



**INÊS
CAMPOS
MONTEIRO
SABINO
DOMINGUES
CARBONE**

**Segmentação do Tracto Vocal a partir de Estudos
Imagiológicos de Ressonância Magnética.**



**INÊS
CAMPOS
MONTEIRO
SABINO
DOMINGUES
CARBONE**

**Segmentação do Tracto Vocal a partir de Estudos
Imagiológicos de Ressonância Magnética.**

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

Dissertação realizada com o apoio financeiro da FCT no âmbito do III Quadro Comunitário de Apoio.

O júri

Presidente

Professora Doutora Maria Beatriz Alves de Sousa Santos
Professora Associada com Agregação da Universidade de Aveiro

Vogais

Professor Doutor Jorge Alves da Silva
Professor Auxiliar da Faculdade de Engenharia da Universidade do Porto

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

Durante os dois anos do desenvolvimento desta tese experimentei um grande desenvolvimento não só a nível profissional, mas também a nível pessoal. Este crescimento não teria sido possível sem o apoio e envolvimento de inúmeras pessoas e instituições. Na impossibilidade de as nomear todas, irei apenas mencionar os nomes mais significativos.

Começo por referir as pessoas ligadas ao progresso científico da tese. Aos meus orientadores, Professor Doutor António Joaquim da Silva Teixeira e Professor Doutor Augusto Marques Ferreira da Silva, que pelas suas personalidades e conhecimentos complementares, fizeram com que este trabalho se desenvolvesse de forma fluida e sem grandes sobressaltos. O primeiro com o seu espírito sempre positivo e incentivador e o segundo mais reservado e até desconfiado, e que me fez ter que provar as minhas capacidades.

À Paula Martins, colega e amiga, pelos vários momentos de desabafo e troca de experiências semelhantes. Também pelo companheirismo e mútua ajuda.

Aos colegas e professores de mestrado, em especial aos do primeiro semestre do ano curricular. Talvez por termos sido poucos, desenvolveu-se uma relação bastante próxima. Listando os nomes, Professora Beatriz, Professor Joaquim, Professor Paulo, Sérgio, Nuno, Luís e Samuel. Um especial agradecimento aos dois últimos porque, além de colegas de mestrado, foram também colegas de trabalho e uma grande ajuda nos momentos de necessidade.

Ao apoio financeiro fornecido pela FCT, permitindo-me usufruir de uma bolsa de iniciação à investigação científica durante dois anos através do projecto "HERON – Um ambiente computacional para investigação em síntese articulatória do Português" com o número POSI/PLP/57680/2004. Agradeço também ao IEETA, que me aliviou o pagamento do segundo ano de propinas.

Uma pessoa que tenho dificuldade em enquadrar no âmbito científico ou pessoal é o primo Pedro Serrano. Foi uma ajuda indirecta mas essencial, especialmente por me ter oferecido o seu livro, que me ajudou a esclarecer bastantes dúvidas sobre este mundo, ainda um pouco estranho para mim, da investigação científica.

Passando ao plano pessoal, começo por falar nos meus queridos pais. Não só por me terem financiado todos os meus estudos, incluindo a parte curricular do mestrado, mas também pelo seu apoio incondicional e incentivo nos momentos mais desesperados. Não posso deixar de mencionar também as correcções, essencialmente ortográficas, por eles efectuadas em todos os meus trabalhos.

À minha mana, simplesmente por ser como é!

A toda a minha família e amigos porque sei que, embora nem sempre fisicamente presentes, estão sempre lá. Sem essa certeza não teria conseguido a paz interior para me dedicar a este trabalho tão absorvente.

Um beijinho especial ao Ricardo, tio Manuel, tio Quim, avô Rui, pai e Fabio pelos comentários e opiniões às primeiras versões desta tese.

Ao meu maridinho Fabio, pelo amor, carinho, companheirismo e paciência demonstrados ao longo deste percurso. A tua presença deu-me a força necessária para levar este projecto até ao fim. As suas 'ideias, descobertas, críticas, sugestões e inspiração (...) são parte inseparável deste livro. É impossível agradecer-lhe a sua contribuição.' (Citando António Damásio, em O Erro de Descartes).

Palavras-chave

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

Resumo

Na área da imagiologia é frequentemente necessário decompor as imagens ou volumes nos seus diversos componentes estruturais, isto é, há necessidade de "segmentar" as imagens. Existe hoje um amplo leque de técnicas de segmentação de imagens, proporcionando níveis de desempenho e eficácia contextualmente variáveis. O objectivo geral deste trabalho é a exploração das potencialidades das técnicas de segmentação aplicada ao estudo da produção da fala no Português Europeu.

Através da segmentação de imagens e volumes de Ressonância Magnética duma das principais estruturas anatómicas relacionadas com a produção da fala, o tracto vocal, pretendeu-se alcançar o objectivo específico de proporcionar novas informações de teor morfo-funcional no conhecimento dos mecanismos de produção do Português Europeu.

A investigação encontra-se inserida num projecto mais amplo e pluridisciplinar denominado HERON, cujo objectivo é a criação dum programa de síntese articulatória para o Português Europeu.

Após a aquisição de imagens 2D e 3D do tracto vocal de dois falantes, foram seleccionadas uma série de técnicas de segmentação a aplicar, nomeadamente *Region Growing*, *Live Wire*, *Level Sets*, e *Reaction Diffusion Active Contours*, tendo-se avaliado as vantagens e desvantagens de cada método consoante a natureza dos dados imagiológicos e o objectivo específico da segmentação.

O método *Region Growing* provou ser eficaz no caso mais simples da segmentação 2D mas em contextos 3D foi necessário recorrer a métodos mais elaborados como os *Level Sets*. Foi ainda proposta uma estratégia inovadora de segmentação semi-automática "2.5D" baseada em reformatação multi-planar ao longo do eixo longitudinal do tracto vocal seguida da aplicação de algoritmos *Live Wire* ou *Level Sets*. Esta metodologia provou viabilizar as aquisições de Ressonância Magnética puramente 3D para os estudos imagiológicos de produção do Português Europeu.

keywords

Segmentation, MRI, Speech production, European Portuguese, Articulatory synthesis

abstract

The main objective of the segmentation techniques is the partition of images or volumes in its structural components. Although its use is of basic importance in the image processing field, the effectiveness and homogeneity of the segmentation techniques still represent a problem in the area of applied engineering whose mathematical solution is yet to be found.

This work focuses on the segmentation of Magnetic Resonance images and volumes of some anatomical structures related with speech production, namely, the vocal tract. Through the segmentation we were able to gather new morph-functional information on the European Portuguese production.

This investigation is part of a broad project, named HERON, whose objective is the development of a speech synthesizer for European Portuguese.

After the comparative analysis of an extensive set of segmentation methods, the ones that were more adequate for our application were chosen. The implemented techniques were: Region Growing; Live Wire; Level Sets; and Reaction Diffusion Active contours. The advantages and disadvantages of each method were also evaluated.

Region Growing proved capable of segmenting the 2D images. For 3D information, the use of more complex methodologies, such as Level Sets, was necessary. A new "2.5D" segmentation strategy was also proposed, based on multi-planar reformatting through the longitudinal axes of the vocal tract, followed by the application of 2D segmentation algorithms as Live Wire and Level Sets. This methodology enabled the 3D Magnetic Resonance acquisitions for imaging studies in the European Portuguese production.

Conteúdo

Lista de Figuras	iii
Lista de Tabelas	vii
Lista de Algoritmos	ix
Lista de Abreviaturas	xiii
Lista de Símbolos	xv
1 Introdução	1
1.1 Estrutura da dissertação	2
1.2 Resultados publicados e submetidos	3
2 Produção de fala e imagem por RM	7
2.1 Estrutura deste Capítulo	7
2.2 Fonética e sons do Português Europeu	8
2.2.1 A Fonética	8
2.2.2 Aparelho fonador e produção da fala	8
2.2.3 Sons do Português Europeu	10
2.2.4 Alfabetos fonéticos	13
2.3 Métodos de recolha de informação sobre a produção de fala	13
2.3.1 Ressonância Magnética	15
2.4 Sequências escolhidas para a aquisição dos dados	23
2.5 Caracterização dos dados (imagens e volumes) adquiridos	26
2.6 Considerações finais	27
3 Segmentação de imagem	29
3.1 Enquadramento geral	29
3.2 Pré-Processamento	30
3.2.1 Delimitação do tracto vocal	30
3.2.2 Atenuação de ruído	31
3.3 Segmentação manual	37
3.4 Segmentação semi-automática	38
3.4.1 Limiarização (<i>Threshold</i>) global	38

3.4.2	Limiarização <i>zoom</i>	38
3.4.3	Limiarização adaptativa	38
3.4.4	Detecção de arestas	39
3.4.5	<i>Region Growing</i>	41
3.4.6	<i>Live Wire (Intelligent Scissors)</i>	42
3.4.7	Contornos Activos	43
3.4.8	<i>Reaction Diffusion Active Contours</i>	57
3.5	Segmentação automática	59
3.6	Métricas para comparação entre contornos	59
3.7	Segmentação em estudos de produção de fala	61
3.8	Comparação das classificações adoptadas com outras presentes na literatura	64
3.9	Considerações finais	66
4	Trabalho experimental	67
4.1	Base de Dados de Ressonância Magnética para o Português Europeu	67
4.1.1	<i>Corpus</i> de Ressonância Magnética	67
4.1.2	Aquisição das imagens	69
4.2	Métodos de processamento de imagem escolhidos	71
4.2.1	Pré-Processamento	71
4.2.2	Segmentação de Imagens e Volumes	73
4.3	Considerações finais	80
5	Resultados	83
5.1	Pré-Processamento	83
5.2	Segmentação de imagens (2D)	85
5.2.1	<i>Region Growing</i>	85
5.2.2	<i>Level Sets</i>	89
5.2.3	<i>Reaction Diffusion</i>	92
5.2.4	Comparação entre as três técnicas de segmentação 2D	92
5.3	Segmentação de imagens de Tempo Real	94
5.3.1	Segmentação manual	95
5.3.2	<i>Region Growing</i>	95
5.3.3	<i>Level Sets</i>	96
5.3.4	<i>Reaction Diffusion</i>	96
5.3.5	Comparação entre as quatro técnicas de segmentação para as imagens de Tempo Real	96
5.4	Segmentação de volumes (3D)	97
5.4.1	Segmentação $2\frac{1}{2}$ D	99
5.4.2	<i>Level Sets</i> 3D	100
5.4.3	Comparação entre as três técnicas de segmentação para os volumes	102
5.5	Discussão	103
5.6	Considerações finais	104

6 Conclusões	107
6.1 Síntese do trabalho desenvolvido	107
6.2 Principais resultados	108
6.3 Trabalho futuro	110
6.4 Considerações finais	111
Bibliografia	111

Lista de Figuras

1.1	Diagrama da evolução do projecto HERON	2
2.1	Aparelho fonador	9
2.2	Diferenças entre um som oral e um nasal	11
2.3	Classificação das vogais segundo a zona de articulação	11
2.4	Classificação das vogais segundo o grau de abertura	11
2.5	Classificação das consoantes segundo o modo de articulação	12
2.6	Classificação das consoantes segundo o ponto de articulação	12
2.7	Representação dos contornos das três vogais cardinais	13
2.8	Funcionamento da ressonância magnética	16
2.9	Representação gráfica do conceito de TR	17
2.10	Representação gráfica do conceito de TE	18
2.11	Representação gráfica do conceito de FA	18
2.12	Representação gráfica do conceito de IT	18
2.13	Representação gráfica do conceito de MX	19
2.14	Representação gráfica do conceito de FOV	19
2.15	Representação gráfica do conceito de ST	19
2.16	Representação gráfica do conceito de SG	19
2.17	Esquematisação da acção dos campos magnéticos associados à aquisição de imagens por RM	21
2.18	Codificação de imagens de RM	22
2.19	Organização do espaço K	23
2.20	Diagrama de pulso de uma sequência TSE	24
2.21	Diagrama de pulsos de uma sequência TrueFISP	25
2.22	Diagrama de pulsos de uma sequência EG Spoiled 2D - FLASH	26
2.23	Artefactos de <i>aliasing</i> 3D	27
3.1	Eliminação do ar exterior	31
3.2	Imagem após o pré-processamento	32
3.3	Filtro de mediana	32
3.4	Filtro de Difusão Anisotrópica	33
3.5	Filtro de <i>Wavelets</i>	35
3.6	<i>Wavelet</i> Haar	36
3.7	<i>Wavelets</i> Daubechies	36

3.8	Periféricos	37
3.9	Exemplo de aplicação de Limiarização	39
3.10	Exemplo de aplicação da Limiarização adaptativa	39
3.11	Operador de Canny	40
3.12	<i>Region Growing</i>	42
3.13	<i>Snake</i>	44
3.14	Ajuste e divisão de um contorno inicial a três regiões de interesse	46
3.15	Segmentação 3D do cérebro através da utilização de <i>Level Sets</i>	47
3.16	Exemplos de funções ϕ	47
3.17	Função de velocidade baseada na intensidade da imagem	49
3.18	Intensidades I_{bgr} e I_{obj}	53
3.19	Detecção de arestas pelo método de <i>Reaction Diffusion</i>	59
4.1	Siemens Magnetom Symphony	70
4.2	Frames [k], [a], [t] e [u]	70
4.3	Visualização de três planos dum volume	71
4.4	Histogramas de algumas imagens RM 2D	74
4.5	Máscara de pós-processamento	75
4.6	Eliminação da sobre-segmentação	76
4.7	Cálculo da grelha adaptativa	78
4.8	Imagem médio-sagital e respectiva grelha adaptativa	79
4.9	Reformatações dum volume	80
5.1	Atenuação de ruído	84
5.2	Efeito da atenuação de ruído em duas linhas de cada imagem	84
5.3	Cálculo da média dos contornos	87
5.4	Média de 100 contornos gerados com o <i>Region Growing</i> 2D	88
5.5	Robustez do <i>Region Growing</i> 2D em relação à posição da semente	88
5.6	Estudo local das zonas com maior variabilidade entre contornos	88
5.7	Avaliação do IP como caracterizador dos sons	89
5.8	Inicialização com um ponto no interior do tracto vocal	89
5.9	Inicialização com uma linha	89
5.10	Inicialização com um círculo	90
5.11	Inicialização com 100 círculos	90
5.12	Segmentação por <i>Level Sets</i> com diferentes funções de velocidade	91
5.13	Segmentação por <i>Level Sets</i> após o pós-processamento	91
5.14	Segmentação por <i>Reaction Diffusion</i>	92
5.15	Segmentação por <i>Reaction Diffusion</i> após o pós-processamento	92
5.16	Comparação entre <i>Reaction Diffusion</i> , <i>Level Sets</i> e <i>Region Growing</i>	93
5.17	Comparação entre segmentações com <i>Region Growing</i> e <i>Level Sets</i>	93
5.18	Comparação de segmentações entre <i>Level Sets</i> e <i>Reaction Diffusion</i>	94
5.19	Comparação de segmentações entre <i>Reaction Diffusion</i> e <i>Region Growing</i>	95

5.20	Segmentação manual dos frames [k], [a], [τ] e [u]	95
5.21	Segmentação por <i>Region Growing</i> 2D após a atenuação de ruído	96
5.22	Comparação entre os vários resultados obtidos para as imagens em Tempo Real	98
5.23	Segmentação por <i>Level Sets</i> após a aplicação do Filtro de Gaussiana às imagens de Tempo Real	99
5.24	Segmentação por <i>Reaction Diffusion</i> após a aplicação do Filtro de Gaussiana às imagens de Tempo Real	99
5.25	Segmentação 2½D por <i>Live Wire</i>	99
5.26	Aplicação de uma <i>mesh</i> à segmentação 2½D por <i>Live Wire</i>	100
5.27	Segmentação 2½D por <i>Level Sets</i>	100
5.28	Aplicação de uma <i>mesh</i> à segmentação 2½D por <i>Level Sets</i>	101
5.29	Efeitos do pós-processamento na segmentação 3D efectuada por <i>Level Sets</i>	101
5.30	Visualização da segmentação 3D por <i>Level Sets</i> das três vogais cardinais	102
5.31	Comparação entre as funções de área obtidas com a segmentação 2½D e com a segmentação 3D	103

