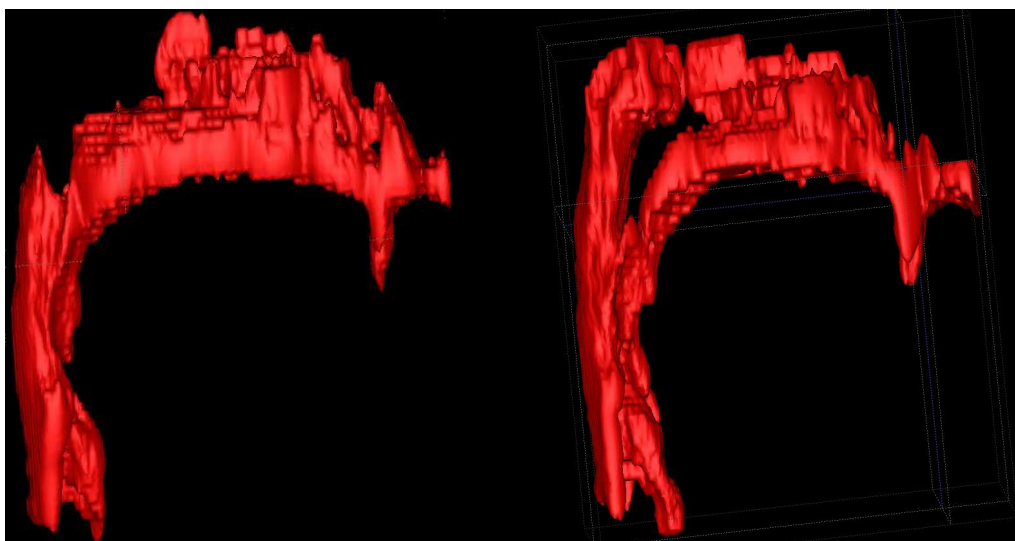


Figura 7. 46 – Visualização 3D, obtida com o ITK Snap, a partir da aquisição em Volume (3D-VIBE) para a vogal oral [6] à esquerda e vogal nasal [6~] à direita, para o falante PAA.

7.4 Corpus Tempo real

Com o corpus dinâmico obtido em tempo real esperava-se obter informação que permitisse observar a variação dos movimentos dos articuladores ao longo do tempo, em especial os movimentos da língua, a posição do velo, e os seus movimentos de abertura/fecho, estes dois últimos em particular, importantes para o estudo dos sons nasais.

Relativamente ao corpus obtido em tempo real, como já referimos, obteve-se uma velocidade de aquisição de 5 imagens por segundo (com uma resolução temporal de aproximadamente 200 ms). Efectuaram-se 100 medidas obtendo-se vinte segundos de aquisição, para cada sequência de

vogais ou para cada palavra repetida. A totalidade dos *itens* adquiridos em tempo real, foi apresentado no capítulo das metodologias.

Apesar de ter sido adquirida grande quantidade de informação em tempo real, no âmbito deste trabalho, só serão apresentados os primeiros resultados obtidos com este corpus. Dada a extensão do corpus não foi viável, no período deste trabalho, o desenvolvimento de técnicas de segmentação automática, ou mesmo semi-automática que permitissem a obtenção de toda a informação.

Uma vez que não foi efectuada gravação em simultâneo do sinal de voz (devido à inexistência de microfone e tecnologia adequada, no momento da aquisição do corpus), não foi possível alinhar as imagens com o sinal áudio, o que permitiria de forma mais rápida e precisa a identificação de cada frame em relação ao som produzido. Assim, as frames apresentadas, foram identificadas por comparação com as imagens estáticas adquiridas na mesma falante (Falante MTC), o que é um trabalho extremamente moroso. De salientar, que este é o método também reportado em outros estudos para retirar informação das imagens em tempo real (Demolin et al., 1997; Demolin, Metens & Soquet, 2000).

Sequência de vogais em Tempo Real

Assim, e relativamente a uma das sequências de vogais [a E i O u] adquiridas, apresenta-se na Figura 7.47, o conjunto de imagens referentes a cerca de 5,4 s de produção. Nestas imagens, observadas em modo-*cine* é possível observar os movimentos dos articuladores, em particular da língua e lábios, com clareza. Mais importante que obter a posição para uma determinada configuração, é visualizar a forma como os articuladores se movem, de forma articulada, de uma configuração para outra. A visualização do velo, foi comprometida por alguns artefactos de susceptibilidade magnética.

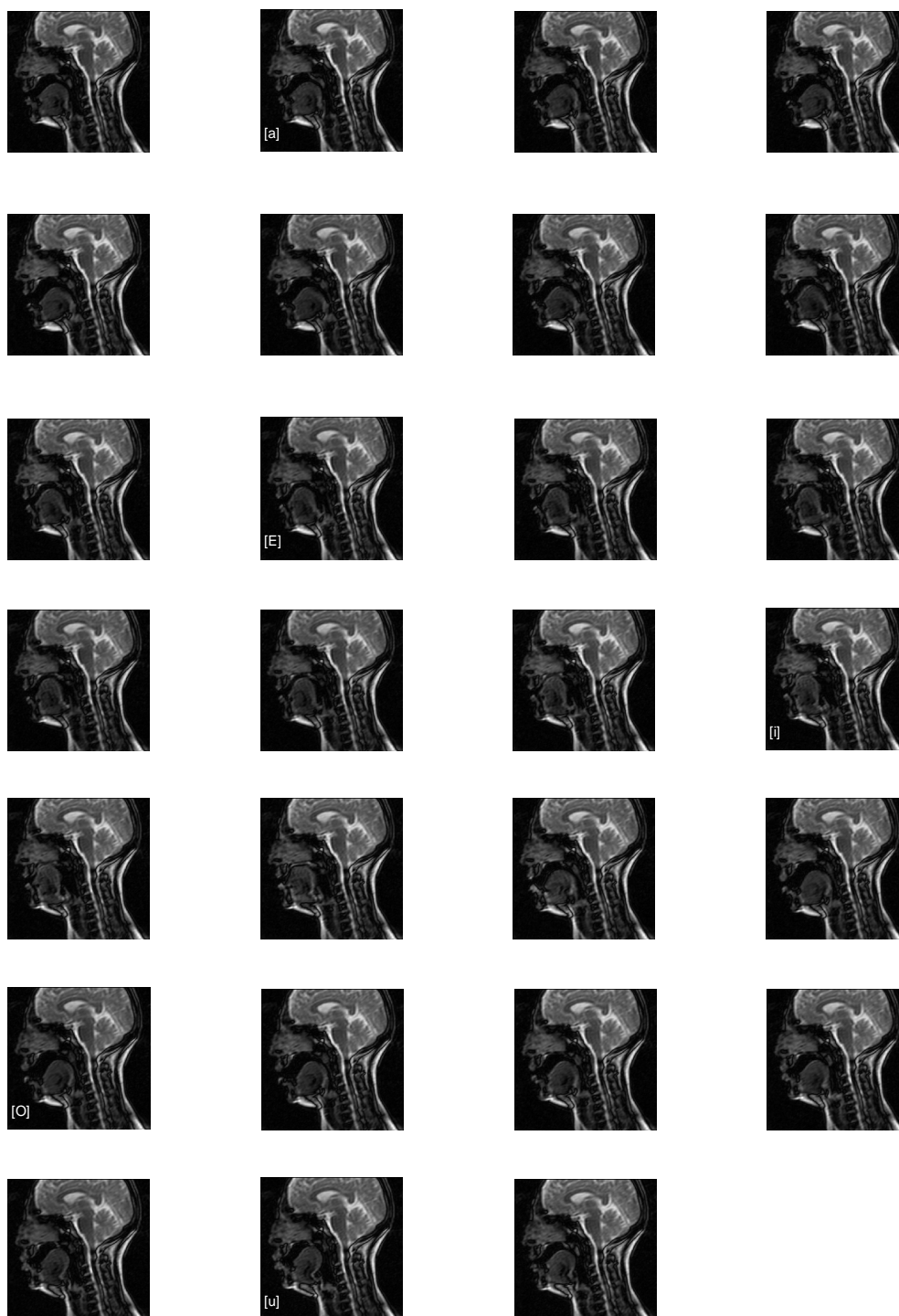


Figura 7. 47 - Vinte e sete frames identificadas, por comparação, referentes à produção da sequência de vogais [a E i O u] e que correspondem a cerca de 5,4 s de produção.

Outra das experiências efectuada foi a marcação de pontos, em três zonas diferentes da língua (ponta da língua, ponto mais alto do dorso da língua e posição da parte mais recuada da língua, a nível do bordo inferior de C2). Este tipo de informação tem correspondência, por exemplo, com a informação relativa aos pontos, onde são colocados os *pellets* ou eléctrodos, no X-Ray Microbeam e nos sistemas EMA, o que pode ser importante.

Identificação de três pontos na língua

Para a repetição da palavra Cato [katu], efectuou-se o seguinte procedimento:

- 1) Identificação de cada *frame* relativa a um determinado item, por comparação com as imagens estáticas (2D no plano sagital) obtidas para o mesmo falante (MTC);
- 2) Em cada *frame* procedeu-se à marcação de três pontos na zona da língua e observou-se a sua variação no tempo durante a repetição de cada palavra;
- 3) Os três pontos marcados foram: ponta da língua, ponto mais alto ao nível do dorso e posição da parte posterior da língua, tendo como referência o bordo inferior do corpo de C2 (ver figura 7.48).

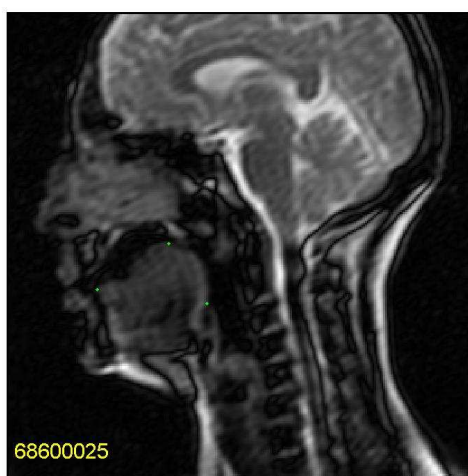


Figura 7. 48 - Marcação de três pontos na língua: (1) ponta da língua, (2) ponto mais elevado do dorso e (3) posição da raiz a nível do bordo inferior do corpo de C2. *Frame* referente à produção da vogal [u] durante a produção da palavra [katu].