Lista de Tabelas

2.1	Comparação entre várias técnicas de aquisição de informação sobre o tracto vocal	14
2.2	Vantagens e desvantagens da utilização da RM	17
3.1	Regras da decisão <i>fuzzy</i>	52
4.1	<i>Corpus</i> 2D	68
4.2	<i>Corpus</i> Tempo Real	68
4.3	<i>Corpus</i> 3D	69
5.1	Avaliação do pré-processamento para as imagens de Tempo Real	85
5.2	Avaliação das técnicas de segmentação	105

Lista de Algoritmos

1	Detector de arestas Canny	40
2	<i>Region Growing</i>	41
3	<i>Level Sets</i>	55
4	<i>check_{in}</i>	56
5	<i>check_{out}</i>	56

Lista de Abreviaturas

AD	<i>Anisotropic Diffusion</i> (Difusão Anisotrópica)
BW	<i>Band Width</i> (largura de banda)
EG	Sequência Eco de Gradiente
EMMA	<i>Electro Magnetic Mid-sagittal Articulator</i>
EPG	Electropalatografia
ES	Sequência Eco de Spin
ETL	<i>Echo Train Length</i>
FA	<i>Flip Angle</i> (ângulo de nutação)
FCT	Fundação para a Ciência e Tecnologia
FG	Filtro de Gaussiana
FID	<i>Free Induction Decay</i> (decaimento de indutância livre)
FLASH	<i>Fast Low Angle Shot</i>
FM	Filtro de mediana
FOV	<i>Field Of View</i> (Campo de Visão)
FSE	Sequência Fast Spin Eco
HERON	Projecto com a referência POSC/PLP/57680/2004, financiado pela FCT
IO	Imagem Original
IP	Índice de Pratt
IT	<i>Inversion Time</i> (Tempo de Inversão)
MRI	Magnetic Resonance Imaging
MSQE	<i>Mean Square Error</i> (erro quadrático médio)
MX	Matriz
NA	<i>Number of Acquisitions</i> (Número de aquisições)
NEX	<i>Number of Excitations</i>
nMSQE	<i>normalized Mean Square Error</i> (erro quadrático médio normalizado)
NSDE	<i>Normalized Step Difference Energy</i>
PE	Português Europeu
PSNR	<i>Peak Signal to Noise Ratio</i>
RARE	<i>Rapid Acquisition Relaxation Enhancement</i>
RF	Onda de radiofrequência
RM	Ressonância Magnética
SAP	Síntese articulatória do Português
SAR	<i>Specific Absorption Rate</i>

SE	Sequência Spin Eco
SG	<i>Slice Gap</i> (Intervalo entre cortes)
SNG	Função que retorna o sinal (positivo ou negativo)
SNR	<i>Signal to Noise Ratio</i>
ST	<i>Slice Thickness</i> (Espessura de corte)
STE	Eco estimulado
Sub-índices	Derivadas, indicando a sua ordem e parâmetro de derivação
SVH	Sistema Visual Humano
T	Limiar (<i>Threshold</i>)
TA	Tempo de Aquisição
TC	Tomografia Computorizada
TE	Tempo de Eco
TR	Tempo de Repetição
True FISP	<i>True Fast Imaging with Steady Precession</i>
TSE	Sequência Turbo Spin Eco
WD	<i>Wavelet Denoising</i>

Lista de Símbolos

(x, y)	coordenadas de cada ponto da imagem de RM
α	constante real positiva
α_T	limiar do termo de propagação ν
β	constante real positiva
ϵk	função de velocidade da curvatura
ϵ	controla o intervalo de valores próximos de T que devem ser considerados no interior da região a segmentar
γ	Coefficiente de amortecimento
γ_1	$I(x + 1, y)$
γ_2	$I(x - 1, y)$
γ_3	$I(x, y + 1)$
γ_4	$I(x, y - 1)$
λ	constante real positiva
∇	gradiente
$\nabla_x D_i$	componente x do gradiente do mapa de distâncias D_i
$\nabla_y D_i$	componente y do gradiente do mapa de distâncias D_i
ν	termo de propagação
Ω_1	conjuntos de pixels dentro do contorno C
Ω_2	conjuntos de pixels fora do contorno C
ϕ	função de uma dimensão superior à da curva C (<i>Level Set</i>)
σ^2	constante positiva
\vec{N}	normal à curva
\wedge	operador “e”
\tilde{I}_{th}	valor estimado de I_{th}
$ \cdot $	valor absoluto
$ G_\phi $	magnitude do gradiente na direcção da superfície do <i>Level Set</i>
A	matriz onde estão representadas as forças internas da <i>snake</i>
$a(x, y)$	média das intensidades da imagem numa vizinhança (pequena) de (x, y)
B_0	campo magnético externo
C	curva
$C(i, j)$	pixels no contorno C
C_2	conjunto de pixels no contorno automático
$d(x, y, t)$	função de difusão

D_i	mapa de distâncias
d_i^2	distâncias entre dois pontos correspondentes
det	detector de arestas, detector de regiões, ou detector de arestas e regiões
div	operador de divergência
DP	densidade de protões
F	conjunto das forças externas
f	saída de um sistema de decisão <i>fuzzy</i>
F_v	função de velocidade
g	função de paragem
$G(I)$	distribuição Gaussiana
G_ϕ	magnitude do gradiente “alvo”
h	passo de discretização
I	intensidade da imagem de RM
I'	imagem após o processo de atenuação de ruído
I_{bgr}	média da intensidade do fundo na vizinhança do contorno
I_{obj}	média da intensidade do objecto de interesse na vizinhança do contorno
I_{th}	limiar da força de propagação em Baillard et al. (2000)
Id	matriz identidade
K	constante de difusão
k	curvatura local
L_{in}	lista dos pontos pertencentes ao interior
L_{out}	lista dos pontos pertencentes ao exterior
M_{tot}	magnetização total
N	número de pares de pontos correspondentes
$n(A)$	cardinal de A
$N(x)$	vizinhança de x
N_g	tamanho do filtro Gaussiano
P	indicador da região
$P(I)$	função normalizada de probabilidade
$p(x_s)$	probabilidade no contorno
$p(y_s x_s)$	probabilidade da região a segmentar
P_G	distribuição Gaussiana modificada
P_T	probabilidade de transição entre o interior e o exterior da estrutura a segmentar
s	variável da parametrização
S_{area}	similaridade de áreas
S_{forma}	similaridade de formas
S_{p-p}	pico da amplitude do sinal
T	controla o brilho da região a segmentar
t	tempo
V	coordenadas dos pontos da <i>Snake</i>
V_{ext}	função de velocidade do gradiente
V_p	função de velocidade regional

Capítulo 1

Introdução

Define-se segmentação de uma imagem como a sua partição em regiões, tendo cada uma das quais um conjunto de características comuns. O processo de segmentação - que se pode realizar de forma manual, semi-automática ou automática - pode ser utilizado para vários fins. Contudo, existem algumas limitações intrínsecas à implementação da grande maioria dos processos de segmentação existentes, tais como o tempo necessário para a sua execução e o baixo nível de precisão dos resultados obtidos. Existem variados métodos de segmentação e há na literatura, uma acesa discussão acerca deste assunto.

O presente trabalho tem como objectivo principal a aplicação de técnicas de segmentação a imagens do tracto vocal humano obtidas através de ressonância magnética (RM), com o intuito final de proporcionar dados para a melhoria do sintetizador de fala do projecto “SAP - Síntese Articulatoria do Português” (Teixeira e Vaz, 1999), um sintetizador articulatorio especialmente vocacionado para a síntese do Português que tem como principal objectivo melhorar o conhecimento dos sons nasais. Foi já modelada a interacção entre a fonte de excitação glotal e as cavidades acima da laringe, e o efeito da dinâmica dos articuladores (Teixeira et al., 2001). No entanto, para desenvolver e completar este sintetizador, são necessários dados reais. É nesta etapa que a presente dissertação se insere. Foram previamente analisados para o efeito, os métodos de segmentação mais utilizados na área, procedendo-se à selecção dos mais apropriados para este fim.

Esta investigação tenciona ser um válido contributo para um maior conhecimento dos mecanismos de produção da fala, através da aplicação das técnicas de segmentação de imagem actualmente conhecidas e mais utilizadas. As segmentações obtidas irão ser utilizadas futuramente para a extracção de parâmetros articulatorios.

O interesse cada vez maior por este estudo é motivado pelas aplicações práticas dos respectivos conhecimentos: menciona-se *in primis* a possibilidade de proporcionar novas oportunidades no campo médico, nomeadamente no que diz respeito à terapia da fala e à resolução de problemas ligados à falta ou escassez auditiva. Na área da aprendizagem da língua por estrangeiros, o conhecimento mais aprofundado dos mecanismos de produção de sons repre-

sentaria ainda um importante auxílio. O enquadramento deste trabalho e os conhecimentos através dele adquiridos podem também ser aplicados no aperfeiçoamento dos sintetizadores de voz já existentes. Estes sintetizadores representam um importante desafio, tendo em conta as simplificações que estes *softwares*, inseridos nas interfaces dos computadores, podem proporcionar na interacção entre homem e máquina, sobretudo para os utentes com limitações físicas.

A presente investigação encontra-se inserida no projecto HERON - Um ambiente computacional para Investigação Articulatoria do Português (Referência POSI / PLP / 57680 / 2004), aprovado pela FCT (Fundação para a Ciência e Tecnologia) em 2004. O objectivo final do projecto HERON é obter um programa de síntese articulatória para o Português, Figura 1.1. Para tal, a primeira tarefa é obter dados. É nesta tarefa que este trabalho se insere.



Figura 1.1: Diagrama da evolução do projecto HERON.

1.1 Estrutura da dissertação

Ao longo do percurso de investigação, tornou-se clara a complexidade da abordagem pluridisciplinar que o estudo necessita. O segundo Capítulo representa, neste sentido, o primeiro desafio. Nele resumem-se as noções fundamentais das duas principais áreas envolvidas: Fonética e Imagiologia. Se do ponto de vista da arquitectura da dissertação, a apresentação destas noções representa a indispensável introdução, para o leitor, ao trabalho específico; do ponto de vista do processo de investigação, representa a fase de aquisição e aprofundamento de uma série de conhecimentos novos para o autor. Destaca-se, na segunda parte deste segundo Capítulo, a abordagem às técnicas de aquisição de imagens do tracto vocal, tendo-se fundamentado a escolha, entre todas, da utilização da RM.

Se no Capítulo segundo são apresentadas as principais áreas que este estudo aborda, antecipando-se, na Secção referida à imagiologia, a escolha da técnica de aquisição de imagens e volumes por RM, o terceiro Capítulo é inteiramente dedicado à apresentação do método escolhido para o processamento das imagens adquiridas: a segmentação.

Neste Capítulo, a análise da literatura existente leva a concluir que, na imagiologia, a segmentação continua a ser um desafio aberto para os investigadores da área. Sobretudo no caso do processamento de imagens médicas, evidenciar contornos de partes anatómicas é uma tarefa que apresenta muitas dificuldades devido não só à complexidade dos contornos e à existência de artefactos nas imagens, mas também às limitações ligadas à aplicabilidade de métodos de aquisição de imagem pouco invasivos e que não proporcionem riscos para o paciente. Evidencia-se também as vantagens que um método de segmentação totalmente

automático, actualmente inexistente, poderia trazer, tendo em conta o crescente aumento de informação bruta que necessita de ser trabalhada de modo a se poder tirar informações importantes a partir dela.

O quarto Capítulo apresenta e fundamenta as escolhas metodológicas adoptadas para o fim proposto para este estudo, a extracção de informação de dados de RM, através da sua segmentação. Não se omite a descrição do aparelho utilizado para a realização da RM e apresenta-se a tipologia de imagens adquiridas (2D, Tempo Real e 3D), e o respectivo *Corpus* utilizado. Optou-se por um *Corpus* bastante abrangente, em detrimento do número de informantes. Isto levou à criação de uma primeira base de dados geral sobre os sons existentes na língua portuguesa.

O Capítulo quinto contém a descrição e discussão dos resultados obtidos através da implementação das diferentes técnicas de segmentação às imagens de RM adquiridas. O Capítulo termina com uma análise comparada da eficácia das diferentes técnicas utilizadas à luz das experiências realizadas.

O último Capítulo representa o epílogo da dissertação contendo uma síntese do trabalho desenvolvido e reflexões sobre os métodos adoptados, destacando-se os principais objectivos alcançados e, por fim, avançando propostas para trabalhos futuros.

1.2 Resultados publicados e submetidos

Durante o decorrer da presente investigação foram já publicados ou submetidos a publicação alguns documentos, listados de seguida:

- Trabalhos publicados ou aceites para publicação internacional:
 - Martins et al. (2007d) -> Martins, P., Carbone, I., Silva, A., e Teixeira, A. (2007d). An MRI study of European Portuguese nasals. Interspeech, Antwerp, BE.
Neste artigo foram apresentados estudos 2D e 3D dos dados relativos às nasais, em especial às vogais nasais. Os resultados 2D foram obtidos através da segmentação por *Region Growing* (apresentados na Secção 5.2.1 da presente dissertação) e os 3D foram obtidos com a técnica de segmentação $2\frac{1}{2}$ D com base no *live wire* 2D (Secção 5.4.1). Os resultados apresentados consistiram na comparação de contornos, comparação de funções de area e comparação de VPOQ (*Velum Port Opening Quotient* - conceito relacionado com a nasalização e proposto em (Engwall et al., 2006)).
 - Martins et al. (2007c) -> Martins, P., Carbone, I., Silva, A., e Teixeira, A. (2007c). MRI study of coarticulation in European Portuguese. 154th Meeting of the Acoustical Society of America, volume 122, página 3031. New Orleans, Louisiana.
Este póster consiste na compilação das inferências relativas à coarticulação no PE. O tratamento dos dados e os resultados apresentados são semelhantes aos já

descritos para o artigo (Martins et al., 2007d), com a diferença que são aplicados aos dados relativos à coarticulação.

- Carbone et al. (2007a) -> Carbone, I., Martins, P., Silva, A., e Teixeira, A. (2007a). Volumetric MRI acquisition and processing. 154th Meeting of the Acoustical Society of America, volume 122, página 3031. New Orleans, Louisiana.

Neste póster são apresentados resultados relativos ao tratamento da informação volumétrica. O tratamento dos volumes é o descrito na Secção 4.2.2 desta dissertação e os resultados são os apresentados na Secção 5.4.

- Martins et al. (2007b) -> Martins, P., Carbone, I., Silva, A., e Teixeira, A. (2007b). Coarticulation effects on European Portuguese: a first MRI study. Workshop AFCP, Montpellier.

Este artigo complementa o póster (Martins et al., 2007c), descrevendo com mais pormenor os resultados e ilações que podem ser tirados sobre a coarticulação no PE, através dos dados de RM adquiridos.

- Material submetido:

- Martins et al. (2007a) -> Martins, P., Carbone, I., Pinto, A., Silva, A., e Teixeira, A. (2007a). European Portuguese MRI based speech production studies. Speech Communication - Special issue on Iberian Languages.

Esta longa publicação compila toda a informação retirada do trabalho efectuado no processamento e interpretação dos dados RM adquiridos no contexto do projecto HERON. O processamento 2D é o descrito na Secção 5.2.1 e o 3D é o presente na Secção 5.4.1. Em termos de resultados são apresentados a localização de alguns pontos articulatorios de especial importância, contornos e sua comparação, visualizações 3D do tracto vocal, funções de área e suas comparações, e o VPOQ.

- Relatórios e publicações nacionais:

- Martins et al. (2006) -> Martins, P., Carbone, I., Teixeira, A., e Silva, A. (2006). Ressonância Magnética estática e dinâmica aplicada ao estudo da Fonética Articulatória do Português Europeu (PE). Projecto de investigação HERON. Relatório técnico. IEETA/Universidade de Aveiro.

Neste relatório, com o objectivo de justificar a aquisição do material de estudo, são descritas as vantagens da utilização da RM no estudo do PE, e possíveis resultados a ser obtidos com esse estudo. É ainda apresentado o *corpus* a adquirir e algumas propostas de sequências a ser utilizadas para a aquisição das imagens, tendo por base uma extensiva revisão bibliográfica da área (também apresentada neste relatório) e algumas sugestões pessoais da Técnica de Radiologia, e investigadora do projecto HERON, Paula Martins.

- Carbone et al. (2006) -> Carbone, I., Martins, P., Teixeira, A., e Silva, A. (2006). Base de Dados RM dos Sons do Português Europeu. Relatório técnico SPL 0605, IEETA/Universidade de Aveiro.

Este relatório teve como principal objectivo actualizar a informação presente em (Martins et al., 2006) após a aquisição dos dados. São descritas as sequências utilizadas para a aquisição dos dados, são apresentadas tabelas com todo o material adquirido (fone, palavra de referencia, codificação manual, e codificação DICOM) e alguma outra informação sobre a recolha, tal como as especificações dos informantes utilizados e os parâmetros das sequências. É ainda descrita a estrutura do DVD onde a informação ficou gravada.

- Carbone et al. (2007b) -> Carbone, I., Martins, P., Teixeira, A., e Silva, A. (2007b). Vocal tract segmentation and analysis over a European Portuguese MRI database. Revista periódica do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro (aceite para publicação).

Neste artigo irá ser apresentado um estudo da robustez do *Region Growing* 2D à localização da semente. É ainda proposta uma métrica - o Índice de Pratt - que poderá ser útil na classificação dos fones do PE. Este trabalho está descrito na Secção 5.2.1.

Capítulo 2

Produção de fala e imagem por Ressonância Magnética

Há milhares de anos os gregos, especuladores de tudo o que faz parte do universo, elaboraram o primeiro tratado conhecido relativo à linguística. No *Dialogo de Crátilo* da autoria de Platão, defendia-se a ideia de que numa época remota tinha existido um “legislador” que a tudo deu o nome correcto e natural. Tal tratado demonstra a grande curiosidade e interesse que o Homem desde sempre teve em explicar e analisar o fenómeno da linguagem e da fala. A razão está no facto de que estes fenómenos representam uma característica única dos seres humanos e directamente responsável pela sua evolução. A história da linguagem está intimamente ligada à história do Homem (Fromkin e Rodman, 1993). Ainda hoje, a linguagem, isto é, a capacidade de articular sons segundo esquemas pré-estabelecidos, é um dos alicerces da comunicação, logo, das relações entre indivíduos. A enorme importância desta característica especificamente humana leva ao interesse crescente pelo estudo aprofundado de determinados aspectos da linguagem e da fala, incluindo a possibilidade da sua reprodução de forma cada vez mais exacta, em sintetizadores para os mais diversos fins.

A presente investigação tem, entre as suas características principais, um carácter fortemente pluridisciplinar isto é, a integração de conhecimentos oriundos de diferentes domínios. Este Capítulo servirá para a contextualização deste estudo no âmbito das principais áreas em que este se encontra inserido, nomeadamente a área da Fonética e Sons do Português Europeu (PE), e a área de imagem, em particular da RM.

2.1 Estrutura deste Capítulo

Na primeira parte do Capítulo, após a explicação do funcionamento dos mecanismos anatómicos de produção de fala, serão expostas as classificações próprias dos sons da língua portuguesa. Na segunda parte, serão apresentadas as técnicas de estudo destes mecanismos, destacando a importância da análise das imagens do aparelho fonador. De seguida, abordar-se-ão as técnicas mais utilizadas de aquisição de imagem, para concluir focando a atenção na RM (enquanto técnica escolhida para a presente investigação) e seu funcionamento.

2.2 Fonética e sons do Português Europeu

A interpretação das imagens de RM, no contexto de produção de fala, pressupõe o conhecimento prévio do aparelho fonador e dos mecanismos envolvidos na produção. De seguida, serão expostas as noções necessárias para o enquadramento do presente trabalho no âmbito da Fonética e dos mecanismos de produção da fala.

2.2.1 A Fonética

A linguagem é uma das características dos seres humanos directamente responsável pela sua evolução. Os estudiosos mostram-se interessados em saber qual a relação entre a linguagem e o desenvolvimento evolucionário da espécie humana. Alguns encaram a linguagem como a diferença qualitativa entre os seres humanos e os primatas; outros encaram a capacidade de linguagem como um salto qualitativo. Ao tentarem compreender o desenvolvimento da linguagem, os estudiosos de todos os tempos têm-se debruçado sobre o papel desempenhado pelo aparelho fonador e pelo ouvido. O linguista Lieberman (1968) sugere, por exemplo, que os primatas não humanos carecem dos mecanismos físicos necessários à produção da variada fala humana.

Falar uma língua pressupõe o conhecimento de uma série de sistemas: o sistema sonoro, chamado fonética; o sistema dos significados, chamado semântica, o sistema de regras de formação de palavras, chamado morfologia, e o sistema de regras de formação de frases, chamado sintaxe. Por fim, deve-se dispor de um vocabulário de palavras, isto é, de um léxico adequado.

A articulação, as características físicas, e a percepção dos sons são o objecto de estudo da Fonética, disciplina que estuda e descreve os sons independentemente da sua função dentro de um ou mais sistemas linguísticos (Moutinho, 2000).

2.2.2 Aparelho fonador e produção da fala

Nos seres humanos, os sons são gerados através da modulação do ar na fase de expiração, por meio de vibração, articulação e ressonância. Tais mecanismos envolvem a acção duma série de órgãos que formam o “aparelho fonador”. Moutinho (2000) explica que o aparelho fonador é formado por:

- Pulmões - geradores do fluxo de ar necessário para fazer vibrar as cordas vocais;
- Laringe - possui as cordas vocais, responsáveis pela vibração;
- Cavidades supralaríngeas - articuladores e ressoadores.

A laringe é constituída pelas pregas ou cordas vocais e pela glote (Moutinho, 2000).

Deste conjunto, ir-se-á aprofundar o conhecimento da anatomia e funcionamento das cavidades supralaríngeas, tendo sido apenas estas objecto de análise nesta investigação. É a

este nível do aparelho fonador (cavidades supralaríngeas) que o ar é modulado e se torna nos diferentes sons de uma determinada língua.

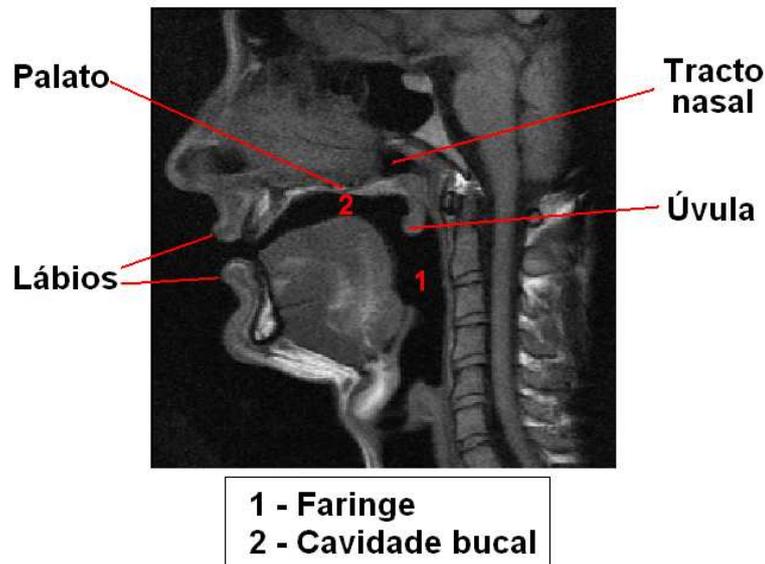


Figura 2.1: Aparelho fonador demonstrado numa imagem de Ressonância Magnética na produção do fonema [Ø].

As cavidades supralaríngeas, apresentadas na Figura 2.1, são formadas por:

- Tracto nasal - possui uma configuração estática e mede entre 10,0 a 12,5 cm (Teixeira et al., 2001);
- Faringe - tracto com cerca de 12 cm de comprimento, (Martins, 2007);
- Cavidade bucal - possui uma configuração dinâmica que leva à modificação das características de ressonância do tracto vocal (formado pela faringe mais a cavidade bucal) logo, à produção da fala, (Moutinho, 2000).

Por sua vez, dentro da cavidade bucal, situam-se os “articuladores”, que também intervêm na produção dos diferentes sons e que, por tal razão, são muitas vezes utilizados nos modelos do aparelho fonador, e posteriormente incorporados nos sintetizadores de fala.

Entre os articuladores, distinguem-se os “activos” e os “passivos”, consoante possuem flexibilidade ou não. São articuladores activos:

- Lábios;
- Língua - é o articulador que pode assumir um maior número de conformações, sendo por isso o mais difícil de modelar;
- Velo - pode encontrar-se elevado ou baixado, permitindo ou não o acoplamento com o restante tracto vocal. A extremidade do velo é denominada úvula (Martins, 2007);

- Mandíbula.

Os articuladores passivos são:

- Sector anterior da arcada dentaria superior (SAADS);
- Alvéolos dentários superiores;
- Palato duro - estrutura curva cujo vértice atinge o ponto mais alto da cavidade oral.

As inúmeras combinações das configurações dos articuladores proporcionam a existência de um vasto leque de sons, analisados e classificados pela comunidade científica. De seguida, ir-se-á abordar tais classificações, focando a atenção nas que mais dizem respeito aos sons da língua portuguesa.

2.2.3 Sons do Português Europeu

Apresentado, na Secção anterior, o funcionamento do aparelho fonador, é agora importante relembrar a classificação dos sons e as características que os diferenciam. É no âmbito da investigação fonética (Secção 2.2.1) que se sentiu a necessidade de criar classificações para os diferentes sons gerados pelo aparelho fonador. De seguida, abordam-se as classificações que dizem respeito à língua portuguesa.

Existem duas classes principais de sons: as vogais e as consoantes. Nas primeiras, os sons são produzidos sem constricções à passagem do ar, sendo que nas segundas existe uma obstrução à passagem do fluxo de ar.

Cada uma destas classes de sons pode ser subdividida, por sua vez, em sons orais ou nasais. A distinção entre os primeiros e os segundos é dada pela posição do velo aquando da produção do som. Quando o velo está levantado, fechando a passagem para o tracto nasal, o ar é expelido na totalidade pela boca, gerando os sons orais. Pelo contrário, quando o velo se encontra numa posição rebaixada, permitindo ao fluxo de ar que sai da faringe o acesso ao tracto nasal, o som produzido é o resultado da emissão do ar, não só através da cavidade bucal, mas também pelas cavidades nasais. Os sons assim produzidos adoptam a classificação de nasais. A Figura 2.2 apresenta um exemplo elucidativo sobre as diferenças descritas.

No que diz respeito às vogais, estas podem ainda ser classificadas segundo três critérios distintos: (1) lugar ou zona de articulação, (2) labialização e (3) grau de abertura (Moutinho, 2000). O primeiro critério divide as vogais em “anteriores”, “centrais” e “posteriores”, de acordo com a zona da cavidade bucal em que estas são articuladas, como se mostra na Figura 2.3. A labialização indica a ausência ou presença da projecção dos lábios, distinguindo as vogais em “não labializadas” e “labializadas”. No PE, só as vogais posteriores são consideradas labializadas (Moutinho, 2000).

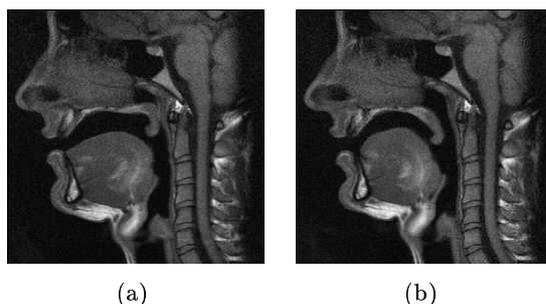


Figura 2.2: Diferenças entre um som oral e um nasal: (a) produção do [6] (som oral); (b) produção do [6~] (som nasal).

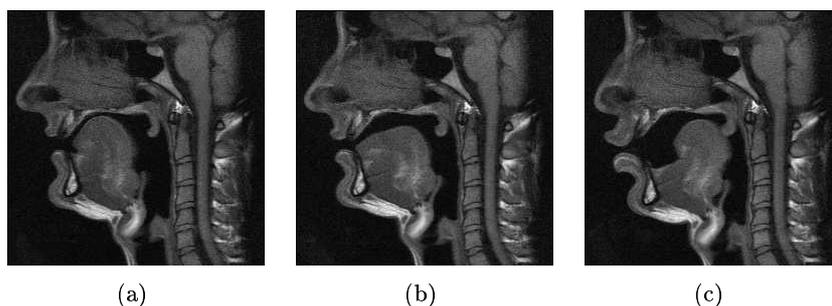


Figura 2.3: Exemplo de classificação das vogais segundo a zona de articulação: (a) Anterior [i], (b) Central [ø], e (c) Posterior [u].

Por fim, o terceiro critério refere-se à abertura do canal bucal e leva à classificação em “abertas”, “semi-abertas”, “semi-fechadas” e “fechadas”, Figura 2.4.

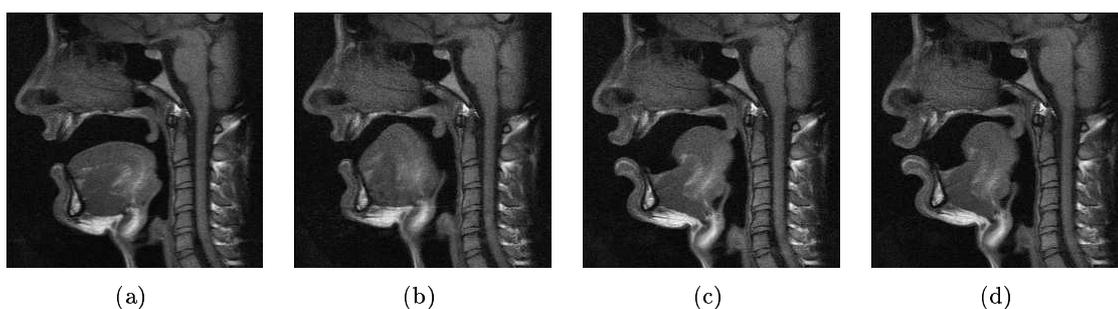


Figura 2.4: Exemplo de classificação das vogais segundo o grau de abertura: (a) Aberta [a], (b) Semi-Aberta [E], (c) Semi-fechada [o], e (d) Fechada [u].

As consoantes são classificadas em função do: (1) modo de articulação e (2) ponto de articulação. O modo da articulação especifica a natureza da obstrução e divide as consoantes em “oclusivas” (onde existe uma obstrução total à passagem de ar), “fricativas” (existe uma constrição do tracto vocal), “laterais” (existe uma obstrução parcial provocada pelo contacto da língua com o palato) e “vibrantes” (onde os sons são articulados com uma constrição parcial que provoca a vibração da língua), Figura 2.5.

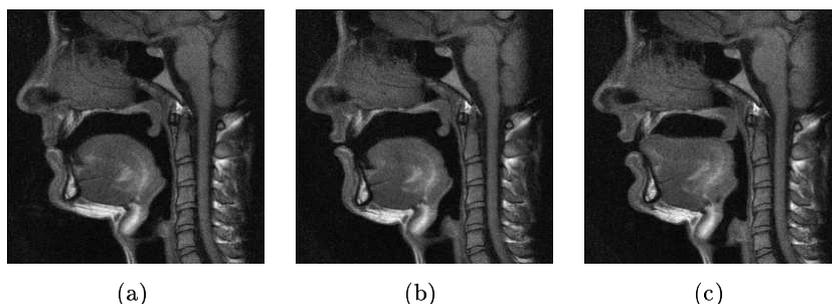


Figura 2.5: Exemplo de classificação das consoantes segundo o modo de articulação: (a) oclusiva [p], (b) fricativa [f], e (c) lateral [l].