Na figura 7.49, observa-se uma das repetições da palavra [katu], utilizada para obter o *plot* referente à variação temporal e espacial dos três pontos marcados na língua, durante a repetição da palavra, nos 20 segundos de aquisição.

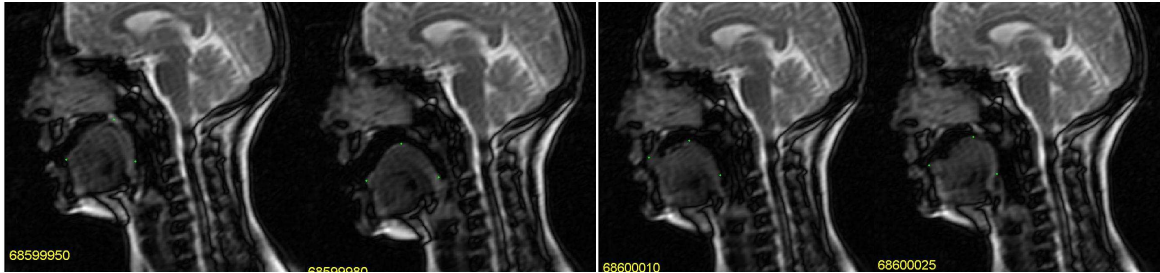


Figura 7. 49 - Frames referentes à produção de [k a t u], à esquerda [k], centro esquerda [a], centro direita [t] e à direita [u] (frames não sequenciais).

Na figura 7.50, observa-se a variação no espaço, dos três pontos marcados na língua durante os 20 segundos de aquisição para a palavra [katu]. É possível, mudando a forma de apresentação, identificar cada um dos diferentes sons, utilizando cor ou símbolos diferentes.

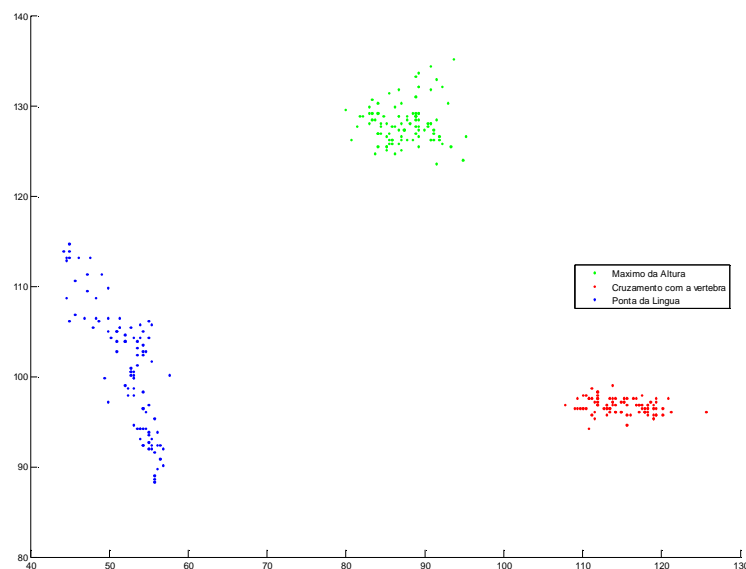


Figura 7. 50 - Variação espacial, dos três pontos marcados ao nível da língua durante a produção da palavra [katu].

Efectuou-se ainda outra abordagem às imagens em tempo real:

Após identificação das frames relativas a cada fone, e das frames intermédias, efectuou-se segmentação do contorno da língua, utilizando a técnica de *Live Wire*. Como exemplo, apresentamos na figura 7.51, a variação da posição da língua durante a produção de uma repetição da palavra [k6~tu], à esquerda e [katu] à direita. Relativamente à palavra [k6~tu], está representada informação relativa a nove frames, correspondendo a informação de cerca de 1,8 s de movimento da língua. Para a palavra [katu], representam-se as 7 frames relativas à sua produção (1,4 s).

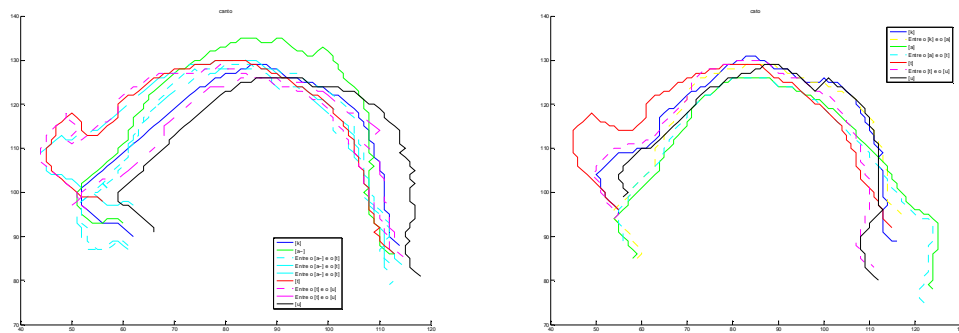


Figura 7. 51 - Pseudo - variação da posição da língua no tempo durante a produção de [k6~tu] (esquerda) e [katu] (direita).

Desta maneira, é possível descrever o movimento da língua, nomeadamente da ponta da língua, dorso e raiz na produção da palavra. Por exemplo, ao passar do [k], para o [6~], não existem alterações de posicionamento significativas a nível da ponta da língua, mas há uma elevação significativa do seu dorso. Depois, há um avanço e subida da ponta da língua, com abaixamento do dorso para a produção da oclusiva [t], produzida com a ponta da língua a tocar nos dentes (ou na região alveolar). Após produção do [t] a ponta da língua e o dorso, em conjunto, recuam para a produção do [u], uma vogal posterior. Como se observa, a zona da língua que menos sofre alterações, excepção feita para o [u], é a sua parte posterior e raiz. É possível também comparar os contornos relativos a cada frame, produzidos em palavras diferentes, como se pode observar na Figura, 7.52. Se, em conjunto com a segmentação da língua, forem segmentados os lábios e a mandíbula, é possível obter informação relativa a vários articuladores, permitindo avaliar a sua coordenação.

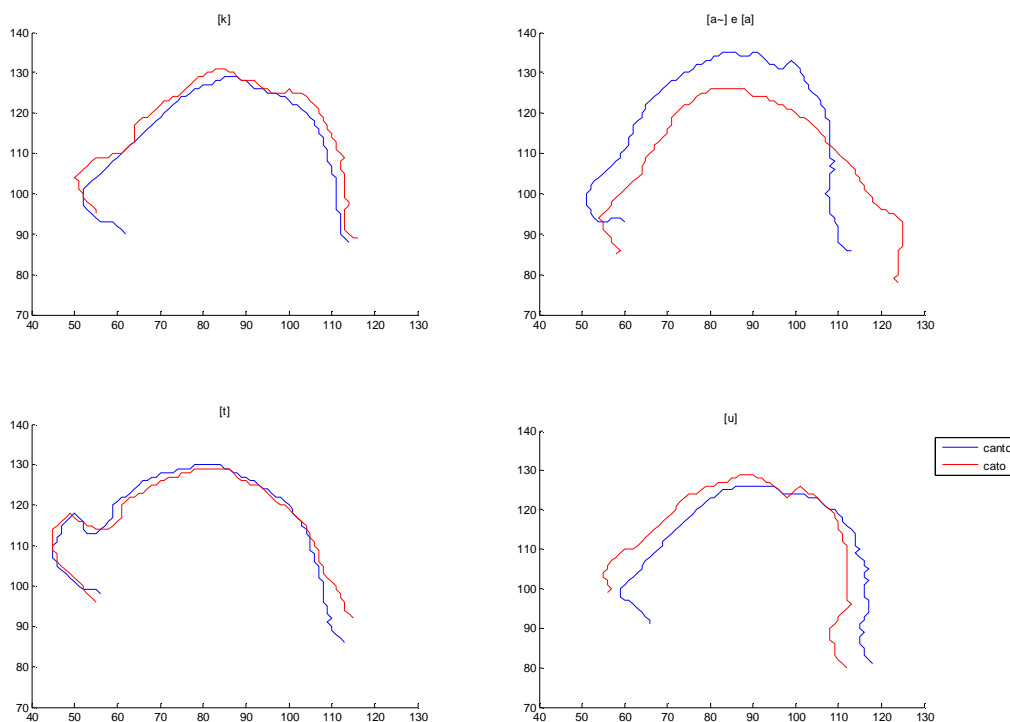


Figura 7. 52 - Sobreposição dos contornos da língua para o [k] (cima à esquerda), [a~] e [a] (cima à direita), [t] (baixo à esquerda) e [u] (baixo à direita) na produção referente a uma frame de [k6~tu] (azul) e [katu] (vermelho).

Qualidade das imagens em Tempo Real

Relativamente às imagens em tempo real, muito trabalho a nível de segmentação há ainda por fazer, para que seja possível apresentar informação temporal de uma forma mais rápida do que a conseguida no presente estudo. Mostrámos alguns possíveis caminhos para obter informação temporal, no entanto os métodos têm indubitavelmente que ser melhorados e automatizados. Será necessário também a implementação de técnicas de pré-processamento de imagem, com recurso a filtros para melhorar a qualidade das imagens e, posteriormente a segmentação.

Com a qualidade obtida com a sequência seleccionada para o estudo foi possível efectuar segmentação utilizando o *Live Wire*, para a zona da língua. Contudo, a zona do velo não foi ainda segmentada devido à presença de artefactos de susceptibilidade magnética (característicos das sequências EG). Quando se observam as imagens em modo *cine*, conseguimos ter uma ideia dos movimentos de velo, no entanto, quando se tenta frame a frame, ainda de que de forma manual, efectuar a sua segmentação é praticamente impossível. A resolução temporal não é ainda a

desejável, havendo necessidade de testar outras possibilidades, ou com outras sequências de pulso, técnicas de aquisição paralela, interpolação etc.