Os pontos de articulação designam-se usando o articulador e o ponto para onde este se move. Podem então ser classificadas em “bilabiais”, “labiodentais”, “dentais”, “alveolares”, “pré-palatais”, “palatais”, “velares” e “úvulares”, Figura 2.6.

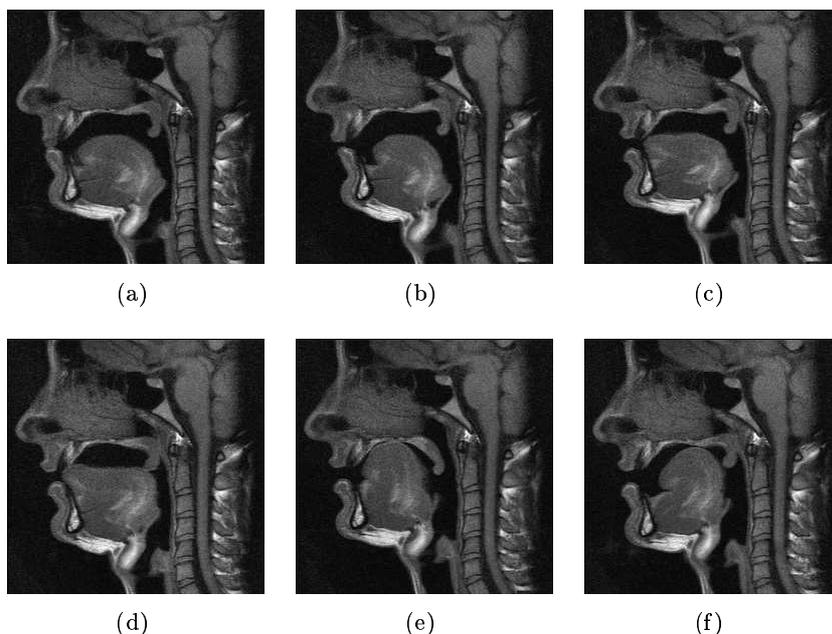


Figura 2.6: Exemplo de classificação das consoantes segundo o ponto de articulação: (a) Bilabial [p], (b) Labiodental [f], (c) Dental [s], (d) Alveolar [l], (e) Palatal [j], e (f) Velar [g].

Termina-se esta Secção mencionando, no âmbito dos mecanismos de produção da fala, um fenómeno de particular importância: a “coarticulação”. Este fenómeno diz respeito aos efeitos exercidos pelos sons adjacentes a um dado fonema, na produção deste último. A título explicativo, menciona-se o exemplo da produção de uma sequência do tipo vogal-consoante-vogal. Neste caso, as posições articulatórias da consoante serão afectadas pela vogal que a segue e, em alguns casos, até pela vogal que a antecede. Isto é, o fenómeno da coarticulação é traduzido pelo “contexto” em que um dado fonema é produzido, conferindo-lhe determinadas características sonoras. Isto é, o contexto em que o fonema é produzido, confere-lhe características

sonoras específicas.

2.2.4 Alfabetos fonéticos

A grande variedade de sons inviabiliza a sua transcrição através dos alfabetos comuns, que apresentam um sistema estabelecido de representar apenas letras. Em muitos casos, a “grafia” de uma dada letra pode não corresponder às múltiplas formas de pronunciar esta mesma letra, isto é, aos diferentes sons. Por essa razão, existem alfabetos mais abrangentes que não se limitam à representação das letras, mas também dos próprios sons, ou fones de uma língua. No caso do Português, por exemplo, no alfabeto fonético constam nove vogais. Destas, o [i], [u] e [a] são designadas vogais cardinais, Figura 2.7.

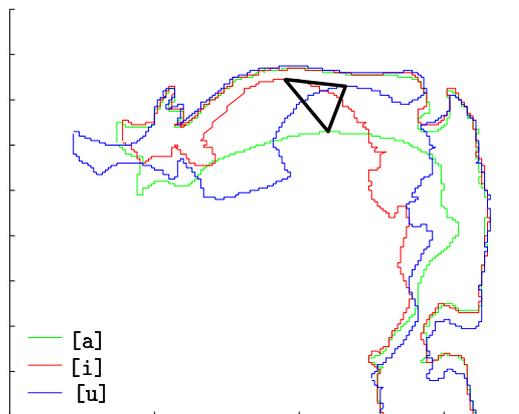


Figura 2.7: Representação dos contornos das três vogais cardinais, [i] (a vermelho), [u] (a azul) e [a] (a verde) e união dos seus pontos mais altos (triângulo a preto).

O alfabeto fonético mais conhecido é o IPA (*International Phonetic Alphabet*). Contudo, este alfabeto não é prático de digitar nos teclados comuns. Tal limitação levou à utilização, na realização do presente trabalho, do alfabeto fonético SAMPA (*Speech Assessment Methods Phonetic Alphabet*¹) (Wells, 1997), outro alfabeto fonético que consiste basicamente na adaptação do alfabeto IPA para código ASCII, característica que o torna de mais fácil escrita com teclados tradicionais.

2.3 Métodos de recolha de informação sobre a produção de fala

O estudo dos sons da fala e dos mecanismos anatómicos que os produzem, tratados na Secção anterior, baseia-se na observação dos mesmos. Os instrumentos de apoio e os métodos utilizados para proceder a tal observação têm sido vários: *Velotrace*, *Electro Magnetic Mid-sagittal Articulometer* (EMMA), *X-ray Microbeam System*, Electropalatografia (EPG), fotografia de alta velocidade, etc. O estudo destes mecanismos foi também suportado e baseado na análise

¹O alfabeto SAMPA para diversas línguas, incluindo o PE, pode ser encontrado em <http://www.phon.ucl.ac.uk/home/sampa/index.html>.

de imagens resultantes de radiografia simples, cineradiografia, videofluoroscopia.

Mais recentemente, os investigadores utilizaram também imagens obtidas através de Tomografia Computorizada (TC), Ressonância Magnética (RM) e Ultrasonografia, (Mohammad, 1999). Destas, só as últimas duas não utilizam radiações ionizantes, apresentando menos riscos para o informante, razão pela qual se tornaram objecto de crescente interesse nesta área.

A literatura mostra uma grande prevalência da utilização das imagens de RM para aquisição de informação relativa ao tracto vocal para diversas línguas: Inglês Americano (Stone e Lundberg, 1996; Story et al., 1996; Alwan et al., 1997; Narayanan et al., 1997, 2004); Inglês Britânico (Greenwood et al., 1992; Mohammad et al., 1997; Mohammad, 1999; Shadle et al., 1999); Francês (Demolin et al., 1996; Badin et al., 1998a, 2000; Soquet et al., 2002; Demolin et al., 2003; Serrurier e Badin, 2005b); Sueco (Engwall e Badin, 1999; Engwall, 2000; Engwall e Badin, 2000; Engwall, 2001, 2003a,b); Japonês (Dang et al., 1998; Dang e Honda, 1997; Takemoto et al., 2003, 2004); Alemão (Hoole et al., 2000; Kröger et al., 2000; Mathiak et al., 2000; Mády et al., 2001, 2002); Coreano (Yang, 1999; Kim, 2004); Tamil (Narayanan et al., 1996, 2004); entre outros. Para o PE, usando RM, encontrou-se apenas um estudo do tracto vocal, (Rua e Freitas, 2006), existindo também algum trabalho realizado para as vogais do Português Brasileiro (Gregio, 2006).

Apresentam-se, na Tabela 2.1, as principais vantagens e desvantagens dos métodos de aquisição de dados sobre o tracto vocal mais comuns.

Tabela 2.1: Comparação entre várias técnicas de aquisição de informação sobre o tracto vocal. Adaptado de (Maeda et al., 2006).

	MRI	EMMA	Ultrason	cine-radiografia	X-ray microbeam
todo o tracto vocal	sim	não	não	sim	não
língua	sim	alguns pontos	sim	sim	alguns pontos
base da língua	sim	não	não	sim	não
velo	sim	sim	não	sim	sim
resolução temporal	-	200 Hz	30-200 Hz	50-60 Hz	40-160 Hz
3D	sim	não	não ²	não	não
efeitos nocivos	não	não	não	sim	não/sim ³
artefactos naturais	sim (posição de supinação)	afecta	sim	sim	afecta
ruído acústico	alto	baixo	aceitável	baixo	aceitável
movimentos da cabeça	restritos	restritos	restritos	livres	livres
portátil	não	não	sim	não	não
custo	alto	alto	baixo	alto	alto

²Stone e Lundberg (1996) referem o desenvolvimento experimental de uma maquina de Ultrason 3D.

³O risco é mais baixo do que no raio X convencional, mas continua a ser utilizada radiação ionizante.

A utilização de imagens obtidas por RM foi a opção feita também para a realização do presente estudo. Por tal razão, e a título de contextualização, passa-se a explicar os mecanismos físicos subjacentes à aquisição de imagens através de RM.

2.3.1 Ressonância Magnética

O hidrogénio é o átomo de estrutura mais simples, entrando por isso na composição de inúmeros compostos, entre os quais a água. A elevada concentração de água no corpo humano, tem como directa consequência a prevalência de hidrogénio face a outros elementos. Isto, e o facto de apresentar um momento magnético elevado, justifica a utilização do átomo de hidrogénio na geração de imagens por RM. Nesta técnica, uma imagem é gerada processando o sinal eléctrico emitido pelos núcleos dos átomos hidrogénio através da indução da sua excitação. Esta Secção apresenta uma descrição desta técnica de aquisição de imagens.

O hidrogénio é constituído apenas por um protão e um electrão, o segundo movendo-se em torno do primeiro. O protão tem igualmente um movimento giratório, mas em torno do seu próprio eixo, gerando um campo magnético, denominado *spin* magnético.

Quando os protões não se encontram sob a influência de campos magnéticos exteriores, o *spin* magnético de cada um deles aponta para uma direcção diferente. Nesta situação, a soma vectorial de todos eles é igual a zero, isto é, a magnetização total M_{tot} é igual a zero ($M_{tot} = 0$).

Para se obter o sinal de RM, isto é, o sinal eléctrico emitido pelos núcleos de hidrogénio, é necessário colocar a amostra a examinar sob o efeito de um campo magnético, que varia geralmente entre 0,2 e 3,0 Tesla (1 Tesla = 10^4 Gauss) dependendo do aparelho. Este campo magnético estático é chamado de campo magnético externo B_0 .

Sob a acção do campo magnético, os *spins* começam a “sentir” o efeito do campo magnético externo B_0 podendo orientar-se paralelamente (Energia E_1) ou anti-paralelamente (Energia E_2) a B_0 . A alteração induzida na direcção do *spin* magnético através da influência do campo magnético externo faz com que a soma vectorial de todos os *spins* seja diferente de zero. Isto é, existe uma magnetização total (M_{tot}) paralela a B_0 (magnetização longitudinal).

De seguida, é necessário medir esta magnetização. Para tal, é necessário deslocar a magnetização para um eixo perpendicular a B_0 , ou seja, o eixo transversal. Para passar a magnetização do eixo longitudinal para o transversal é necessário emitir uma onda electromagnética, isto é, onda de radiofrequência (RF) normalmente em torno de 63 MHz (para um campo de 1,5 Tesla), com campo magnético B_1 perpendicular a B_0 . A aplicação do pulso RF, resulta num movimento de precessão em torno de B_1 , no eixo transversal, sem deixar de girar em torno de B_0 , Figura 2.8(a).

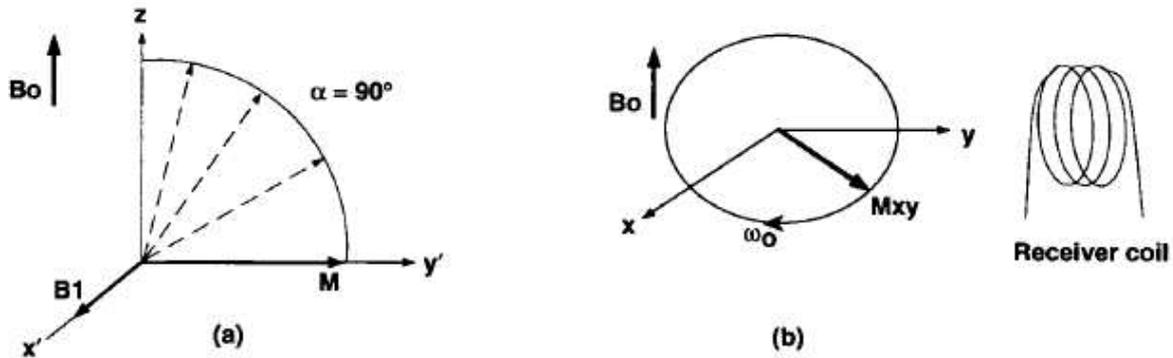


Figura 2.8: Funcionamento da ressonância magnética: (a) precessão da magnetização M em torno da RF. O *frame* de referencia $x'y'z$ roda em torno de B_0 , (b) a rotação de M_{xy} é proporcional à magnitude de B_0 . Esta oscilação da magnetização é medida pela bobine receptora. Adaptado de (Wright, 1994).

Quando cessa o pulso de RF, os núcleos excitados tendem a retornar ao nível de energia normal, mais baixo, com *spins* maioritariamente paralelos ao campo externo, emitindo a mesma frequência absorvida (fenómeno de ressonância). Finalmente, este sinal, designado por decaimento de livre indução (FID), é medido por uma bobina ou antena, Figura 2.8(b). A rapidez com que o conjunto de *spins* retorna ao estado de equilíbrio depende do tipo de tecido a que eles pertencem.

Existem dois processos de relaxamento: relaxamento longitudinal ($M_z \rightarrow M_{tot}$) e relaxamento transversal ($M_{xy} \rightarrow 0$), que são descritos pelas constantes de tempo T_1 e T_2 (medidas em milissegundos). O parâmetro T_1 representa o tempo que os prótons demoram a recuperar 63% da magnetização longitudinal. O T_2 é o tempo de decaimento da magnetização transversal para 37%, resultante das interações *spin-spin*. O relaxamento transversal decai ainda através de um mecanismo adicional denominado T_2^* , resultante da falta de homogeneidade do campo magnético. T_1 , T_2 e T_2^* dependem do tipo de tecido. T_2^* conta ainda com os efeitos de heterogeneidade do campo magnético, susceptibilidade magnética e desvio químico.

Em suma, os passos mais importantes no processo de aquisição de uma imagem de RM são:

1. Colocação do paciente num campo magnético intenso;
2. Envio de um pulso de oscilações de campo magnético na gama da RF;
3. Cessamento do pulso de RF;
4. Produção de um sinal de RF pelo paciente;
5. Localização, amplificação e desmodulação do sinal obtido;
6. Reconstrução de uma imagem do paciente.

As principais vantagens e desvantagens, no contexto do estudo da fala, deste método de aquisição de dados são apresentadas na Tabela 2.2.

Tabela 2.2: Vantagens e desvantagens da utilização da RM

Vantagens	Desvantagens
<ul style="list-style-type: none"> - Método relativamente inócuo - Boa resolução de contraste ao nível dos tecidos moles - Apresenta capacidade multiplanar - Permite modelação 3D - Visão do tracto vocal em toda a sua extensão 	<ul style="list-style-type: none"> - Tempo de aquisição longo - Baixa resolução temporal - Não há distinção entre osso, dentes e ar - O processo de aquisição produz muito ruído acústico (não confundir com ruído na imagem)

Durante a fase de aquisição de imagens de RM torna-se necessário tomar uma série de opções tendo em consideração a finalidade das imagens a adquirir. Têm, por exemplo, que se estabelecer compromissos entre a resolução espacial, a relação Sinal/Ruído e a resolução temporal. Estas e outro tipo de escolhas são feitas através da manipulação de uma série de parâmetros, descritos na próxima Secção.