7.5 Sumário

Neste capítulo, apresentaram-se alguns dos resultados já extraídos da base de dados recolhida para os sons do PE. Grande parte da informação apresentada referiu-se ao Corpus 2D, onde se procurou descrever do ponto de vista articulatorio todos os sons do PE (com excepção das vibrantes), efectuando também comparações entre classes de sons, quando adequado. A qualidade das imagens permitiu atingir os objectivos definidos para o Corpus, havendo espaço para melhorias se for necessário obter informação com mais falantes. Os métodos de segmentação utilizados, revelaram-se também adequados, permitindo extrair informação de forma rápida e precisa, viabilizando a aquisição de maior número de informantes.

Relativamente ao Corpus 3D, apresentámos alguns resultados, focalizando o nosso interesse nos sons nasais e em sons cuja informação pudesse complementar a informação 2D. A informação obtida a partir das imagens 3D, obtida com método de aquisição volumétrica, permitiu a obtenção das funções de área do tracto de forma directa e mais rápida do que o habitualmente descrito. O tipo de aquisição efectuada, em conjunto, com as técnicas de segmentação permitiu obter áreas que foram obtidas ortogonalmente a cada ponto do tracto, o que é relevante.

Do corpus obtido em tempo real, apresentámos apenas alguns resultados preliminares, não tendo sido ainda possível incluir informação relevante, como os movimentos do velo, particularmente nos sons nasais, dadas as limitações referidas no ponto anterior. Sem dúvida que algum trabalho de melhoria, quer na aquisição das imagens, quer na sua segmentação deve ser efectuado.

Capítulo 8

Conclusões e Trabalho Futuro

Neste oitavo e último capítulo, apresenta-se uma síntese do trabalho efectuado (secção 8.1), referem-se os principais resultados e conclusões (secção 8.2), apontam-se algumas das limitações do estudo (secção 8.3) e, por fim, na secção 8.4 sugerem-se alguns tópicos que poderão ser objecto de desenvolvimentos futuros dentro desta área de investigação.

8.1 Resumo do trabalho realizado

Para que fosse viável a aquisição de dados anatómicos e fisiológicos, obtidos através da Ressonância Magnética, durante a produção de fala (PE), foi necessário um período de revisão bibliográfica, fundamental para nos integrarmos, numa área do conhecimento que não é próxima da nossa área de formação de base e que permitiu também conhecer o que, ao longo das últimas duas décadas, foi efectuado para outras línguas, na área da produção. Só conhecendo as especificidades desta área do conhecimento, seria possível adaptar ou sugerir um protocolo que nos permitisse obter a informação pretendida com a qualidade desejada. A integração numa equipa multidisciplinar e no projecto HERON, além de ser factor de motivação, facilitou-nos a tarefa, uma vez que alguns dos objectivos a alcançar se encontravam já bem delineados e estruturados.

Assim, uma das primeiras tarefas realizadas foi a elaboração do Corpus, que deveria incluir a maior parte dos sons do PE, mas com possibilidade de ser adquirido e tratado, pelo menos em parte, dentro do tempo disponível para a realização deste trabalho.

Foram testadas e optimizadas sequências e procedimentos que permitiram estabelecer o protocolo de aquisição de imagem, seguido na sessão final de recolha de dados.

Após aquisição e processamento das imagens, obtiveram-se e compararam-se contornos, obtiveram-se distâncias médio-sagitais (2D), extraíram-se parâmetros articulatorios e funções de área do tracto vocal (3D), para parte significativa dos sons do PE. Obteve-se informação em tempo real relativa a uma grande quantidade de sons, a uma razão de cinco frames por segundo.

Em particular, obteve-se informação relevante relativa aos sons nasais do PE (por exemplo, determinação da área de abertura do velo, o coeficiente de abertura do velo em vogais e consoantes nasais) dada a especificidade destes sons, ao nível da sua produção.

Testaram-se vários métodos de segmentação, em função da qualidade das imagens, essencialmente métodos semi-automáticos, o que nos pareceu ser uma mais valia com vista a futuros desenvolvimentos deste trabalho.

A sistematização de procedimentos, a aquisição de conhecimentos em áreas específicas e a resolução das dificuldades com que nos deparámos, por certo contribuirão para que seja possível, numa próxima oportunidade, melhorar e agilizar o processo de aquisição.

8.2 Principais resultados e conclusões

O resultado mais visível deste trabalho é a existência de uma **base de dados RM** relativa à produção do Português Europeu, cuja informação poderá ser utilizada na melhoria e validação do processo de síntese dos sons do PE, permitindo a continuidade do trabalho iniciado e desenvolvido na Universidade de Aveiro, por Teixeira e colaboradores na área da síntese articulatória (Teixeira, 2000; Teixeira & Vaz, 2001; Teixeira et al., 2004; Teixeira et al., 2005).

Assim, existe agora informação detalhada acerca de **várias classes de sons do Português Europeu**, nomeadamente dos sons **nasais, laterais e fricativas**, que esperamos possa contribuir para a melhoria do processo de síntese dos sons nasais e de fricativas, e permitir a expansão dos sons a serem sintetizados, nomeadamente a inclusão das laterais, cujos primeiros resultados apontam para a existência de diferenças em relação ao reportado para outras línguas, por exemplo o Inglês Americano (Narayanan, Alwan & Haker, 1997; Bangayan, Alwan & Narayanan, 1996).

Foi possível também, obter informação relativa à influência do contexto na produção dos sons (**coarticulação**), em particular para as **oclusivas e fricativas**. Apesar das imagens se referirem a configurações resultantes de produções sustidas (estáticas), o método utilizado na segmentação e comparação de contornos foi sensível a diferenças devidas a efeitos coarticulatórios. Os primeiros resultados corroboram, no geral, o obtido para outras línguas, nas classes de sons estudadas.

A informação obtida com a base de dados, e que será oportunamente disponibilizada à comunidade científica, permitirá também uma análise por parte de investigadores de outras áreas, podendo contribuir ainda que modestamente, para um aumento dos conhecimentos na área da Fonética, ser uma mais valia no ensino da mesma, ou também, e não menos importante, a

possibilidade de aplicação na área da Terapia da Fala, podendo as imagens obtidas ou os modelos construídos através delas, vir a ser utilizadas como apoio na correcção de algumas perturbações articulatórias ou serem utilizadas como ajudas visuais para o desenvolvimento da fala em surdos.

A nível técnico e relacionado com a aquisição de imagem, foi introduzido um método de aquisição diferente do usualmente referenciado nos estudos de produção de fala, uma vez que, neste estudo, quando se pretendeu obter **informação tridimensional** se recorreu a **uma sequência volumétrica** otimizada tendo em atenção o compromisso entre resolução espacial, S/R e tempo de aquisição. A utilização de uma sequência 3D, constitui um dos maiores desafios, não porque não seja utilizada como rotina num sector de RM, mas pelas exigências relativas à aquisição durante a fonação, nomeadamente a necessidade de um tempo de aquisição curto e grande volume de informação.

Com este método de aquisição foi possível a obtenção das funções de área do tracto vocal após segmentação, podendo, portanto concluir que o método é válido para aplicação nestes estudos. Em nossa opinião, constitui uma mais valia, uma vez que permite ultrapassar algumas das limitações apontadas ao método de aquisição tradicional: (1) permite um maior número de cortes (com menor tempo de aquisição) e obtenção de cortes mais finos, (2) como consequência, permite uma modelação mais adequada e fidedigna da função de área ao longo do tracto, (3) cada um dos cortes obtidos por segmentação é perfeitamente ortogonal ao tracto vocal, (4) é versátil, podendo em qualquer momento reconstruir e apresentar a informação de forma diferente e (5) permite obter visualização 3D, após segmentação, de uma forma muito rápida (pouco mais de um minuto). A implementação do método foi possível graças ao trabalho de segmentação desenvolvido em paralelo (Carbone, 2007).

8.3 Limitações do estudo

Como na maior parte dos estudos, existem limitações e uma das que pode ser apontada é sem dúvida alguma, o facto de a informação ser relativa apenas a um falante. Com um só falante é difícil efectuar generalizações, uma vez que não podem ser estudadas características da fala independentes do sujeito falante. No entanto, como primeiro estudo pareceu-nos uma melhor estratégia, tendo em atenção a escassez ou mesmo inexistência de dados para o PE e tendo em consideração o tempo disponível para efectuar este trabalho. Por outro lado, permitiu a constituição de um corpus mais alargado, permitindo num estudo preliminar obter informação relativa a praticamente todos os sons do PE, com excepção das vibrantes.

A resolução temporal obtida com as imagens em tempo real, é ainda insuficiente quando comparada com a velocidade de alguns articuladores durante a produção de fala. No entanto, pensamos que a maior limitação do corpus obtido em tempo real, e que deverá ser alvo de

atenção no futuro é a fraca qualidade das imagens particularmente, na zona do velo. Esta degradação da qualidade ficou a dever-se, essencialmente, à presença de artefactos de susceptibilidade magnética nessa zona. Sem dúvida que, para o estudo dos sons nasais é fundamental a obtenção de imagens em tempo real que permitam segmentar a zona do velo e observar a sua dinâmica (movimentos de abertura/fecho). Uma vez que a sequência utilizada neste estudo é muito sensível a diferenças de susceptibilidade magnética (sequência EG), dever-se-á, em abordagens futuras, otimizar uma sequência igualmente rápida, mas menos sensível a artefactos de susceptibilidade magnética como, por exemplo, utilização de sequência TSE Zoom, modo de aquisição radial, ou técnicas de segmentação do espaço K para diminuir o tempo de aquisição com sequências TSE.

8.4 Desenvolvimentos Futuros

Na sequência do trabalho desenvolvido no âmbito da presente dissertação e, num enquadramento um pouco mais alargado do Projecto HERON pensamos que, das muitas e possíveis alternativas para a melhoria e seguimento lógico deste trabalho, se deveria:

1) Terminar a análise dos dados obtidos e sua divulgação

Esta seria a primeira tarefa a desenvolver, uma vez que, grande parte da informação adquirida não foi ainda analisada e apresentada (por exemplo, o corpus obtido em tempo real) ou foi apresentada, apenas, de forma parcial (corpus 3D).

2) Estudos dirigidos para classes específicas de sons

Será também importante a realização de estudos mais específicos para determinadas classes de sons, com um maior número de falantes, e com maior número de repetições por falante, o que permitiria efectuar generalizações e quantificar variações inter e intra-sujeito. Os resultados das primeiras análises do nosso estudo apontam já para a necessidade de elaborar um corpus dirigido para o estudo de uma classe específica de sons do PE – as laterais. Esta tarefa, poderá, em nosso entender, ser facilmente realizada num curto espaço de tempo.