Parâmetros Manipuláveis na Aquisição de Imagens por Ressonância Magnética

Durante a aquisição de imagens de RM tem que se proceder a uma série de escolhas tendo em consideração a finalidade das imagens a adquirir. Veja-se, por exemplo, o aumento de resolução espacial, em detrimento do tempo de aquisição. Esta e outro tipo de escolhas são feitas através da manipulação de uma série de parâmetros programáveis. De seguida, faz-se uma breve descrição dos parâmetros considerados mais importantes:

- TR = Tempo de Repetição: Intervalo de tempo entre dois pulsos RF, Figura 2.9

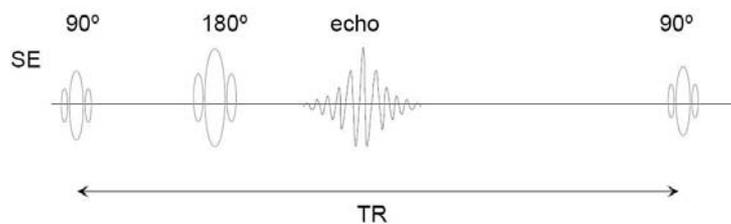


Figura 2.9: Representação gráfica do conceito de TR. Adaptado de (Blink, 2004).

- TE = Tempo de Eco: Intervalo de tempo entre a aplicação do pulso RF e o pico do eco, Figura 2.10
- FA = Ângulo de Nutação (a sigla vem do inglês *Flip Angle*): Determina o ângulo de nutação (ângulo de Ernst) da magnetização em relação ao plano XY, Figura 2.11

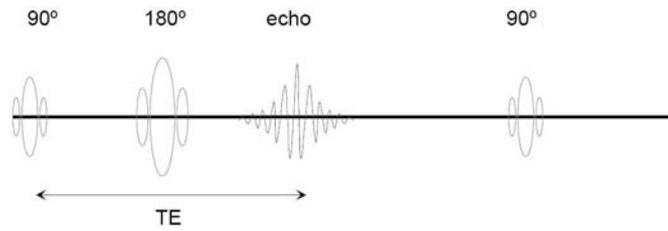


Figura 2.10: Representação gráfica do conceito de TE. Adaptado de (Blink, 2004).

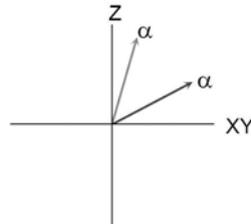


Figura 2.11: Representação gráfica do conceito de FA. Adaptado de (Blink, 2004).

- IT = Tempo de Inversão (a sigla vem do inglês *Inversion Time*): Intervalo de tempo entre o pulso de 180° e o pulso de 90°, Figura 2.12

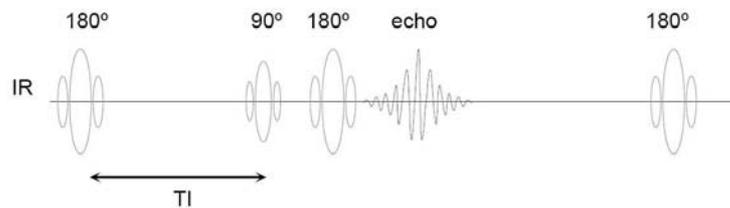


Figura 2.12: Representação gráfica do conceito de IT. Adaptado de (Blink, 2004).

- NA = Número de Aquisições: Número de aquisições. É necessário ter em atenção que seleccionar um NA=2 dobra o tempo necessário para a aquisição, mas aumenta a relação sinal/ruído somente em $\sqrt{2}$ (Wright, 1997).
- MX = Matriz: Determina, entre outros, a resolução espacial da imagem, Figura 2.13
- FOV = Campo de Visão (a sigla vem do inglês *Field Of View*): Região anatómica visualizada, Figura 2.14
- ST = Espessura de corte (a sigla vem do inglês *Slice Thickness*): Veja a Figura 2.15

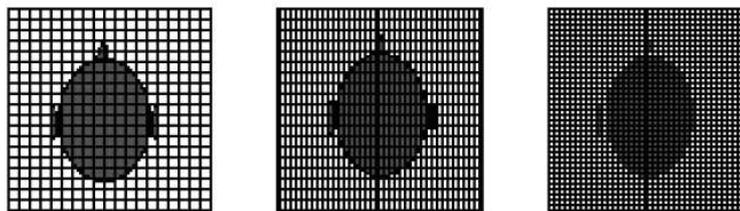


Figura 2.13: Representação gráfica do conceito de MX. Adaptado de (Blink, 2004).

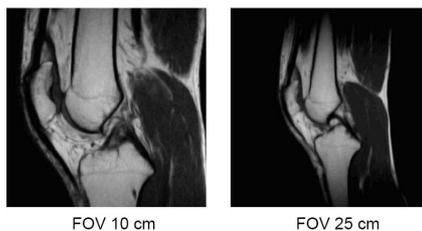


Figura 2.14: Representação gráfica do conceito de FOV. Adaptado de (Blink, 2004).

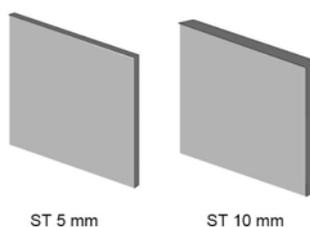


Figura 2.15: Representação gráfica do conceito de ST. Adaptado de (Blink, 2004).

- SG = Intervalo entre cortes (a sigla vem do inglês *Slice Gap*): Espaçamento entre fatias, Figura 2.16

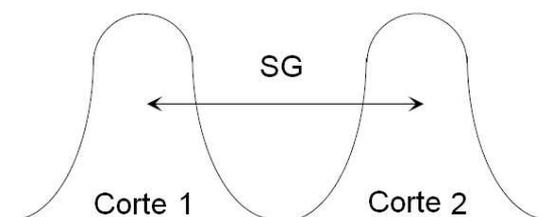


Figura 2.16: Representação gráfica do conceito de SG. Adaptado de (Blink, 2004).

- Codificação de fase (*Phase Encoding*): alteração da fase das rotações (*spins*) numa dimensão com um gradiente de pulso do campo magnético ao longo dessa dimensão antes da aquisição do sinal.

- BW = largura de banda (a sigla vem do inglês *Band Width*): Largura de banda do receptor = frequência de amostragem = 1 / tempo de aquisição Normalmente é medida em Hertz (Hz).

Juntando alguns dos conceitos definidos acima, e observando que a largura de banda do receptor é em Hz e o gradiente em Hz/cm, obtém se a relação entre o FOV, a BW e o G_{R0} (*Read Out gradient* ou codificação em frequência):

$$FOV = \frac{Sample\ BW}{G_{R0}\ Strength}$$

A largura de banda do receptor define o intervalo de frequências de uma ponta do FOV à outra (Blink, 2004).

Uma adequada manipulação dos parâmetros descritos permite o realce das características de interesse dos tecidos humanos (contraste). Algumas das relações mais importantes são:

- TR longo reduz o efeito de T_1 ;
- TR curto aumenta o contraste devido a T_1 ;
- TE curto reduz o efeito de T_2^* .

Note que T_1 e T_2 são propriedades inerentes dos tecidos, enquanto que TR e TE são parâmetros controláveis. A discussão anterior leva-nos à noção de ponderação:

- Se TR e TE são curtos, a ponderação é em T_1 ;
- Se TR e TE são longos, a ponderação é em T_2 ;
- Se TR é (muito) longo e TE é (muito) curto, ponderação é em DP (densidade de prótons).

Considera-se agora oportuno introduzir a expressão geral do sinal de ressonância FID (Hashemi et al., 1997):

$$FID \propto DP(1 - e^{-T_1/TR})e^{-T_2^*/TE} \cos(\omega_0 T) \quad (2.1)$$

Codificação do Sinal

A localização do sinal de RM é efectuada através dos gradientes (Kornprobst et al., 2002; Greenspan et al., 2002). Estes são variações lineares da intensidade, no espaço de reconstrução de imagem, do campo magnético ao longo de uma determinada direcção. Podem ser de três tipos, dependendo do eixo: G_x , G_y e G_z . Num paciente em supinação (decúbito dorsal), a direcção x é normalmente a direcção esquerda-direita do paciente e a direcção y é escolhida como sendo a direcção anterior-posterior; z corresponde à direcção longitudinal, paralela ao campo magnético principal. Normalmente, z é horizontal e corresponde à direcção cabeça-pés (ou pés-cabeça), quando se está a adquirir informação relativa ao corpo (podendo

ser diferente quando se pretende estudar, por exemplo, o cérebro).

O gradiente de selecção do corte (que excita uma determinada região espacial) pode ser aplicado ao longo de qualquer eixo, em função da orientação desejada. Veja-se a Figura 2.17 onde se aplica um gradiente de selecção do corte ao longo do eixo longitudinal.

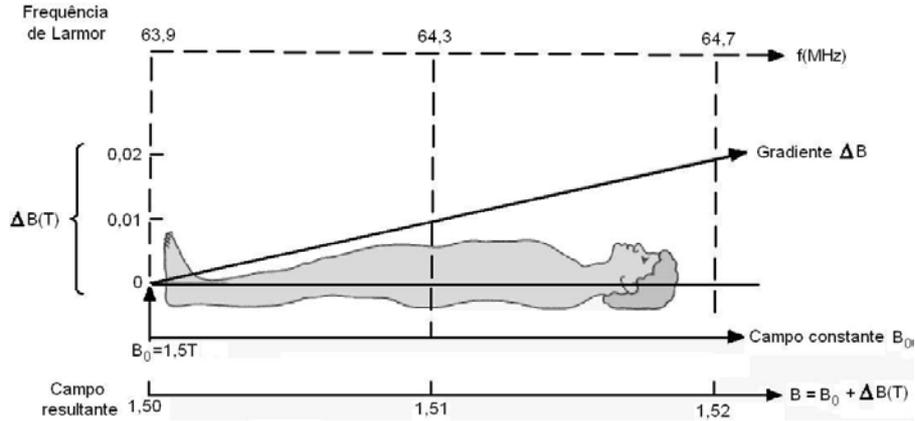


Figura 2.17: Esquematização da acção dos campos magnéticos associados à aquisição de imagens por RM. Adaptado de (Silva, 2003).

A variação da espessura do corte pode ser dada pela:

- variação da largura de banda - quanto menor a largura de banda, menor a espessura;
- variação da intensidade do gradiente G_z - quanto maior o gradiente, menor a espessura.

Após se ter excitado um corte, é necessário posicionar a informação dentro do corte (ou seja, dentro de cada corte, é necessário saber de onde vem o sinal). São para isso utilizados mais dois gradientes, gradiente de codificação de frequência (codificação no eixo dos x) e gradiente de codificação de fase (codificação no eixo dos y). A primeira cria uma dependência linear entre a localização espacial e a frequência de oscilação. A segunda é idêntica à primeira e resulta de uma pré-codificação de frequência durante um certo intervalo de tempo.

O sinal de RM é codificado no chamado espaço K, também conhecido, na área da engenharia, como espaço de Fourier, Figura 2.18.

A primeira transformação é relacionada com a forma como os dados são recolhidos. A segunda determina a forma como os dados adquiridos são processados de forma a formar a imagem. A sequência RM é então:

$$\vec{\mu} \rightarrow \vec{M} \rightarrow \vec{M}_{xy} \rightarrow S(t) \rightarrow S(\vec{k}) \rightarrow I(\vec{x})$$

onde $\vec{\mu}$ são os momentos magnéticos microscópicos, \vec{M} é a *bulk* de magnetização, \vec{M}_{xy} é a magnetização transversa, $S(t)$ é o sinal eléctrico, $S(\vec{k})$ é o sinal no espaço K, e $I(\vec{x})$

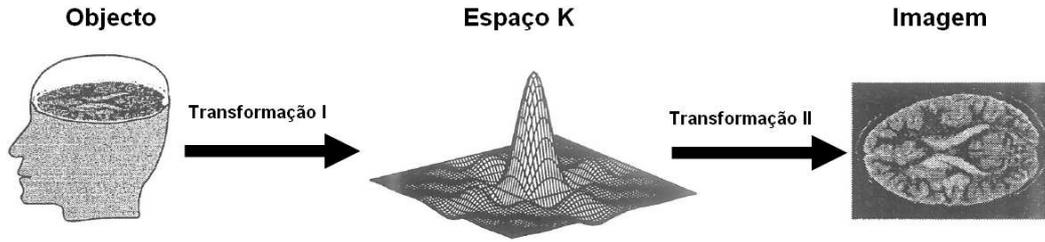


Figura 2.18: Codificação de imagens de RM. Adaptado de (Liang e Lauterbur, 1999).

é a imagem RM. A transformação $\vec{\mu} \rightarrow \vec{M}$ é dada pela exposição do objecto ao campo B_0 , $\vec{M} \rightarrow \vec{M}_{xy}$ é feita através de excitações RF, $\vec{M}_{xy} \rightarrow S(t)$ é baseado na lei de indução de Faraday, $S(t) \rightarrow S(\vec{k})$ envolve o uso de gradientes magnéticos de forma a codificar a informação espacial nas respostas de um sistema de *spins* através de excitações RF, finalmente, $S(\vec{k}) \rightarrow I(\vec{x})$ é dada pela Transformada de Fourier Inversa.

A informação relativa às baixas frequências é concentrada no centro do espaço K, enquanto que a informação relativa a frequências altas é agrupada na sua zona periférica. Sendo assim, o contraste e as formas gerais da imagem de RM estão relacionados com o centro do espaço K, tendo as frequências baixas amplitudes altas, o que origina grandes alterações nos níveis de cinzento da imagem final. Contrariamente, as frequências altas são caracterizadas por amplitudes baixas, não tendo nenhum efeito no contraste ou forma geral da imagem de RM. Esta informação é no entanto importante para o aumento da resolução espacial da imagem final. A Figura 2.19 ilustra estes fenómenos.

Sequências de pulsos

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, afectando as características das imagens. Existe, hoje em dia, um vasto número de sequências, tornando-se por vezes difícil a sua classificação e normalização, uma vez que cada fabricante utiliza nomenclaturas diferentes para sequências que são basicamente idênticas (Marques, 2003).

Contudo, pode-se dividir as sequências de pulso em três grandes grupos (Marques, 2003):

1. Sequências Eco de Spin (ES) ou Spin Eco (SE);
2. Sequências Eco de Gradiente (GE);
3. Sequências híbridas.

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

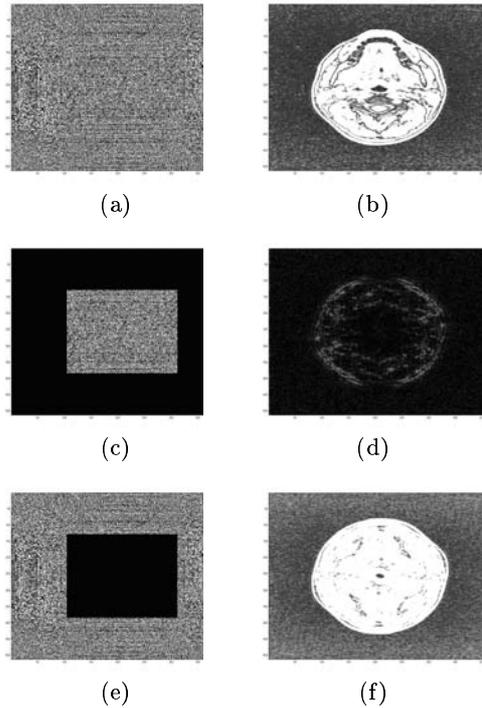


Figura 2.19: Organização do espaço K: (a) espaço K original; (b) imagem de RM original; (c) espaço K alterado onde se mantiveram as baixas frequências; (d) reconstrução da imagem de RM a partir de (c); (e) espaço K alterado onde se mantiveram as altas frequências; (f) reconstrução da imagem de RM a partir de (e).

ou Fast GE e sequências Turbo ou Fast IR.