A introdução de indivíduos com diferente proveniência geográfica seria também de considerar e permitiria averiguar a existência de variações dialectais.

3) Optimização das sequências utilizadas

O recurso a uma técnica de aquisição volumétrica mostrou-se bastante vantajosa, como já se referiu. Contudo, a qualidade das imagens obtidas poderá ser significativamente melhorada, em particular, a obtenção de melhor resolução no plano coronal (cavidade bucal e cavidade oral) e melhor SNR (o que facilitará a segmentação), pelo que algum trabalho deverá ser dispensado na optimização de sequências 3D, para utilizar nestes estudos, quando se pretenda modulação 3D do tracto vocal.

Neste estudo, obtiveram-se imagens em tempo real utilizando RM, com uma velocidade de 5 *frames/s*. Com esta rapidez foi possível extrair alguma informação relativa à dinâmica de alguns articuladores, mas manifestamente insuficiente para estudar, por exemplo, os movimentos da ponta da língua ou do velo. Assim, é fundamental melhorar a resolução temporal obtida e testar utilização de sequências menos sensíveis a artefactos de susceptibilidade magnética, em particular quando se pretenda estudar o velo. Esta optimização poderá passar também pela utilização de bobines que permitam utilizar técnica de imagem paralela (PAT) ou bobines dedicadas (o que não foi possível neste estudo) que permitam a diminuição do tempo de aquisição, ou por outro lado, utilizar técnicas de interpolação (pós aquisição) que permitam aumentar o nº de imagens por segundo, tal como descrito em Narayanan et al (2004) ou Takemoto et al. (2003).

A melhoria da qualidade das imagens no geral, e o aumento da resolução temporal das imagens em tempo real constituem desafios que muito gostaríamos de ultrapassar, pois permitiria a continuação nesta linha de investigação na área das Ciências da Fala, capitalizando os conhecimentos adquiridos durante a fase dos estudos Pós-graduados, mas também, não menos importante, manter o contacto com a prática e investigação na área da Imagiologia, uma vez que, a nossa formação de base é como Técnica de Radiologia.

4) Aprofundar conhecimentos relativos à coarticulação

Neste trabalho, começámos por tentar avaliar fenómenos de coarticulação nalguns sons do PE, nomeadamente em fricativas e oclusivas, tendo obtido alguns resultados interessantes e encorajadores. Muito ficou ainda por fazer, e este nosso modesto trabalho pode apenas ser encarado como um pequeno estudo exploratório no âmbito do estudo da coarticulação para o PE. Do que foi exposto, no capítulo de resultados, ficou patente a necessidade de incluir maior variedade de contextos, nomeadamente contextos assimétricos, que permitam avaliar os diferentes efeitos coarticulatórios (DE) e (ED). Além disso, será importante avaliar não só o efeito das vogais nos sons consonânticos como também o efeito contextual das consoantes sobre as vogais, à semelhança do que já foi efectuado para outras línguas (mais com EPG e EMA do que com RM).

Será também desejável, a escolha de contextos menos artificiais, procurando construir um corpus mais baseado em palavras reais do que em combinações artificiais. A inclusão de outros sons consonânticos como as laterais e vibrantes que não foram contempladas neste trabalho, bem como de sons nasais, seria importantíssimo para poder conhecer de uma forma mais assertiva e completa a produção e interacção dos diversos segmentos durante a fala. A extensão de um corpus desta natureza e o difícil acesso a equipamentos de RM, limitaria por certo o nº de informantes a incluir, mas idealmente deveria contemplar mais do que um falante do PE. A utilização de outras técnicas exploratórias, em paralelo com RM, como a EPG e EMA deveria

também ser ponderada. Uma vez que a coarticulação é um fenómeno dinâmico, o seu estudo em tempo real seria o mais indicado.

5) Gravação do sinal de voz e sincronização

Será também fundamental a gravação do sinal de voz durante a aquisição de imagens em tempo real, ou em aquisições dinâmicas, o que só será possível com a utilização de equipamento adequado (microfone óptico com sistema de cancelamento de ruído) devido ao ambiente extremamente ruidoso durante a aquisição de imagens RM, em particular com técnicas de aquisição rápida. Com o sinal de voz obtido em simultâneo com a aquisição de imagem é possível, em cada momento, associar o som a uma determinada configuração na imagem. Se além desta gravação do sinal áudio, se sincronizar a aquisição da imagem, com a produção do falante, durante a repetição de palavras ou sílabas, um determinado nº de vezes, será possível mediante o recurso a técnicas de pós processamento de imagem obter uma resolução temporal (aparente) muito elevada.

6) Melhoria das técnicas de segmentação

Com o aumento do nº de imagens por segundo obtidas nos estudos em tempo real aumenta também o nº de imagens que é necessário segmentar, o que efectuado de forma manual se torna moroso e praticamente inexequível num período de tempo razoável (Narayanan et al 2004; Bresch et al., 2006). Assim, só fará sentido uma tentativa de aumento da resolução temporal e do número de imagens por segundo, se forem desenvolvidas em simultâneo técnicas de segmentação automática ou quasi-automática.

O desenvolvimento destas técnicas de segmentação é já objecto de investigação a nível de um projecto de Mestrado em Engenharia Electrónica em curso no Departamento de Electrónica e Telecomunicações da UA (Carbone, 2007) e deverá ter continuidade.

7) Indivíduos com perturbações da articulação

Neste estudo, considerou-se apenas a normalidade. A realização de estudos de produção com RM em indivíduos com perturbações articulatorias de etiologia diversa (como glossectomizados, por exemplo, tal como referido por Mády, 2000) deveria ser também considerada, uma vez que poderá contribuir para desenvolver estratégias terapêuticas mais adequadas.

De salientar que, este trabalho beneficiou com a constituição de uma equipa multidisciplinar e também com o desenvolvimento tecnológico verificado nos últimos anos na área da RM, tendo sido possível obter, num período de tempo relativamente curto, o que muitos foram construindo de forma incremental ao longo de alguns anos.

Finalmente, pensamos ter dado um pequeno passo para a divulgação da utilização da RM e de outras técnicas de imagem em estudos de produção de fala, em Portugal. Esperamos também, que a constituição de uma equipa multidisciplinar na área das Ciências da Fala, tenha aberto o caminho a projectos futuros.

Referências Bibliográficas

- Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 641-647.
- Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *Journal of the Acoustical Society of America (JASA)*, 101(2), 1078-1089.
- Andrade, A., & Viana, M. C. (1996). Fonética. In I. H. F. Faria, E. R. Pedro, I. Duarte & C. Gouveia (Eds.), *Introdução à Linguística Geral e Portuguesa* (pp. 113-167): Caminho.
- Andrade, A. (1997). Variação fonética de /l/ em ataque silábico em Português europeu. Proc XIII Encontro Nacional da Associação Portuguesa de Linguística, Lisboa (Faculdade de Letras da Universidade de Lisboa).
- Andrade, A. (1999). On /l/ velarization in European Portuguese. Proc International Conference of Phonetics (ICPhS), San Francisco.
- Apostol, L., Perrier, P., Raybaudi, M., & Segebarth, C. (1999). *3D geometry of the vocal tract and inter-speaker variability*. Proc. of the XIVth International Congress of Phonetic Sciences (ICPhS).
- Araújo, P. A. G. M. d. (2007). A Influência de Diferentes Tipos de Oclusão Dentária na Produção de Sons da Fala. Dissertação de Mestrado. Universidade de Aveiro, Aveiro.
- Arnal, A., Badin, P., Brock, G., Connan, P.-Y., Florig, E., Perez, N., et al. (2000). *An X-ray Database for French*. Proc 5th Seminar on Speech Production.
- Badin, P., Bailly, G., Raybaudi, M., & Segebarth, C. (1998). *A three-dimensional linear articulatory model based on MRI data*. Proc 5th International Conference on Spoken Language Processing (ICSLP).
- Badin, P., Borel, P., Bailly, G., Revéret, L., Baciuo, M., & Segebarth, C. (2000). *Towards an Audiovisual Virtual Talking Head: 3D Articulatory Modeling of Tongue, Lips and Face Based on MRI and Video Images*. Proc 5th Seminar on Speech Production.
- Badin, P., Pouchot, L., Bailly, G., Raybaudi, M., Segebarth, C., Lebas, J. F., et al. (1998). *Un Modèle Articulaire Tridimensionnel du Conduit Vocal Base Sur des Données IRM*. Proc Journées d'Etude sur la Parole (JEP '98).
- Baer, T., Gore, J. C., Boyce, S., & Nye, P. W. (1987). Application of MRI to the analysis of speech production. *Mag. Res. Imag.*, 5(1), 1-8.
- Baer, T., Gore, J. C., Gracco, L. C., & Nye, P. W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *Journal of the Acoustical Society of America (JASA)*, 90(2), 799-828.

- Bangayan, P., Alwan, A., & Narayanan, S. (1996). From MRI and acoustic data to articulatory synthesis: A case study of the lateral approximants in American English. *Proc of the International conference of Spoken Language Processing*, 793-796.
- Barbosa, J. M. (1994). *Introdução ao Estudo da Fonologia e Morfologia do Português*. Coimbra: Livraria Almedina.
- Bitar, R., Leung, G., Perng, R., & Tadros, S. (2006). MR Pulse Sequences: What every Radiologist wants to know but is afraid to ask. *Radiographics*, 26(2), 513-537.
- Bresch, E., Nielsen, J., Nayak, K., & Narayanan, S. (2006). Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *Journal of the Acoustical Society of America (JASA)*, 120(4), 1791-1794.
- Brown, M. A., & Semelka, R. C. (1999). MRI Imaging Abbreviations, Definitions and Descriptions: A Review. *Radiology*, 213, 647-662.
- Brown, M. A., & Semelka, R. C. (2003). *MRI: Basic Principles and Applications*. Wiley.
- Bushberg, J., Seibert, A., Jr, E. L., & Boone, J. (2002). *The Essential Physics of Medical Imaging* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.
- Bushong, S. (2001). *Radiologic Science for Technologists - Workbook and Laboratory Manual* (Seventh ed.): Mosby.
- Carbone, I., Martins, P., Teixeira, A., & Silva, A. (2006). Base de Dados RM dos Sons do Português Europeu (Relatório Técnico). Aveiro: Universidade de Aveiro.
- Carbone, I. (2007 (não publicado)). Processamento de imagens de RM do tracto vocal. Universidade de Aveiro, Aveiro.
- Carroll, T., Sakaie, K., Wielopolski, P., & Edelman, R. (2006). Advanced Imaging Techniques, Including Fast Imaging. In J. H. R. Edelman, M. Zlatkin & J. Crues (Ed.), *Clinical Magnetic Resonance Imaging* (Vol. 1): Philadelphia: Elsevier.
- Caseiro Alves, F. (1998). Ressonância Magnética da Cirrose Hepática: Contributo dos novos agentes de contraste na avaliação morfológica e funcional. Tese de doutoramento. Universidade de Coimbra. Coimbra.
- Chiba, T., & Kajiyama, M. (1941). The Vowel: Its Nature and Structure. *Tokyo-Kaiseikan Publishing Company, Ltd.*
- Chodorowski, A., Mattsson, U., Langille, M., & Hamarneh, G. (2005). Color lesion boundary detection using live wire. *Proc SPIE*.
- Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: MIT Press.
- Crary, M. A., Gorham, M., Burton, S., Kotzur, I. M., & Gauer, J. (1996). Dynamic magnetic resonance imaging in the study of vocal tract configuration. *Journal of Voice*, 10(4), 378-388.
- Cruz-Ferreira, M. (1999). European Portuguese. In *Handbook of The International Phonetic Association*. Cambridge: Cambridge University Press.
- Dang, J., & Honda, K. (1994). MRI measurements and acoustic of the nasal and paranasal cavities. *Journal of the Acoustical Society of America (JASA)*, 94(3, Pt 2), 1765.

- Dang, J., & Honda, K. (2002). Improvements of a Physiological Articulatory Model in Construction and Control Strategy. *First Pan-American/Iberian Meeting on Acoustics*, 12(5), 2418-2419.
- Dang, J., & Honda, K. (2004). Construction and control of a physiological articulatory model. *Journal of the Acoustical Society of America (JASA)*, 115(2), 853-870.
- Dang, J., Honda, K., & Perrier, P. (2004). Implement of Coarticulation in Physiological Articulatory Model. *International Congress of Acoustics 2004*, 1335-1336.
- Dang, J., Honda, K., & Suzuki, H. (1994). Morphological and acoustical analysis of the nasal and the paranasal cavities. *Journal of the Acoustical Society of America (JASA)*, 96(4), 2088-2100.
- Dang, J., Shadle, C., Kawanishi, Y., Honda, K., & Suzuki, H. (1998). An experimental study of the open end correction coefficient for side branches with an acoustic tube. *Journal of the Acoustical Society of America (JASA)*, 104(2), 1075-1084.
- Davison, N. E., Eviatar, H., & Somorjai, R. L. (2000). Snakes simplified. *Pattern Recognition*, 33(10), 1651-1664.
- Dale, B. Schultz-Haakh, H., Priatna, A. (2006). VIBE: Fast Gradient Echo Imaging for every part of the body. *Magneton Flash*, (3), 56-59.
- Delattre, P. (1968). Divergences entre nasalités vocalique et consonantique en Français. *Word*, 24, 64-73.
- Deller, J., Proakis, J., & Hansen, H. (1993). *Discrete-Time Processing of Speech Signals*: Macmillan.
- Delvaux, V., Metens, T., & Soquet, A. (2002a). French nasal vowels: acoustic and articulatory properties. *Proc 7th International Conference on Spoken Language Processing (ICSLP)*, Denver.
- Delvaux, V., Metens, T., & Soquet, A. (2002b). Propriétés acoustiques et articulatoires des voyelles nasals du français. *Proc XXIVèmes Journées d' Étude sur la Parole*, Nancy.
- Delvaux, V. (2003). Perceptual cues for French nasal vowels: Evidences from fixed-classification and identification experiments. *Phonology. From representations to constraints*.
- Demolin, D., Metens, T., & Soquet, A. (1996). *Three-dimensional measurement of the vocal tract by MRI*. *Proc 4th International Conference on Spoken Language Processing (ICSLP)*.
- Demolin, D., George, M., Lecuit, V., Metens, T., Soquet, A., & Raeymaekers, H. (1997). *Coarticulation and articulatory compensations studied by dynamic MRI*. *Proc 5th European Conference on Speech Communication and Technology (Eurospeech)*.
- Demolin, D., Lecuit, V., Metens, T., Nazarian, B., & Soquet, A. (1998). Magnetic Resonance Measurements of the Velum Port Opening. *Proc 5th International Conference on Spoken Language Processing (ICSLP)*.
- Demolin, D., Metens, T., & Soquet, A. (2000). Real Time MRI and Articulatory Coordinations in Vowels. *Proc 5th Seminar on Speech Production*, Germany.
- Demolin, D., Delvaux, V., Metens, T., & Soquet, A. (2003). Determination of velum opening for French nasal vowels by magnetic resonance imaging. *Journal of Voice*, 17(4), 454-467.

- Drake, R. L., Vogl, W., & Mitchell, A. W. M. (2005). *Gray's anatomy for students*. Philadelphia (PA)
- Elster, A., & Burdette, J. (2001). *Questions & Answers in Magnetic Resonance Imaging* (2nd ed.). St. Louis: Mosby Inc.
- Engwall, O. (2000a). *A 3D tongue model based on MRI data*. Proc 6th International Conference on Spoken Language Processing (ICSLP).
- Engwall, O. (2000b). *Are static MRI measurements representative of dynamic speech? Results from a comparative study using MRI, EPG and EMA*. Proc. of the 6th International Conference on Spoken Language Processing (ICSLP).
- Engwall, O. (2000c). *Dynamical aspects of coarticulation in Swedish fricatives – a combined EMA & EPG study*. *TMH-QPSR*(4).
- Engwall, O. (2001). *Making the Tongue Model Talk: Merging MRI & EMA Measurements*. Proc 7th European Conference on Speech Communication and Technology (Eurospeech).
- Engwall, O. (2002). *Tongue Talking: studies in intraoral speech synthesis*. Tese de doutoramento, KTH - Royal Institute of Technology, Stockholm.
- Engwall, O. (2003a). *Combining MRI, EMA and EPG measurements in a three-dimensional tongue model*. *Speech Communication*, 41(2-3), 303-329.
- Engwall, O. (2003b). *A revisit to the Application of MRI to the Analysis of Speech Production - Testing our assumptions*. *Proceedings of the 6th Seminar on Speech Production*, 43-48.
- Engwall, O. (2004). *From real-time MRI to 3D tongue movements*. Proc 8th International Conference on Spoken Language Processing (ICSLP).
- Engwall, O., Delvaux, V., & Metens, T. (2006). *Interspeaker Variation in the Articulation of Nasal Vowels*. Proc *ISSP*.
- Engwall, O., & Badin, P. (1999). *Collecting and analysing two and three dimensional MRI data for Swedish*. *Speech Transmission Laboratory: Quarterly Progress and Status Report (STL-QPSR)*, 11-38.
- Ericsson, C. (2005). *Articulatory-Acoustic Relationships in Swedish Vowel Sounds*. Tese de Doutoramento. Stockholm University, Stockholm.
- Falcão, A. X., Udupa, J. K., Samarasekera, S., Sharma, S., Hirsch, B. E., & Lotufo, R. d. A. (1998). *User-steered image segmentation paradigms: Live wire and live lane*. *Graphical Models and Image Processing*, 60(4), 233-260.
- Fant, G. (1960). *Acoustic Theory of Speech Production* (Second ed.). The Netherlands: Mouton and Co., Gravenhage.
- Fant, G. (2001). *On the Speech code*. *TMH-QPSR*, Vol. 42: 61-67
- Farnetani, E. (1999). *Coarticulation and Connected Speech Processes*. In W. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell Publishers.
- Farnetani, E. & Recasens, D. (1999). *Coarticulation models in recent speech production theories*. In W. J. a. H. Hardcastle, N. (Ed.), *Coarticulation: Theory, Data, Techniques* (pp. 31-68). Cambridge: Cambridge University Press.

- Fletcher, J., & Harrington, J. (1999). Lip and jaw coarticulation. In W. J. H. a. N. Hewlwt (Ed.), *Coarticulation: theory, data and techniques* (pp. 164-175). Cambridge: Cambridge University Press.
- Foldvick, A. K., Kristiansen, U., & Kvoerness, J. (1995). *Three dimensional ultrasound and magnetic resonance imaging: a new dimension in phonetic research*. Proc XIIIth International Congress of Phonetic Sciences (ICPhS).
- Fujita, S., & Honda, K. (2005). An experimental study of acoustic characteristics of hypopharyngeal cavities using vocal tract solid models. *Acoustical Science and Technology*, 26(4), 353-357.
- Gibbon, F., & Nicolaidis, K. (1999). Palatography. In W. J. H. N. Hewlett (Ed.), *Coarticulation: Theory, Data and Techniques* (pp. 229- 244.): Cambridge: CUP.
- Gick, B., Kang, A. M., & Whalen, D. H. (2000). *MRI and X-ray evidence for commonality in the dorsal articulations of English vowels and liquids*. Proc 5th Seminar on Speech Production.
- Gick, B., Kang, A. M., & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, 30(3), 357-371.
- Glokner, J., Hu, H., Standley, D., Angelos, L., & King, K. (2005). Parallel MR Imaging: A user's guide. *Radiographics*, 25(5), 1279-1297.
- Graetzer, S. (2006). *Consonantal coarticulation resistance in vowel-consonant-vowel sequences in two Australian languages*. Proc Eleventh Australasian International Conference on Speech Science and Technology Auckland, New Zealand.
- Greenwood, A. R., Goodyear, C. C., & Martin, P. A. (1992). *Measurements of vocal tract shapes using magnetic resonance imaging*. Proc IEEE Communications Speech & Vision.
- Gregio, F. N. (2006). *Estudo das vogais do Português Brasileiro com técnicas de MRI*. Dissertação de Mestrado, Pontificia Universidade Catolica De Sao Paulo, Brasil, S. Paulo.
- Hashemi, R., & Bradley, Jr. W. (1997). *MRI the basics*. Philadelphia: Lippincott Williams &Wilkins.
- Heidemann, R. M., Özsarlak, Ö., Parizel, P. M., Michiels, J., Kiefer, B., Jellus, V., et al. (2003). A brief review of parallel magnetic resonance imaging. *European Radiology*, 13(10), 2323-2337.
- Honda, K. (2002). Evolution of vowel production studies and observation techniques. *Acoustical Science and Technology*, 23(4), 189.
- Hoole, P. (1993). Methodological considerations in the use of electromagnetic articulography in phonetic research. *FIPKM*, 31, 43-64.
- Hoole, P., & Nguyen, N. (1999). Electromagnetic articulography. In W. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 260-269). Cambridge: Cambridge University Press.
- Hoole, P., Wismüller, A., Leinsinger, G., Kroos, C., Geumann, A., & Inoue, M. (2000, May 01-04). Analysis of tongue configuration in multi-speaker, multi-volume MRI data. Proc. of the 5th Seminar on Speech Production, Seeon, Germany.
- Hornack. (1996). Physics of MRI. Retrieved 22-09-2006, 2006, from www.cis.rit.edu/htbooks/mri