De seguida, são apresentadas algumas das sequências existentes para a aquisição de imagens 2D, sequências de imagens de Tempo Real, e aquisições volumétricas. São focadas apenas as sequências utilizadas na aquisição dos dados tratados neste trabalho.

2.4 Sequências escolhidas para a aquisição dos dados

Dentro do variado leque de sequências de aquisição disponíveis, escolheu-se uma sequência por cada tipo de dados a adquirir (Martins, 2007). Os parágrafos seguintes descrevem as sequências escolhidas para este trabalho.

Imagens 2D: Sequência TSE

Esta sequência (que pertence ao grupo SE) é uma versão otimizada da técnica RARE (*Rapid Acquisition Relaxation Enhancement*) (Hashemi et al., 1997). 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, como mostra a Figura 2.20.

Trata-se de uma técnica multi-eco em que cada SE produzido é codificado com fase dife-

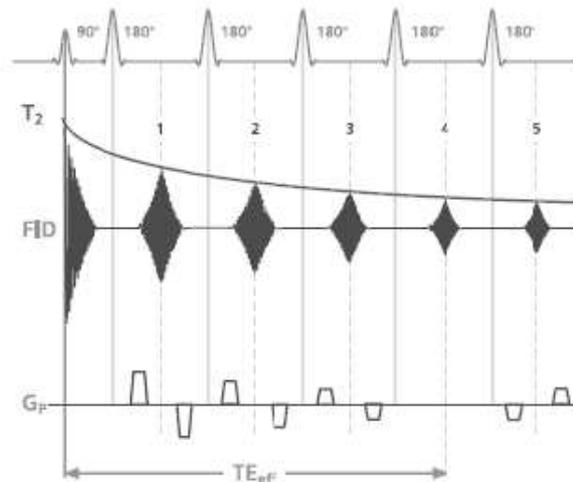


Figura 2.20: Diagrama de pulso de uma sequência TSE. Adaptado de (Siemens, 2003).

rente. Como resultado da mudança de fase entre ecos, através da aplicação do gradiente, podem ser preenchidas várias linhas do espaço K (ou espaço de Fourier) por Tempo de Repetição (TR), encurtando o Tempo de Aquisição (TA). O número de ecos por TR designa-se por *Echo Train Length* (ETL) ou *turbo Factor* (Brown e Semelka, 2003; Carroll et al., 2006; Elster e Burdette, 2001; Marques, 2003). É necessário ter em atenção que, quanto maior o factor turbo, menor o número de cortes por TR permitidos. Encurtando o espaçamento entre ecos, a duração do trem de ecos é menor, diminuindo o esborratamento na imagem, característico desta sequência (Carroll et al., 2006). O contraste na imagem é determinado pelo Tempo de Eco (TE) efectivo, sendo possível obter ponderação T_1 , DP e T_2 .

As grandes vantagens das sequências TSE, além da diminuição do TA, são a diminuição dos efeitos de susceptibilidade magnética e o facto de tornar possível aquisição 3D TSE com ponderação em T_2 , com um TA aceitável. Uma outra vantagem é que, dado 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 obter matrizes de 512×512) e diminuição da espessura de corte. Como se pode aumentar as NEX (*Number of Excitations*), a relação Sinal/Ruído é boa mesmo com cortes finos. Por outro lado, uma das grandes desvantagens, e que pode ser condicionante, é o aumento da deposição de RF, com aumento do SAR (*Specific Absorption Rate*) devido ao maior número de pulsos de 180° utilizados. Além deste, os efeitos de transferência de magnetização e o esborratamento observado nas imagens, em particular para um TE curto, constituem outras desvantagens (Hashemi et al., 1997).

Conjuntos de imagens de Tempo Real: Sequência TrueFISP

Na sequência True FISP (*Fast Imaging with Steady Precession*) ambos componentes do sinal

(FID e SE/STE²) são utilizados, sendo obtido um eco combinado simples (Elster e Burdette, 2001; Scheffler e Lehnhardt, 2003; Marques, 2003). O diagrama de pulsos deste tipo de seqüências é apresentado na Figura 2.21.

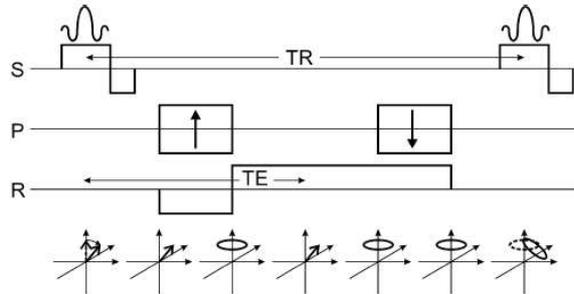


Figura 2.21: Diagrama de pulsos de uma seqüência TrueFISP. Adaptado de (Scheffler e Lehnhardt, 2003).

Esta seqüência foi descrita em 1958 por Carr e proposta por Oppelt em 1986, 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) de forma a minimizar as bandas de interferência (*dark-phase dispersion bands*) características nesta seqüência, resultantes de falta de homogeneidade de campo induzida pelo paciente (Scheffler e Lehnhardt, 2003).

Aquisições volumétricas: Seqüência 3D VIBE

A seqüência 3D VIBE é uma seqüência FLASH (*Fast Low Angle Shot*) em volume. A seqüência FLASH é uma seqüência EG ultra-rápida (Weishaupt et al., 2003). O 3D VIBE é uma seqüência EG *spoiled*, o que significa que a magnetização transversal residual é destruída antes da aplicação do pulso RF seguinte. O diagrama de pulsos deste tipo de seqüências é apresentado na Figura 2.22.

Nestas seqüências, após a aplicação de um pulso de excitação RF, é aplicado um gradiente que permite a refocalização dos *spins*, obtendo-se um eco (neste caso eco de gradiente).

Esta seqüência, por ser do tipo EG, é mais sensível à falta de homogeneidade do campo e a fenómenos de susceptibilidade magnética. Como vantagem, as seqüências 3D EG *spoiled* possibilitam a recolha de um conjunto de cortes muito finos, o que permite a reconstrução dos dados em qualquer direcção.

²STE significa Eco estimulado

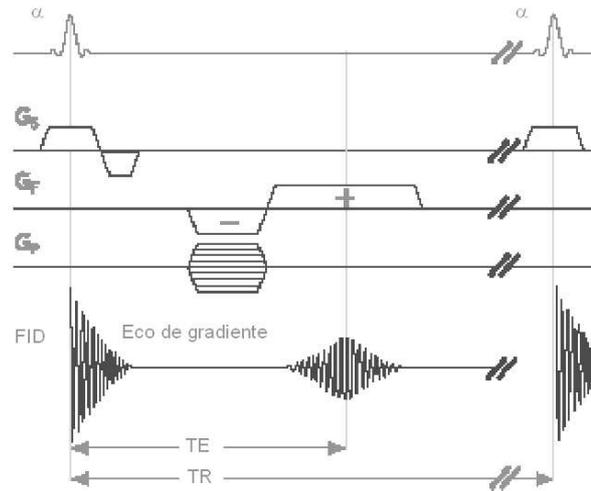


Figura 2.22: Diagrama de pulsos de uma sequência EG Spoiled 2D - FLASH. Adaptado de (Siemens, 2003).

2.5 Caracterização dos dados (imagens e volumes) adquiridos

As imagens 2D possuem uma relação sinal/ruído bastante boa e não apresentam artefactos significativos. Apesar disso, a segmentação do tracto vocal apresenta algumas dificuldades. Uma destas dificuldades é a distinção entre os dentes e o ar, dado que a falta de átomos de hidrogénio nos dentes faz com que estes não sejam visíveis nas imagens de RM. A delimitação do tracto vocal, nas aquisições onde a produção dos fonemas é feita com a boca aberta, é uma outra dificuldade, pois o contorno tem tendência a parar no limite da imagem, quando deveria terminar no limite dos lábios. Ainda na delimitação do tracto vocal, desta vez inferiormente, encontraram-se dificuldades em distinguir a glote. A glote é importante por ser um dos pontos de referência normalmente utilizados nos estudos de produção de fala. Os limites do tracto vocal, perto da zona do cérebro, não são suficientemente marcados, o que pode fazer com que os contornos “derramem” para esta zona.

As imagens adquiridas em Tempo Real são muito ruidosas, sendo que o contorno do tracto vocal pode até ser descontínuo. Uma zona particularmente ruidosa é a zona do velo, devido à presença de artefactos de susceptibilidade magnética (artefactos causados pela existência de duas zonas com frequências muito distintas) (Martins, 2007). Para além destas desvantagens, as imagens de Tempo Real “herdam” todos os problemas já descritos para as imagens 2D.

Os volumes, neste caso, têm a desvantagem de possuir um voxel anisotrópico. Ou seja, o tamanho do voxel não é igual em todas as direcções, sendo a resolução espacial diferente nos vários planos, tal como já observado em (Martins, 2007). Este facto provoca dificuldades na segmentação das imagens ao nível da cavidade oral, onde a resolução no plano coronal é inferior. Um outro problema é a presença de artefactos de *aliasing* 3D, o que não facilitou a segmentação na zona da glote (Martins, 2007), Figura 2.23.

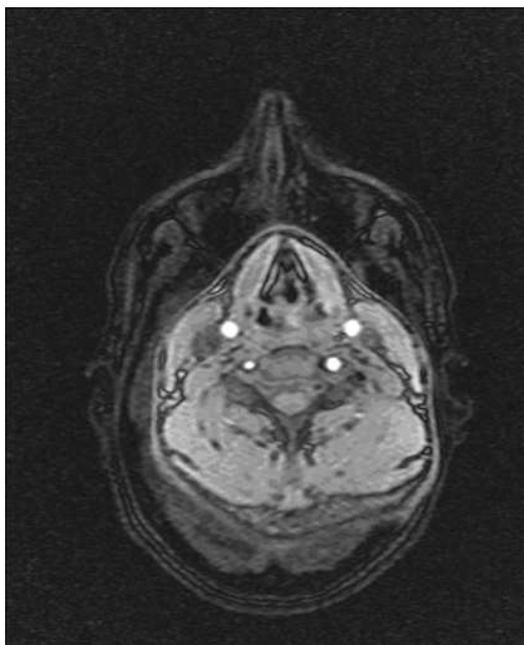


Figura 2.23: Artefacto de “enrolamento” (*wrap around* ou *aliasing*) numa aquisição 3D - observa-se sobreposição dos últimos cortes do volume obtido, sobre os primeiros. Adaptado de (Martins, 2007).

Apesar destas limitações relacionadas com a qualidade dos dados, a maior dificuldade é a forma anatómica do tracto vocal. Este é uma estrutura altamente flexível que (1) varia consideravelmente ao longo do seu comprimento, e (2) mostra um elevado grau de assimetria (Soquet et al., 2002). Para além disso, apresenta uma curvatura de praticamente 90° , o que dificulta o processo de segmentação e posterior análise dos contornos. O ordenamento dos pontos do contorno apresenta uma dificuldade acrescida devido a este factor.

2.6 Considerações finais

O objectivo deste Capítulo foi a contextualização do presente estudo nas áreas com ele relacionadas, nomeadamente a fonética e sons do PE, seus mecanismos de produção e suas técnicas de estudo. Foram abordados e analisados temas como o intrincado mecanismo do sistema fonador e as técnicas de aquisição de dados sobre o funcionamento do tracto vocal. Entre estas foi avaliada como particularmente vantajosa, para a área dos estudos de produção, a aquisição de imagens por RM. Concluiu-se com uma descrição pormenorizada do seu funcionamento.

Todas estas noções representarão a plataforma de conhecimentos necessária ao enquadramento e compreensão da presente investigação, cujos métodos e resultados serão expostos nos próximos Capítulos.

Capítulo 3

Segmentação de imagem

No Capítulo anterior, a investigação foi contextualizada no que diz respeito à sua ligação com as áreas da fonética e imagem médica. Isto é, proporcionaram-se algumas noções sobre as duas áreas principais que se intersectam na presente investigação: abordaram-se conceitos básicos de fonética e sons do PE, desde a apresentação da anatomia e dos mecanismos do aparelho fonador, até aos alfabetos fonéticos; e introduziram-se algumas noções sobre o funcionamento da RM, enquanto método de aquisição de imagens do tracto vocal. O presente Capítulo servirá para apresentar o método escolhido para a análise e processamento destas imagens: a segmentação.

Numa primeira parte, após uma descrição geral da técnica de segmentação de imagens, serão ilustradas de forma pormenorizada as várias fases de implementação desta técnica: uma fase preliminar (pré-processamento) e o processo de segmentação propriamente dito, do qual se apresentarão algumas das diferentes técnicas existentes. Será dada particular atenção às técnicas posteriormente utilizadas para a segmentação do tracto vocal nas imagens e volumes adquiridos. Por fim, será abordada a utilização desta técnica de processamento de imagem na área da investigação fonética.

3.1 Enquadramento geral

Na área de processamento de imagem, entende-se por “segmentação” a partição de uma imagem em regiões que não se intersectem e que sejam homogéneas em relação a uma determinada característica como, por exemplo, a intensidade ou a textura, (Pham et al., 2000). Esta técnica é amplamente utilizada para a análise de imagens médicas e tem sido bastante estudada e debatida, (McInerney e Terzopoulos, 1996), embora ainda não se tenha encontrado uma forma ideal para a sua execução.

Se, por um lado, alguns autores defendem que a segmentação manual é a mais correcta, por ser pouco sensível ao ruído, tolerante à falta de informação e suficientemente precisa (por utilizar a experiência e conhecimento contextual do problema do perito) (Székely et al., 1998; Nascimento, 1998; Olabarriaga, 1999). Por outro lado, outros sustêm que este tipo de técnica,

mesmo feita por especialistas, é muito subjectiva, levando a valores elevados de variância, tanto inter como intra-sujeito (Santos et al., 2004; Mendonça et al., 2007; Poon et al., 2007). Isto é, na execução de uma segmentação manual, dois especialistas podem delinear contornos diferentes a partir de uma mesma imagem; ou até o mesmo especialista pode, em momentos distintos, produzir contornos significativamente diferentes a partir de uma mesma imagem. Para além desta desvantagem, a segmentação manual não é reprodutível e demora muito tempo (Székely et al., 1998; Poon et al., 2007). Em muitos casos, contudo, continua a ser preferível a utilização de métodos manuais de segmentação, devido à difícil utilização dos métodos semi-automáticos desenvolvidos até ao momento e aos resultados, pouco satisfatórios, obtidos com os mesmos.

A inexistência de um método de segmentação “standard” na área da imagem médica prende-se com as metodologias de aquisição de dados, que visam sempre a optimização da qualidade num universo tipicamente 4D (x,y,z,t). Há claramente necessidade de, dadas as actuais limitações tecnológicas, estabelecer compromissos entre as dimensões espacial e temporal, isto é, na aquisição de uma imagem ter que se optar entre ou uma melhor resolução espacial ou temporal. Com efeito, até ao momento, as tentativas de obter dados com amostragem temporal satisfatória têm implicado sacrifícios ao nível da qualidade espacial da imagem. Poderá, então, haver necessidade de se proceder à elaboração de processos de melhoria de imagem numa fase que se denomina “pré-processamento”, que permitam a eliminação ou, na maioria dos casos, a redução de eventuais ruídos, artefactos ou outro tipo de distorção.

Descrever-se-á nas próximas Secções os tratamentos aplicados às imagens na fase de pré-processamento. De seguida, apresentar-se-ão de forma exaustiva os métodos mais frequentemente utilizados na segmentação de imagens médicas, que podem ser classificados, com base na interacção com o utilizador, em manuais, semi-automáticos, e automáticos.

3.2 Pré-Processamento

Antes de se efectuar a segmentação, é necessário efectuar alguma manipulação às imagens. Uma das manipulações características na segmentação do tracto vocal é a eliminação do ar exterior ao tracto vocal. Este assunto será tratado na Secção 3.2.1. Outro tipo de pré-processamento, mais abrangente e utilizado em vários tipos de imagem é a atenuação de ruído (Secção 3.2.2).

3.2.1 Delimitação do tracto vocal

No processo de segmentação de imagens de RM do tracto vocal, uma das principais dificuldades que foi preciso contornar foi a delimitação do tracto vocal. Para fazer com que os contornos gerados pelas técnicas de segmentação não se expandissem à área da imagem exterior ao tracto vocal, optou-se por recorrer a um estratagema prévio à aplicação da técnica de segmentação (pré-processamento). Este pré-processamento foi baseado no único trabalho

conhecido que foca esta problemática, o trabalho de Behrends et al. (2003).

Através da adopção de um *kernel*, Equação 3.1, operou-se uma convolução a duas dimensões na parte mais à esquerda da imagem. Verificou-se empiricamente que o tamanho de 30 para o *kernel* é o que resulta num “esborratamento” mais apropriado. Processando desta forma a imagem, esta torna-se desfocada. Uma das consequências é a formação de uma “mancha” que cobre toda a área em que se encontra o falante, distinguindo-se do fundo negro, delimitando assim a área do tracto vocal.

$$K = \begin{pmatrix} 0 & \dots & 0 & 1 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 1 & 0 & \dots & 0 \end{pmatrix} \quad (3.1)$$

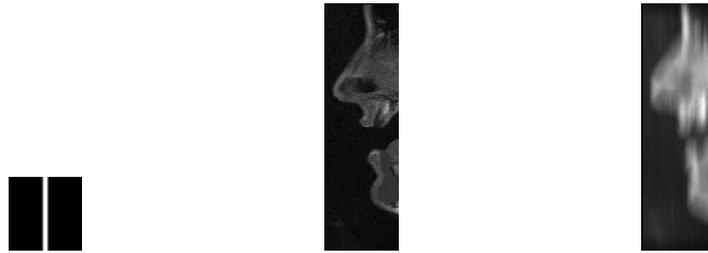


Figura 3.1: Eliminação do ar exterior: *kernel* * Porção à esquerda da imagem = Resultado da Convolução.

Ao resultado desta convolução é aplicado o algoritmo *Region Growing* (que será abordado de forma mais aprofundada na subsecção 3.4.5), no qual a semente foi posicionada num ponto pertencente ao fundo da imagem, logo, na área exterior ao tracto vocal. O valor de limiar utilizado foi 0,07. Os resultados obtidos foram, então, assim processados: todos os valores detectados como pertencentes à mesma categoria da semente dada foram substituídos pelo valor 0. Os restantes substituídos pelo valor 1. Obteve-se assim a indicação dos pixels pertencentes ao fundo da imagem. Excluídos estes pixels, os algoritmos de segmentação são aplicados a imagens como a que se mostra na Figura 3.2.

3.2.2 Atenuação de ruído

As técnicas de segmentação de imagem médicas nem sempre levam à obtenção de contornos suficientemente exactos, sendo estas, na maioria das vezes, corrompidas por ruído e artefactos, (Lin et al., 2004a). Por esta razão, é necessário efectuar um tratamento prévio à imagem a



Figura 3.2: Imagem após o pré-processamento

segmentar, através da utilização de filtros, cujo funcionamento examinar-se-á mais adiante. É importante, contudo, realçar que determinadas características da imagem, essenciais à sua posterior segmentação, não devem ser alteradas pela utilização destes filtros. Perona e Malik (1990) apontam como critérios que qualquer candidato a paradigma deve satisfazer para gerar descrições de escala múltipla “com significado semântico”:

- causalidade - não devem ser gerados detalhes ausentes na imagem original;
- localização imediata - o limite da região na imagem filtrada deve coincidir com o respectivo limite na imagem original;
- parcialmente suave - não deve ocorrer suavização inter-regiões, mas sim dentro duma região.

Nesta Secção, serão descritas algumas técnicas de pré-processamento usualmente utilizadas.

Filtro de mediana

O filtro de mediana é um filtro não linear onde o valor de cada pixel de saída é determinado como sendo a mediana dos pixels vizinhos, Equação 3.2. Este filtro mantém as arestas, mas resulta numa perda de resolução por suprimir os pequenos detalhes, Figura 3.3.

$$I(x, y) = \text{mediana}(I(x - k, y - l)) \quad (3.2)$$

onde $(k, l) \in w$ e w é uma janela adequada.

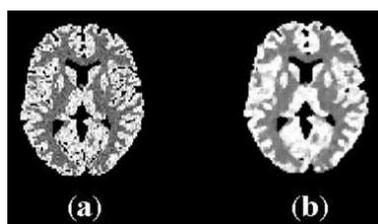


Figura 3.3: Exemplificação do filtro de mediana: (a) imagem original, (b) imagem processada com um filtro de mediana 3×3 . Adaptado de (Murase et al., 2001).

Filtro de Difusão Anisotrópica

O Filtro de Difusão Anisotrópica caracteriza-se por reduzir o ruído sem esborratar os detalhes morfológicos. Este filtro não linear foi inspirado na difusão anisotrópica do calor e foi descrito pela primeira vez por Perona e Malik (1990).

Foi posteriormente aplicado a imagens de RM em (Gallo et al., 1997; Atkins e Mackiewicz, 1998; Murase et al., 2001), entre outros. Veja a Figura 3.4 para uma ilustração deste método.

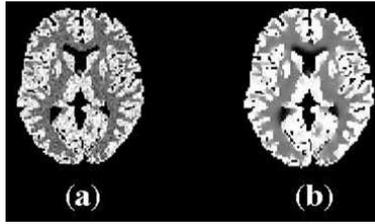


Figura 3.4: Exemplificação do Filtro de Difusão Anisotrópica: (a) imagem original, (b) imagem processada com o Filtro de Difusão Anisotrópica. Adaptado de (Murase et al., 2001).

O método é caracterizado pela Equação 3.3

$$\frac{\partial}{\partial t}I(x, y, t) = \text{div}(d(x, y, t)\nabla I(x, y, t)) \quad (3.3)$$

onde (x, y) são as coordenadas de cada ponto da imagem de RM, representada por I , t corresponde à iteração, $d(x, y, t)$ é uma função monotonamente decrescente, denominada função de difusão, div é o operador de divergência, e o símbolo ∇ representa o gradiente.

O nome anisotrópico é devido à função de difusão $d(\cdot)$ não ser constante, mas uma função do gradiente ∇I . Contrariamente, caso a difusão fosse igual em todos os pontos da imagem, o processo seria denominado isotrópico (Voci et al., 2004).

Perona e Malik (1990) propõem para a função de difusão a expressão dada pela Equação 3.4

$$d(x, y, t) = e^{-(\|\nabla I(x, y, t)\|/K)^2} \quad (3.4)$$

Esta função de difusão privilegia arestas com contrastes altos em detrimento das arestas com contrastes baixos.

Outra função de difusão comum é dada em (Perona e Malik, 1990) por:

$$d(x, y, t) = \frac{1}{1 + (\nabla I(x, y, t)/K)^2} \quad (3.5)$$

Esta função de difusão privilegia regiões grandes, em detrimento de regiões mais pequenas.

O comportamento do filtro depende do valor de K , a constante de difusão. Gerig et al. (1992) propõem a utilização da função de difusão (3.4) com uma constante $K = 2,0$. As experiências levadas a cabo por esta equipa em imagens de RM de cérebros concluíram que este valor, com um número de iterações igual a 3, seria o que conduzia a resultados mais satisfatórios. Para volumes, com a mesma função de difusão, os autores utilizaram $K = 5,0$ e um número de iterações igual a 3.

No mesmo artigo, são definidos valores máximos para os passos de integração: $1/5$ para dados 2D, considerando uma vizinhança de 4 pixeis; $1/7$ para dados 2D, considerando uma vizinhança de 8 pixeis; $1/7$ para dados 3D, considerando uma vizinhança de 6 pixeis; e $3/44$ para dados 3D, considerando uma vizinhança de 26 pixeis.

Note-se ainda que se se utilizar $d(x, y, t) = 1$, obtém-se o filtro de Gaussiana, que, apesar de ser um dos filtros mais utilizados, tem como desvantagem o facto de “esborratar” os objectos. Isto significa que a suavização não é feita apenas dentro de cada região, mas também entre regiões.

No que diz respeito à sua implementação, a Equação 3.3 pode ser discretizada, através do método das diferenças finitas, da seguinte maneira (Helbing e Orglmeister, 1993):

$$I_x = \frac{1}{2h}(I(x+h, y) - I(x-h, y)) \quad (3.6)$$

$$I_y = \frac{1}{2h}(I(x, y+h) - I(x, y-h)) \quad (3.7)$$

$$I_{xx} = \frac{1}{4h^2}(I(x+2h, y) - 2I(x, y) + I(x-2h, y)) \quad (3.8)$$

$$I_{yy} = \frac{1}{4h^2}(I(x, y+2h) - 2I(x, y) + I(x, y-2h)) \quad (3.9)$$

onde h é o passo de discretização e os sub-índices representam derivadas, indicando a sua ordem e parâmetro de derivação.

A função discreta d , dada pela Equação 3.4, é calculada separadamente nas direcções x e y :

$$d(I_x) = e^{-(|I_x|/\sqrt{2K})^2} \quad (3.10)$$

$$d(I_y) = e^{-(|I_y|/\sqrt{2K})^2} \quad (3.11)$$

A função discreta d , dada pela Equação 3.5, é também calculada separadamente nas direcções x e y :

$$d(I_x) = \frac{1}{1 + (|I_x|/K)^2} \quad (3.12)$$

$$d(I_y) = \frac{1}{1 + (|I_y|/K)^2} \quad (3.13)$$

onde $|\cdot|$ denota o valor absoluto. d é calculado em cada iteração e a imagem é calculada (também em cada iteração) segundo,

$$I^{t+1} = I^t + d(I_x) \cdot I_{xx}^t + d(I_y) \cdot I_{yy}^t \quad (3.14)$$

O processo aqui explicado foi implementado em Matlab segundo as instruções presentes em (Helbing e Orglmeister, 1993).

Wavelets

As *Wavelets* proporcionam uma análise das médias e dos detalhes a várias escalas, permitindo um processamento localizado em espaço e frequência. A eliminação do ruído resulta, normalmente, de, para uma dada escala, se considerarem irrelevantes os coeficientes, anulando-os a partir de operações de Limiarização, Figura 3.5.

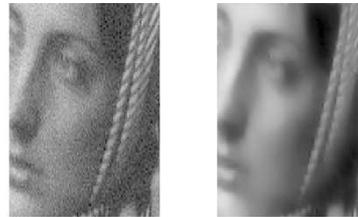


Figura 3.5: Exemplificação do filtro de *Wavelets*: (a) imagem original, (b) imagem processada com um filtro de *Wavelets*. Adaptado de (Misiti et al., 2007).

A principal diferença deste método para os outros dois descritos nesta Secção, o filtro de mediana e o filtro de Difusão Anisotrópica, é que este actua num espaço transformado, enquanto que os outros dois actuam no espaço da imagem. No entanto, *Wavelets* e a Difusão Anisotrópica, partilham o facto de ambos serem representações multi-escala.

Em (Silva, 2005) foi escolhida como função de base a função Haar e a decomposição foi efectuada até ao nível dois. A *Wavelet* Haar é descontínua e assemelha-se a uma função degrau, Figura 3.6. Esta é a única *Wavelet* ortogonal, simétrica, de valores reais, suportada de forma compacta (Silva, 2006).

A função Haar pertence à família das *Wavelets* ortonormais suportadas de forma compacta, com fase extrema e o maior número possível de momentos para um dado tamanho de suporte. Na Figura 3.7 são apresentadas outras nove funções deste tipo (normalmente denominadas “Daubechies”).

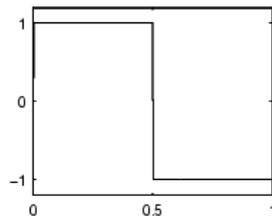


Figura 3.6: *Wavelet* Haar. Adaptado de (Misiti et al., 2007).

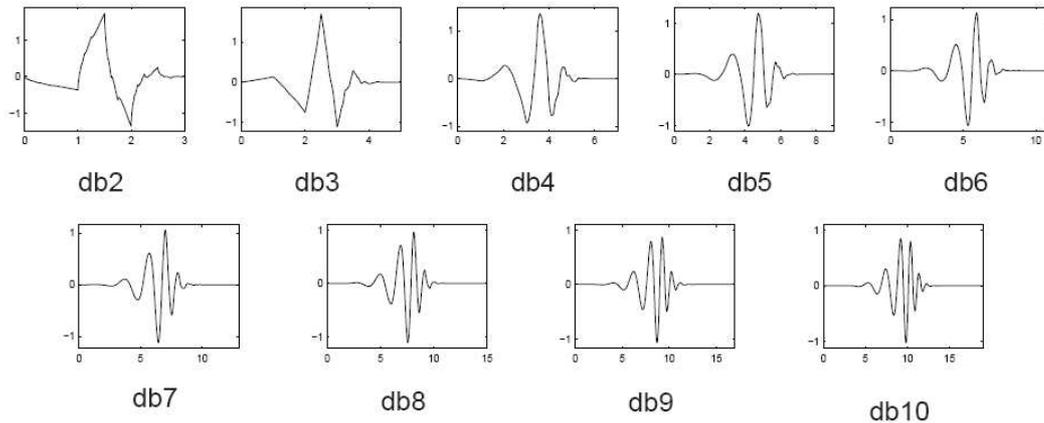


Figura 3.7: *Wavelets* Daubechies. Adaptado de (Misiti et al., 2007).

Métricas de avaliação do efeito dos métodos de atenuação de ruído

A avaliação do pré-processamento consiste na comparação entre a imagem original e a imagem obtida após a aplicação dos filtros, em termos de quantificação do ruído presente numa imagem, antes e depois do seu pré-processamento. Este procedimento não representa uma tarefa de fácil execução: ainda hoje, a inspeção visual continua a ser a melhor técnica para avaliar a diminuição do ruído e verificar a conservação das características essenciais da imagem.

Contudo, existe a possibilidade de implementar métricas que quantifiquem a atenuação de ruído. Apresentam-se de seguida algumas destas métricas.

Diferenças pixel a pixel

Este é o método mais simples e consiste no cálculo das diferenças absolutas entre pixels correspondentes nas imagens antes e depois de serem suavizadas.

Erro quadrático médio

O erro quadrático médio (MSQE) é dado por

$$MSQE = \frac{1}{N} \sum_{i=1}^N (I(x_i, y_i) - I'(x_i, y_i))^2 \quad (3.15)$$

onde I' é a imagem após o processo de atenuação de ruído.

Peak signal to noise ratio

O *Peak signal to noise ratio* (PSNR) representa a diferença entre a imagem original e a imagem após o processo de atenuação de ruído (Ichigaya et al., 2006).

$$PSNR = 20 \log_{10} \frac{S_{p-p}}{\sqrt{MSQE}} \quad (3.16)$$

onde S_{p-p} é o pico da amplitude do sinal e corresponde a 255 no caso de se estar a trabalhar com uma representação em 8 bits.

3.3 Segmentação manual

A segmentação manual consiste na identificação de contornos feita por especialistas na área em questão ou por pessoas consideradas como tendo experiência suficiente para classificar uma região da imagem como pertencente, ou não, à estrutura que se deseja identificar. Esta segmentação poderá ser feita manualmente e depois digitalizada (Soquet et al., 2002; Demolin et al., 2003) para futuro processamento. Pode ainda ser feita através da utilização de *softwares* de desenho em combinação com periféricos como *trackball* (Greenwood et al., 1992; Baer et al., 1991), *graph pen* (Baer et al., 1991) ou *graphic pad* (Demolin et al., 1996). Na Figura 3.8 estão ilustrados alguns exemplos destes aparelhos.



Figura 3.8: Periféricos: Microsoft Trackball Explorer (esquerda), GENIUS Graph Pen (centro), TechDis graphic pad (Direita).