- Huettel, S., Song, A., & McCarthy, G. (2004). *Functional Magnetic Resonance* (Vol. 1). Massachusetts: Sinauer Associates Inc.
- Hunold, P., Maderwald, S., Ladd, M., Jellus, V., & Barkhausen, J. (2004). Parallel acquisition techniques in cardiac cine magnetic resonance imaging using TrueFISP sequences: comparison of image quality and artifacts. *Journal of Magnetic Resonance Imaging*, 20(3), 506-511.
- Jesus, L. M. T. (1997). Speech coding and synthesis using parametric curves. Dissertação de Mestrado. University of East Anglia, Norwich, UK.
- Jesus, L. M. T. (2001). *Acoustic Phonetics of European Portuguese Fricative Consonants*. Tese de doutoramento, University of Southampton, Southampton.
- Johansson, C., Sundberg, J., Wilbrand, H., & Ytterbergh, C. (1983). From sagittal distance to area. *STL-QPSR*, 39-49.
- Karantanas, A., Papanikolaou, N., Kalef-Ezra, J., Challa, A., Gourtsoyiannis, N. (2000). Blueberry juice used per os in upper abdominal MR imaging: composition and initial clinical data. *European Radiology*, 10, 909-913.
- Kent, R. D. (1997). *The speech sciences*. New York (NY): Thomson
- Kim, H. (2004). Stroboscopic-Cine MRI data on Korean coronal plosives and affricates: Implications for their place of articulation as alveolar. *Phonetica*, 61(4), 234-251.
- Kiritani, S., Takenaka, E., & Sawashima, M. (1978). Computer tomography of the vocal tract. *Ann. Bull. RILP*(12), 1-4.
- Kröger, B. J., Winkler, R., Mooshammer, C., & Pompino-Marschall, B. (2000). *Estimation of vocal tract area function from Magnetic Resonance Imaging: Preliminary results*. Proc 5th Seminar on Speech Production.
- Ladefoged, P. (2001). *Vowels and Consonants: An introduction to the sounds of languages*. Oxford: Blackwell Publishers, Inc.
- Lafuente, J. (2000). *Atlas de Tecnología de la Resonancia Magnética*. Madrid: Mallinckrodt publications.
- Laver, J. (1980). *The phonetic description of voice quality* Cambridge: Cambridge University Press.
- Laver, J. (1995). *Principles of phonetics*. Cambridge: Cambridge University Press.
- Liang, Z.-P., & Lauterbur, P. (2000). *Principles of Magnetic Resonance Imaging- A signal processing perspective*. New York: IEEE Press.
- Lima, J. J. P. (1995). *Física dos Métodos de Imagem com Raios X*: Edições ASA.
- Mády, K., Sader, R., Zimmermann, A., Hoole, P., Beer, A. J., & Zeilhofer, H. (2002). *Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction*. Proc International Speech Motor Conference.
- Magen, H. S., Kang, A. M., Tiede, M. K., & Whalen, D. H. (2003). Posterior pharyngeal wall position in the production of speech. *Journal of Speech Language Hearing Research*(46), 241-251.

- Marques, C. B. (2003). Ressonância Magnética. In F. d. M. Gaivão (Ed.), *Imagiologia Clínica: Princípios e técnicas* (1st ed., pp. 369). Coimbra: Serviço de Imagiologia-Hospitais da Universidade de Coimbra.
- Martins, M. (2003). Ultrasonografia. In F. M. Gaivão (Ed.), *Imagiologia Clínica: Princípios e técnicas*. Coimbra: Serviço de Imagiologia- Hospitais da Universidade de Coimbra.
- Martins, M. R. (1988). *Ouvir falar - Introdução à Fonética do Português* (2nd ed.). Lisboa: Editorial Caminho.
- Martins, P., Carbone, I., Pinto, A., Silva, A., & Teixeira, A. (2007). European Portuguese MRI Based Speech Production Studies. Paper submitted to SPEECH COMMUNICATION (waiting for Committee Review).
- Martins, P., Carbone, I., Silva, A., & Teixeira, A. (2007). An MRI study of European Portuguese nasals. Proc at the Interspeech 2007.
- Masaki, S., Tiede, M., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., et al. (1997). MRI observation of dynamic articulatory movements using a synchronized sampling method. *Journal of the Acoustical Society of America (JASA)*, 102(5:2), 3166.
- Masaki, S., Tiede, M., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., et al. (1999). MRI-based speech production study using a synchronized sampling method. *Journal of the Acoustical Society of America (JASA)*, 20, 375-379.
- Mateus, M. H. M., Andrade, A., Viana, M C., & Villalva, A. (1990). *Fonética, Fonologia e Morfologia do Português*: Universidade Aberta.
- Mateus, M. H. M., Falé, I., & Freitas, M. J. (2005). *Fonética e Fonologia do Português*. Lisboa: Universidade Aberta
- Mathiak, K., Hertrich, I., Kincses, W. E., Klose, U., Ackermann, H., & Grodd, W. (2000). Stroboscopic articulography using fast magnetic resonance imaging. *International journal of language & communication disorders/Royal College of Speech & Language Therapists*, 35(3), 419-425.
- McInerney, T., & Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2), 91-108.
- Menzerath, P., & Lacerda, A. d. (1933). *Koartikulation, Steuerung und Lautabgrenzung*: Berlin Bonn (Dümmler).
- Mermelstein, P. (1973). Articulatory model for the study of speech production. *Journal of The Acoustic Society of America (JASA)*, 53(4), 1070-1082.
- Mohammad, M. A.-J. (1999). *Dynamic measurements of speech articulators using MRI*. Tese de Doutoramento. University of Southampton, Southampton, England.
- Moll, K. (1960). Cinefluorographic techniques in speech research. *Journal of Speech and Hearing Research*(3), 227-241.
- Moore, C. A. (1992). The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images. *Journal of Speech and Hearing Research*, 35(5), 1009-1023.

- Mortensen, E. N., Morse, B. S., & Barrett, W. A. (1992). *Adaptive boundary detection using 'live-wire' two-dimensional dynamic programming*. Proc IEEE Proceedings Computers in Cardiology, Durham, NC, USA.
- Moutinho, L. C. (2000). *Uma Introdução ao estudo da fonética e fonologia do Português*. Plátano Editora.
- Munhall, K., Vatikiotis-Bateson, E., & Tohkura, Y. (1995). X-ray film database for speech research. *Journal of Acoustical Society of America*, 98(2), 1222-1224.
- Narayanan, S., Alwan, A., & Haker, K. (1995a). An Articulatory Study of Fricative Consonants using MRI. *Journal of the Acoustical Society of America (JASA)*, 98(3), 1325-1347.
- Narayanan, S., Alwan, A., & Haker, K. (1995b). *An articulatory study of Liquid approximants in American English*. Proc XIIIth International Congress of Phonetic Sciences (ICPhS).
- Narayanan, S., Alwan, A., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *Journal of the Acoustical Society of America (JASA)*, 101(2), 1064-1077.
- Narayanan, S., Kaun, A., Byrd, D., Ladefoged, P., & Alwan, A. (1996). *Liquids in Tamil*. Proc 4th International Conference on Spoken Language Processing (ICSLP).
- Narayanan, S., Nayak, K., Lee, S. B., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *Journal of the Acoustical Society of America (JASA)*, 115(4), 1771-1776.
- Nayak, K., Hargreaves, B., Hu, B., Nishimura, D., Pauly, J., & Meyer, C. (2005). Spiral Balanced Steady-State Free Precession Cardiac Imaging. *Magnetic Resonance in Medicine*, 53, 1468-1473.
- NessAiver, M., Kahana, Y., Stone, M., Parthasarathy, V., & Paritsky, A. (2006). Recording high quality speech during tagged cine-MRI studies using a fiber optic microphone. *Journal of Magnetic Resonance Imaging*, 23(1), 92-97.
- Nguyen-Trong, N., Hoole, P., & Marchal, A. (1991). Articulatory-acoustic correlations in the production of fricatives. Proc XIIth Int. Cong. Phonetic Sciences.
- Nguyen, N. (2000). Review of Massaro. In *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*: MIT Press.
- Nguyen, N., Gibbon, F., & Hardcastle, W. J. (1996). Articulatory and perceptual aspects of fricative-stop coarticulation: a pilot study. *QMC Working Papers in Speech and Language Sciences* 1, 13-25.
- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *J. Acoust. Soc. Am.*, 92, 3078-3096.
- Perkell, J. (1999). Articulatory Processes. In W. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell Publishers, Inc.
- Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge: MIT press.

- Pires, J. C. (2005). Introdução à RMN. On *Métodos e Técnicas Avançadas em Radiologia II*. Aveiro.
- Pfitzinger, H. R. (2005). *Concatenative Speech Synthesis with Articulatory Kinematics obtained via Three-Dimensional Electro-Magnetic Articulography*. Proc DAGA Munich.
- Pompino-Marschall, B. (2005). Von Kempelen et al.- Remarks on the history of articulatory-acoustic modelling. *ZAS Papers in Linguistics*, 40, 145-159.
- Pooley, R. (2005). AAPM/RSNA Physics Tutorial for Residents: Fundamental Physics of MR imaging. *RadioGraphics*(25), 1087-1099.
- Pritchard, D. (2002). *Vocal Tract Visualisation* (EECE593 Winter 2002 Project Report).
- Pruessmann, K., Weiger, M., Scheidegger, M., & Boesiger, P. (1999). SENSE: Sensitivity Encoding for Fast MRI. *Magnetic Resonance in Medicine*, 42, 952-956.
- Pruthi, T. (2006). *Analysis, Vocal Tract Modeling and Automatic Detection of Vowel Nasalisation*. Proposta de tese de doutoramento. University of Maryland, Maryland.
- Recasens, D. (1983). Place cues for nasal consonants with special reference to Catalan. *Journal of the Acoustical Society of America*, 73(4), 1346-1353.
- Recasens, D. (1984). V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study. *Journal of Phonetics*, 12, 61-73.
- Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech*, 28(2), 97-114.
- Recasens, D. (1999). Lingual coarticulation. In W. J. a. H. Hardcastle, N. (Ed.), *Coarticulation*. Cambridge: Cambridge University Press.
- Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *Journal of the Acoustical Society of America*, 111, 2828-2841.
- Recasens, D. (2004). The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters). *Journal of Phonetics*, 32 435-453.
- Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America (JASA)*, 102, 544-561.
- Recasens, D., & Espinosa, A. (2006). Articulatory, positional and contextual characteristics of palatal consonants: Evidence from Majorcan Catalan. *Journal of Phonetics*, 34(iii).
- Ribeiro, C. M. (2003). *Produção de Fala. On Processamento digital de fala*. Lisboa.
- Rochette, C. (1973). *Les groupes de consonnes en Français*. Québec.
- Rossato, S., Teixeira, A., & Ferreira, L. (2006). *Les nasales du Portugais et du Français: une étude comparative sur les données EMMA*. Proc XXVI^{es} Journées d'études sur la parole, Dinard.
- Sá-Nogueira, R (1938). *Elementos para um tratado de Fonética Portuguesa*. Imprensa Nacional. Lisboa.
- Scheffler, K., & Lehnhardt, S. (2003). Principles and applications of balanced SSFP techniques *European Radiology* 13(11), 2409-2418.
- Seeley, R., Stephens, T., & Tate, P. (2005). *Anatomia e Fisiologia* (6ed.): Lusociência - Edições Técnicas e Científicas Lda.

- Serrurier, A., & Badin, P. (2005a). *A Three-Dimensional Linear Articulatory Model of Velum Based on MRI Data*. Proc at the Interspeech. Lisbon.
- Serrurier, A., & Badin, P. (2005b). Towards a 3D articulatory model of the velum based on MRI and CT images. *ZAS Papers in Linguistics*, 40, 195-211.
- Shadle, C. H., Mohammad, M., Carter, J. N., & Jackson, P. J. B. (1999). Multi-planar dynamic magnetic resonance imaging: new tools for speech research. *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS 99)*, 623-626.
- Shellock, F. (2002). Magnetic Resonance Safety update 2002: Implants and devices. *J Magnetic resonance Imaging*(16), 485-496.
- Shiga, Y. (2005). *Precise Estimation of Vocal Tract and Voice Source Characteristics*. Tese de doutoramento, University of Edinburgh, Edinburgh.
- Siemens. (2003). *Spins, Magnets and Resonances - An introduction to the basics of Magnetic Resonance*. Erlanger.
- Siemens. (2004). *Motion, Flows and Artifacts*. Erlanger.
- Silva, A. (2003). Formação de Imagem em Ressonância Magnética. On *Tecnologias de Imagem III*. Aveiro: Universidade de Aveiro.
- Silva, J. S. S. d. (2005). *Segmentação Pulmonar em Estudos de Tomografia Axial Computorizada*. Tese de Doutoramento, Universidade de Aveiro, Aveiro.
- Simões, F. (1999). *Implementação de um sistema de conversão texto-fala para o Português do Brasil*. Tese de Mestrado, Universidade Estadual de Campinas, Campinas.
- Sodickson, D. (2006). Parallel Imaging Methods. In R. Edelman, J. Hesselink, M. Zlatkin & J. Crues (Eds.), *Clinical Magnetic Resonance Imaging* (Third Edition ed., Vol. 1). Philadelphia: Elsevier Inc.
- Soquet, A., Lecuit, V., Metens, T., & Demolin, D. (2002). Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI. *Speech Communication*, 36, 169-180.
- Stevens, K. N., & Öhman, S. E. G. (1963). *Cineradiographic studies of speech* (Quarterly Progress and Status Report No. 2/63). Stockholm: Speech Transmission Laboratory.
- Stone, M. (2005). A Guide to Analyzing Tongue Motion from Ultrasound Images. *Clinical linguistics and phonetics*, 19(6-7), 455-502.
- Stone, M. (1999). Laboratory Techniques for Investigating Speech Articulation. In W. H. a. J. Laver (Ed.), *The Handbook of Phonetic Sciences* (pp. 11-32): Blackwell.
- Stone, M., & Lundberg, A. J. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America (JASA)*, 99(6), 3728-3737.
- Story, B. H. (1995). *Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract*. Tese de Doutoramento. University of Iowa, Iowa.
- Story, B. H., & Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. *Journal of the Acoustical Society of America (JASA)*, 97(2), 1249-1260.

- Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America (JASA)*, 100(1), 537-554.
- Stevens, P. (1954). Some observations on the phonetics and pronunciation of modern Portuguese. *Revista do Laboratório de Fonética Experimental de Coimbra II*, 5-29.
- Takemoto, H., Honda, K., Masaki, S., Shimada, Y., & Fujimoto, I. (2003). *Measurement of temporal changes in vocal tract area function during a continuous vowel sequence using a 3D cine-MRI technique*. Proc. of the 6th Seminar on Speech Production.
- Takemoto, H., Kitamura, T., Nishimoto, H., & Honda, K. (2004). A method of tooth superimposition of MRI data for accurate measurement of vocal tract shape and dimensions. *Acoustical Science and Technology*, 25(6), 468-474.
- Teixeira, A. (2000). *Síntese articulatória das vogais nasais do Português Europeu*. Tese de Doutoramento. Universidade de Aveiro, Aveiro.
- Teixeira, A. (2003). Speech Synthesis. NATO RTO EN Educational Notes.
- Teixeira, A., Moutinho, L. C., & Coimbra, R. L. (2003b). Production, acoustic and perceptual studies on European Portuguese nasal vowels height. Proc. of the 15th International Congress of Phonetic Sciences (ICPhS).
- Teixeira, A., Ferreira, L., Lisboa, R., Coimbra, R. L., & Moutinho, L.C. (2003a). *Novo Corpus acústico para Estudo das Vogais Nasais do Português Europeu Contemplando Variedades Regionais e contextuais* (Relatório Técnico No. FAPL 1/2003). Aveiro: CLC/IEETA, Universidade de Aveiro.
- Teixeira, L. (2003). Tomografia Computorizada. In F. M. Gaivão (Ed.), *Imagiologia Clínica-Princípios e técnicas*. Coimbra: Serviço de Imagiologia-Hospitais da Universidade de Coimbra.
- Teixeira, A., Martinez, R., Jesus, L. M. T., & Vaz, F. (2004a). Articulatory speech synthesis of Portuguese. Proc International Workshop in Phonetics Dedicated to the Memory of Farkas Kempelen, Budapest, Hungary.
- Teixeira, A., Ferreira, L., Moutinho, L.C., Coimbra, R. L., & Lisboa, R. (2004b). *An Acoustic Corpus Contemplating Regional Variation for Studies of European Portuguese Nasals*. Proc. of the Fourth International Conference on Language Resources and Evaluation (LREC).
- Teixeira, A., Martinez, R., Silva, L. N., Jesus, L. M. T., Príncipe, J. C., & Vaz, F. (2005). Simulation of Human Speech Production Applied to the Study and Synthesis of European Portuguese. *EURASIP Journal on Applied Signal Processing*, 2005(9), 1435-1448.
- Teixeira, A., Moutinho, L., & Coimbra, R. L. (2001). Acerca das vogais nasais do Português Europeu. *Revista da Universidade de Aveiro-Letras*(18), 241-274.
- Teixeira, A., & Vaz, F. (2001). European Portuguese Nasal Vowels: An EMMA Study.
- Thimm, G., & Luetin, J. (1999). *Extraction of articulators in X-Ray image sequences*. Proceedings of the European Conference on Speech Communication and Technology.
- Tiede, M., Masaki, S., & Vatikiotis-Bateson, E. (2000). *Contrasts in speech articulation observed in sitting and supine conditions*. Proc the 5th Seminar on Speech Production.

- Tom, K., Titze, I., Hoffman, E., & Story, B. (1999). 3-D Vocal Tract Imaging and Formant Structure: Varying Vocal Register, Pitch and Loudness. *NCVS Status and Progress Report*(14), 101-113.
- Tuller, E., Harris, K. S., & Gross, R. (1981). Electromyographic study of the jaw muscles during speech. *Journal of Phonetics*, 9, 175-188.
- Veloso, J. (1999). *Na Ponta da Língua: Exercícios de Fonética do Português*. Porto: Granito.
- Viana, M. (2001). Síntese de fala. In E. M. Ranchhod (Ed.), *Tratamento das línguas por computador- Uma introdução à linguística Computacional e suas aplicações: Caminho - Série Linguística*.
- Weissleder, R. (1999). Molecular Imaging: Exploring the Next Frontier. *Radiology*(212), 609-614.
- Westbrook, C., & Kaut, C. (1998). *MRI in Practice* (2nd ed.). Oxford: Blackwell Science.
- Whalen, D. H., Kang, A. M., Magen, H. S., Fulbright, R. K., & Gore, J. C. (1999). Predicting Midsagittal Pharynx Shape From Tongue Position During Vowel Production. *Journal of Speech, Language, and Hearing Research*, 42(3), 592-603.
- Wrench, A., McIntosh, A., Watson, C., & Hardcastle, W. J. (1998) Optopalatograph: Real time Feed-back of tongue movement in 3D: ICSP.
- Yang, B. (1999). *Measurement and synthesis of the vocal tract of Korean monophthongs by MRI*. Proc XIVth International Congress of Phonetic Sciences (ICPhS).
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., et al. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved
- Zemlin, W. R. (2000). *Princípios de Anatomia e Fisiologia em Fonoaudiologia* (Fourth ed.). Porto Alegre: ARTMED.
- Zierdt, A., Hoole, P., Honda, M., Kaburagi, T., & Tillmann, H. (2000). Extracting tongues from moving heads. Proc 5th Seminar on Speech Production, Germany.
- Zierdt, A., Hoole, P., & Tillmann, H. (1999). Development of a System for Three-Dimensional Fleshpoint Measurement of Speech Movements. Proc XIVth ICPhS.

Referências Eletrônicas

<http://webanatomy.net>

<http://www.idiap.ch/pages/contenuTxt/Demos/demo6/xray.html>

<http://www.linguistics.ucla.edu/faciliti/facilities/physiology/epg.html>

<http://www.medsch.wisc.edu/ubeam/gen/generator.html>

<http://www.speech.umarylnad.edu>

<https://mywebpages.comcast.net>

www.articulograph.com

www.cis.rit.edu/htbooks/mri

www.gbmc.org/voice

www.ph.ludwigbsurg.de.htm

www.phon.ucl.ac.uk

www.phon.ucl.ac.uk/Sampa/port.htm/

www.speechmaryland

Anexo I

a) International Phonetic Alphabet

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC)

© 2005 IPA

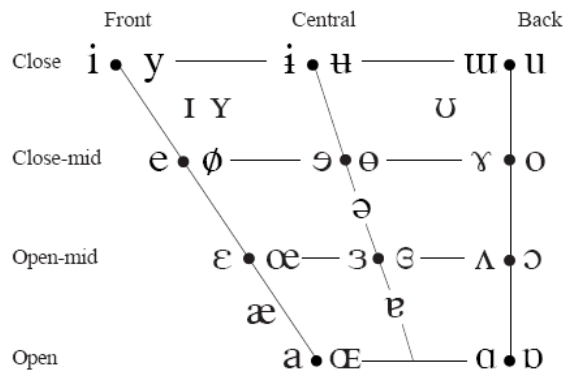
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ʀ					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
◌ǀ Bilabial	◌ɓ Bilabial	◌ʼ Examples:
◌ǃ Dental	◌ɗ Dental/alveolar	◌pʼ Bilabial
◌ǂ (Post)alveolar	◌ɟ Palatal	◌tʼ Dental/alveolar
◌ǁ Palatoalveolar	◌ɠ Velar	◌kʼ Velar
◌ǁ Alveolar lateral	◌ɣ Uvular	◌sʼ Alveolar fricative

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

b) SAMPA - Português Europeu

Consoantes

Oclusivas

Symbol	Word	Transcription
p	pai	pa j
b	barco	"barku
t	tenho	"teJu
d	doce	"dos@
k	com	ko~
g	grande	"gr6nd@

fricativas

f	falo	"falu
v	verde	"verd@
s	céu	sEw
z	casa	"kaz6
S	chapéu	S6"pEw
Z	jóia	"ZOj6

nasais

m	mar	mar
n	nada	"nad6
J	vinho	"viJu

líquidas

l	lanche	"l6nS@
L	trabalho	tr6"baLu
r	caro	"karu
R	rua	"Ru6

Vogais e ditongos

i	vinte	"vint@
	lápis	"lapiS
e	fazer	f6"zer
E	belo	"bElu
a	falo	"falu
6	cama	"k6m6
	madeira	m6"d6jr6
O	ontem	"Ont6~j~
o	lobo	"lobu
u	jus	ZuS
	futuro	fu"turu
@	felizes	f@"liz@S
i~	fim	fi~
e~	emprego	e~"pregu (or em-)
6~	irmã	ir"m6~
o~	bom	bo~
u~	um	u~
aw	mau	maw etc.: iw, ew, Ew, (ow)
aj	mais	majS etc.: ej, Ej, Oj, oj,
6~j~	têm	t6~j~ etc.: e~j~, o~j~, u~j~

Fonte: www.phon.ucl.ac.uk/Sampa/port.htm/

Anexo II

Corpus 2D e 3D

TIPO DE SOM	FONE	PALAVRA REFERÊNCIA	CODIFICAÇÃO	Nº ITEM
Vogal oral	[a]	pato	tse t1 2D vo - pato	1
	[ɒ]	cada	tse t1 2D vo - cada	2
	[@]	devi	tse t1 2D vo - devi	3
	[E]	leva	tse t1 2D vo - leva	4
	[e]	pêca	tse t1 2D vo -pêca	5
	[i]	pipo	tse t1 2D vo - pipo	6
	[O]	pote	tse t1 2D vo - pote	7
	[o]	tôpo	tse t1 2D vo - topo	8
	[u]	buda	tse t1 2D vo - buda	9
Vogal nasal	[6~]	canto	tse t1 2D vn - canto	10
	[e~]	pente	tse t1 2D vn - pente	11
	[i~]	pinta	tse t1 2D vn - pinta	12
	[o~]	ponte	tse t1 2D vn - ponte	13
	[u~]	punto	tse t1 2D vn - punto	14
	[6~]	manto	tse t1 2D vn - manto	15
	[i~]	minto	tse t1 2D vn - minto	16
Cons. nasais	[m]	cama	tse t1 2D cn - cama	17
	[n]	cana	tse t1 2D cn - cana	18
	[J]	canha	tse t1 2D cn - canha	19
Fricativas (início de palavra)	[f]	fala	tse t1 2D frip- fala	21
	[s]	sala	tse t1 2D frip-sala	22
	[S]	chá	tse t1 2D frip-chá	23
	[v]	vaca	tse t1 2D frip- vaca	24
	[z]	zarpa	tse t1 2D frip-Zarpa	25
	[Z]	jacto	tse t1 2D frip-jacto	26
Fricativa [f] V ₁ CV ₂ V ₁ =V ₂ =[i,u,a]	[f]	[afa]	tse t1 2D frc-afa	27
		[ifi]	tse t1 2D frc-ifi	28
		[ufu]	tse t1 2D frc-ufu	29
Fricativa [s]	[s]	[asa] “assa”	tse t1 2D frc-asa	30
		[isi] “issi”	tse t1 2D frc-isi	31
		[usu] “ussu”	tse t1 2D frc-usu	32
Fricativa [S]	[S]	[aSa]	tse t1 2D frc-aSa	33
		[iSi]	tse t1 2D frc-iSi	34
		[uSu]	tse t1 2D frc-uSu	35
Fricativa [v]	[v]	[ava]	tse t1 2D frc-ava	36
		[ivi]	tse t1 2D frc-ivi	37
		[uvu]	tse t1 2D frc-uvu	38
Fricativa [z]	[z]	[aza]	tse t1 2D frc-aza	39
		[izi]	tse t1 2D frc-izi	40
		[uzu]	tse t1 2D frc-uzu	41
Fricativa [Z]	[Z]	[aZa]	tse t1 2D frc-aZa	42
		[iZi]	tse t1 2D frc-iZi	43
		[uZu]	tse t1 2D frc-uZu	44
Oclusivas surdas V ₁ CV ₂ V ₁ =V ₂ =[i,u,a]	[p]	[apa]	tse t1 2D ocl s-apa	45
		[ipi]	tse t1 2D ocl s-ipi	46
		[upu]	tse t1 2D ocl s-upu	47
Oclusiva [t]	[t]	[ata]	tse t1 2D ocl s-ata	48

		[iti]	tse t1 2D ocl s-iti	49
		[utu]	tse t1 2D ocl s-utu	50
Oclusiva [k]	[k]	[aka]	tse t1 2D ocl s-aka	51
		[iki]	tse t1 2D ocl s-iki	52
		[uku]	tse t1 2D ocl s-uku	53
Oclusiva sonora [b]	[b]	[aba]	tse t1 2D ocl v-aba	54
		[ibi]	tse t1 2D ocl v-ibi	55
		[ubu]	tse t1 2D ocl v-ubu	56
Oclusiva sonora [d]	[d]	[ada]	tse t1 2D ocl v-ada	57
		[idi]	tse t1 2D ocl v-idi	58
		[udu]	tse t1 2D ocl v-udu	59
Oclusiva sonora [g]	[g]	[aga]	tse t1 2D ocl v-aga	60
		[igi]	tse t1 2D ocl v-igi	61
		[ugu]	tse t1 2D ocl v-ugu	62
Laterais	[l]	laço	tse t1 2D lat-l - laço	63
	[L]	falha	tse t1 2D lat L - falha	64

AnexoII Tabela 1 - Descrição completa do corpus 2D, com codificação.

TIPO DE SOM	FONE	PALAVRA REFERÊNCIA	CODIFICAÇÃO	Nº ITEM
Vogal oral	[a]	pato	3D_vo_pato	1
	[6]	cada	3D_vo_cada	2
	[i]	pipo	3D_vo_pipo	3
	[u]	buda	3D_vo_buda	4
	[o]	tôpo	3D_vo_topo	5
	[O]	Pote	3D_vo_pote	6
	[e]	pêca	3D_vo_peca	7
Vogal nasal	[6~]	manto	3D_vn_manto	8
	[6~]	canto	3D_vn_canto	9
	[e~]	Pente	3D_vn_pente	10
	[i~]	minto	3D_vn_minto	11
	[i~]	pinto	3D_vn_pinto	12
	[o~]	ponte	3D_vn_ponte	13
	[u~]	punto	3D_vn_punto	14
Consoante nasal	[m]	cama	3D_cn_cama	15
	[n]	cana	3D_cn_cana	16
	[J]	canha	3D_cn_canha	17
Fricativa [f] V ₁ CV ₂ V ₁ =V ₂ =[i,u,a]	[f]	[afa]	3D_fric_afa	18
		[ifi]	3D_fric_ifi	19
		[ufu]	3D_fric_ufu	20
Fricativa [s]	[s]	[asa]	3D_fric_asa	21
		[isi]	3D_fric_isi	22
		[usu]	3D_fric_usu	23
Fricativa [S]	[S]	[aSa]	3D_fric_aSa	24
		[iSi]	3D_fric_iSi	25
		[uSu]	3D_fric_uSu	26
Laterais	[l]	laço	3D_lat_laço	27
		pala	3D_lat_pala	28
		mal	3D_lat_mal	29
	[L]	palha	3D_lat_palha	30
		ralha	3D_lat_ralha	31

AnexoII Tabela 2 - Descrição completa do corpus 3D, com codificação.