Em muitos estudos, alguns deles na área da fala (Shadle et al., 1999; Engwall, 2000; Pluempitwiriwaj et al., 2005), este tipo de segmentação é considerado o mais fidedigno. No entanto, entre as principais desvantagens encontra-se a “variabilidade inter-sujeitos” e até “intra-sujeito” (Secção 3.1). Silva (2005) verificou que especialistas que trabalham em sítios diferentes e com diferentes anos de serviço têm tendência para traçar delimitações diferentes (variabilidade inter-observador), e também que um mesmo especialista, ao fazer o contorno de uma mesma imagem em períodos diferentes, efectua contornos significativamente diferentes (variabilidade intra-observador).

3.4 Segmentação semi-automática

Os métodos de segmentação semi-automática são assim chamados pois necessitam de alguma intervenção do utilizador na sua inicialização e na definição de alguns parâmetros fixos. Nestes métodos de segmentação, e após a intervenção inicial do utilizador, o contorno ajusta-se à região de interesse, com uma fiabilidade variável de acordo com a experiência do utilizador e a qualidade do método.

Ao longo desta Secção, explicar-se-ão alguns métodos de segmentação semi-automática, começando pelos mais simples e antigos, como a Limiarização e a detecção de arestas, passando por alguns outros métodos de grau mais complexo, como *Region Growing* e *Live Wire*, e por outros cuja matemática em que são baseados é mais avançada, como os Contornos Activos e *Snakes*, e terminando nos métodos considerados como “estado da arte”: Modelos geométricos deformáveis e *Reaction Diffusion Active Contours*.

3.4.1 Limiarização (*Threshold*) global

Na técnica de Limiarização com um nível fixo, o que é feito é escolher o valor de cinzento que corresponde à separação entre a zona que se deseja delimitar e o resto da imagem. Sendo assim, todos os pixels da área em questão devem ter um valor de cinzento superior (ou inferior) ao nível escolhido (Mendonça et al., 2007). Embora o nível (limiar) seja normalmente escolhido de forma manual, através de tentativa/erro, existem alguns métodos de selecção automática (Hentea, 1993; Wilkinson et al., 2003; Ye et al., 2005).

A imagem é transformada segundo a Equação 3.17:

$$I(x, y) = \begin{cases} 1 & \text{se } I(x, y) > \text{nível} \\ 0 & \text{se } I(x, y) \leq \text{nível} \end{cases} \quad (3.17)$$

Na Figura 3.9 apresenta-se uma aplicação desta técnica de segmentação.

Esta técnica, por si só, separa a imagem em regiões, mas não dá informação sobre os contornos. Por isso, é por vezes combinada com técnicas de detecção de contornos.

3.4.2 Limiarização *zoom*

Em (Soquet et al., 1998), é também referido um método de segmentação denominado por Limiarização *zoom*. Este método consiste no aumento da imagem a segmentar (*zoom*) e subsequente esbatimento da imagem. De seguida, é aplicada a Limiarização, exactamente da mesma forma como foi descrito na sub-Secção anterior.

3.4.3 Limiarização adaptativa

Neste tipo de algoritmos, o nível é calculado para cada pixel, tendo por base algumas propriedades locais na vizinhança do pixel em questão. Sendo assim, o nível é transformado

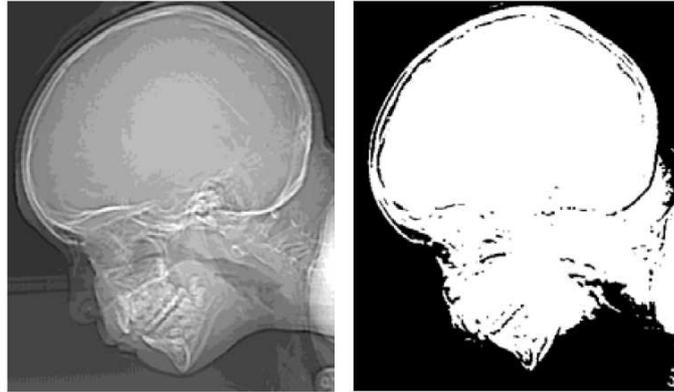


Figura 3.9: Exemplo de aplicação de Limiarização: a) Imagem original; b) Imagem após a aplicação da Limiarização. Adaptado de (Gonzalez e Woods, 2002).

numa função que depende das coordenadas (x, y) de cada pixel.

Sezgin e Sankur (2004) dividem as propriedades locais da imagem que determinam o valor para o nível em cada posição em: (1) variância local, (2) contraste local, (3) “vizinhos do centro” (*center-surround*) (4) ajuste de superfície, e (5) método de *Kriging*. Este tipo de algoritmos é, usualmente, classificado como segmentação automática. No entanto, das 5 classes descritas por (Sezgin e Sankur, 2004) só duas (pontos (4) e (5)) é que não necessitam do ajuste de parâmetros. Decidiu-se, por isso, incluir este tipo de métodos na secção de técnicas de segmentação semi-automáticas.

A principal utilidade das técnicas de segmentação baseadas em Limiarização adaptativa é, tal como se exemplifica na Figura 3.10, a sua aplicação a imagens com iluminação não-uniforme.

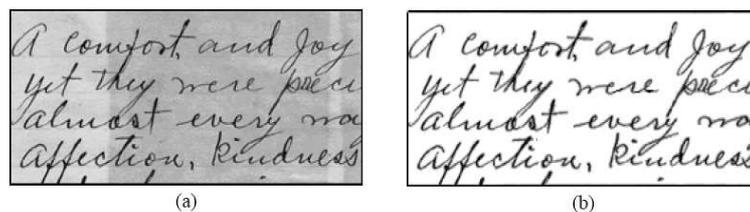


Figura 3.10: Exemplo de aplicação da Limiarização adaptativa: a) Imagem original; b) Imagem após a aplicação da Limiarização adaptativa. Adaptado de (Mitra, 2005).

3.4.4 Detecção de arestas

Os métodos de detecção de arestas determinam a localização de arestas numa imagem. Por arestas entende-se descontinuidades na imagem em relação a uma certa propriedade (cor, textura, etc). A sua detecção é feita através da utilização de operadores como o gradiente, operador de Roberts, operador de Sobel, operador de Prewitt, operador de Canny, etc.

Por ser um dos métodos mais utilizados, apresenta-se no Algoritmo 1, o algoritmo do detector de arestas Canny (Sonka et al., 1999).

Algoritmo 1 - Detector de arestas Canny

- 1: Convolução da imagem com a Gaussiana, G , com desvio padrão σ ;
 - 2: Estimação das direcções normais às arestas locais, Equação 3.18, para cada pixel da imagem I ;
 - 3: Cálculo da Equação 3.19 para estimação das arestas locais;
 - 4: Cálculo da magnitude da aresta através da Equação 3.20;
 - 5: Aplicação de um limiar às arestas da imagem de forma a eliminar respostas em excesso;
 - 6: Repetição dos passos 1 a 5 para valores de σ cada vez maiores;
 - 7: Junção da informação final sobre as arestas em diferentes escalas.
-

$$n = \frac{\nabla(G * I)}{|\nabla(G * I)|} \quad (3.18)$$

$$\frac{\partial^2}{\partial n^2} G * I = 0 \quad (3.19)$$

$$|G_n * I| = |\nabla(G * I)| \quad (3.20)$$

onde $G_n = \frac{\partial G}{\partial n} = n \cdot \nabla G$.

A Figura 3.11 apresenta um exemplo de aplicação deste detector de arestas.

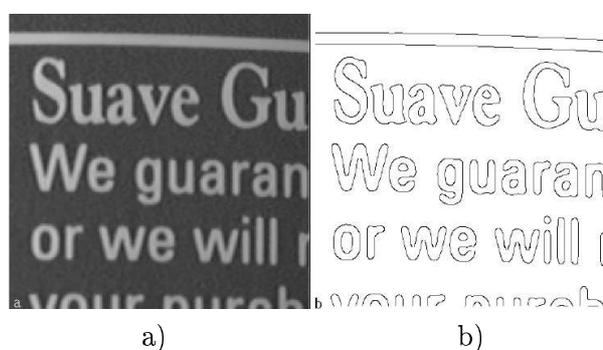


Figura 3.11: Operador de Canny. a) Imagem original; b) Imagem após a aplicação do detector de arestas com o operador de Canny. Adaptado de (Ding e Goshtasby, 2001).

Um outro detector de arestas interessante, e que se irá descrever por ser necessário para um outro método de segmentação descrito na Secção 3.4.6, é o Laplaciano. O Laplaciano detecta arestas após a filtragem da imagem com o Laplaciano do filtro de Gaussiana, e é por isso frequentemente denominado por LoG. A máscara de convolução do operador LoG é (Sonka et al., 1999):

$$h(x, y) = \left(\frac{x^2 + y^2 - \sigma^2}{\sigma^4} \right) e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (3.21)$$

onde σ é o desvio padrão do *kernel* Gaussiano.

Os problemas relacionados com este tipo de técnicas são a geração de falsas arestas, ou o não reconhecimento de arestas reais.

Este tipo de técnicas, tal como as de Limiarização, são consideradas operações de baixo nível, que classificam cada pixel de acordo com as propriedades locais da imagem e, por si só, não extraem a região de interesse (Silva et al., 1996). Necessitam por isso da posterior aplicação de uma rotina de junção de arestas de modo a formarem um método de segmentação (Abrantes e Marques, 1995a,b, 1996; Nascimento, 1998).

3.4.5 *Region Growing*

Neste método, o utilizador escolhe um ponto denominado “semente” e, com base nas características da imagem, a região vai-se “dilatando” e ajustando aos limites da região de interesse, Figura 3.12. As características utilizadas são, tipicamente, a intensidade dos pixels, ou as arestas calculadas por um método de detecção de arestas.

A implementação deste método foi feita em Matlab e pode ser dada pelo pseudo-código 2:

Algoritmo 2 - *Region Growing*

```
Etiqueta(Semente)=Interior;
Media=media(Semente)
SSL=Vizinhos(Semente);
enquanto SSL não estiver vazio fazer
  PontoTeste=primeiro ponto do SSL
  SSL=SSL - primeiro ponto
  para cada Vizinho do PontoTeste que ainda não esteja etiquetado fazer
    se abs(vizinho-Media)<limiar então
      Etiqueta(vizinho)=Interior;
      Actualizar a media
      SSL=SSL + vizinho
    senão
      Etiqueta(vizinho)=Exterior;
  fim se
fim para
fim enquanto
```

Este método é muito sensível a ruído presente na imagem, mas tem a vantagem de ser facilmente generalizado para aplicação à segmentação de volumes, embora com elevados custos computacionais. Outros problemas relatados na literatura são a impossibilidade de mudança de topologia (cada parte de um mesmo objecto tem que ser inicializada com uma semente diferente), a dificuldade da escolha da semente inicial, e a definição do critério de paragem (Silva et al., 1996).

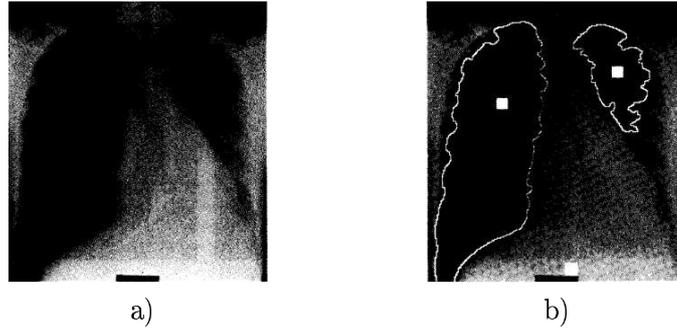


Figura 3.12: *Region Growing*. a) Imagem original; b) Os quadrados a branco representam a semente e a linha a branco representa o contorno final. Adaptado de (Adams e Bischof, 1994).

3.4.6 *Live Wire (Intelligent Scissors)*

A “detecção inteligente” baseia-se na ideia de procura num grafo bidimensional, tirando partido das capacidades humanas em reconhecer objectos e das capacidades computacionais na delimitação dos mesmos (Falcão et al., 2000).

O utilizador começa por escolher um ponto inicial (semente) no limite do objecto de interesse. De seguida, move o cursor ao longo da imagem. A rotina calcula um contorno (ou *Live Wire*) desde o ponto inicial até ao ponto onde se encontra o cursor, que é dinamicamente disponibilizado ao utilizador. Se o contorno disponibilizado é aceitável, o utilizador “clica” com o rato, e o ponto é transformado numa nova semente. O contorno torna-se então parte do limite do objecto. Sendo assim, o contorno final é óptimo por zonas (isto é, óptimo entre sementes escolhidas pelo utilizador) (Liang et al., 2006). Este método é então apropriado para a segmentação de imagens não homogéneas (McInerney e Terzopoulos, 2000).

Fazem parte desta ferramenta o *boundary cooling* (geração automática de sementes) e o treino *on-the-fly* (adaptação do contorno dinâmico a arestas de interesse), que reduzem a participação do utilizador e adaptam dinamicamente a ferramenta aos tipos específicos de contornos pretendidos (Barrett e Mortensen, 1997).

O algoritmo calcula o caminho óptimo desde uma semente (definida pelo utilizador) até todos os pontos na imagem (Mortensen et al., 1992). Este caminho óptimo é calculado com base em funções de custo. Em (Chodorowski et al., 2005) é definida a função de custo associada ao trajecto do pixel p ao pixel q , $C(p, q)$, como:

$$C(p, q) = w_Z f_Z(q) + w_C f_C(q) + w_G f_G(q) + w_D f_D(p, q) \quad (3.22)$$

onde cada termo de custo é pesado por uma constante, w_Z , w_C , w_G e w_D . Os termos $f_Z(q)$ e $f_C(q)$ são os detectores de arestas Laplaciano e Canny, respectivamente, e foram já explicados na Secção 3.4.4. De forma a que a arestas fortes correspondam custos baixos, estes dois termos

foram invertidos. O termo $f_G(q)$ é dado por:

$$f_G(q) = 1 - G/\max(G) \quad (3.23)$$

onde $G(q)$ é a magnitude do gradiente de cor no pixel q e $\max(G)$ representa o gradiente com maior magnitude da imagem. O termo $f_D(p, q)$ é dado por:

$$f_D(p, q) = \text{acos}(D_x(p)/G(p) * D_x(q)/G(q) + D_y(p)/G(p) * D_y(q)/G(q))/\pi \quad (3.24)$$

onde $D_x(p)$ e $D_y(p)$ são os vectores próprios correspondentes ao maior valor próprio para os gradientes nas direcções x e y do pixel p , respectivamente.

A função $C(p, q)$, sofre por fim uma transformação de escala por $\sqrt{2}$ se o pixel q é um vizinho na diagonal de p .

A principal desvantagem deste algoritmo é a grande necessidade de interacção com o utilizador, apesar de se traduzir num relevante poder de controlo sobre o contorno final. Uma outra dificuldade é relacionada com o nível de detalhe desejado. Barrett e Mortensen (1997) dão o exemplo da extracção de cabelo numa imagem. Esta tarefa aumenta consideravelmente a intervenção do utilizador. Por fim, e embora já haja algum esforço nesse sentido (Poon et al., 2007), esta técnica não é facilmente extensível a 3D.

Outras aplicações desta técnica, que não a segmentação, são a composição de imagens digitais e extracção de linhas a partir de documentos digitalizados.

3.4.7 Contornos Activos

Contornos activos são contornos que se deformam segundo a influência de forças internas e externas de forma a localizar os limites de um objecto (Kass et al., 1987). Podem ser classificados em Contornos Activos Paramétricos ou Contornos Activos Geométricos, segundo a sua representação e implementação (El-Zehiry et al., 2005).

Contornos Activos Paramétricos ou *Snakes*

Os contornos activos paramétricos são curvas paramétricas que se deformam consoante as propriedades da imagem (como, por exemplo, linhas ou arestas de um objecto na imagem) até ser atingida a estabilidade, Figura 3.13.

A adaptação da curva às propriedades da imagem é feita através da minimização de uma função de energia. Esta função depende de vários parâmetros que têm que ser ajustados dependendo do problema em questão.

Na Equação 3.25 pode-se verificar que a função de energia está dividida em duas partes: uma dependente da curva - energia interna (primeiros dois termos) e outra dependente da imagem

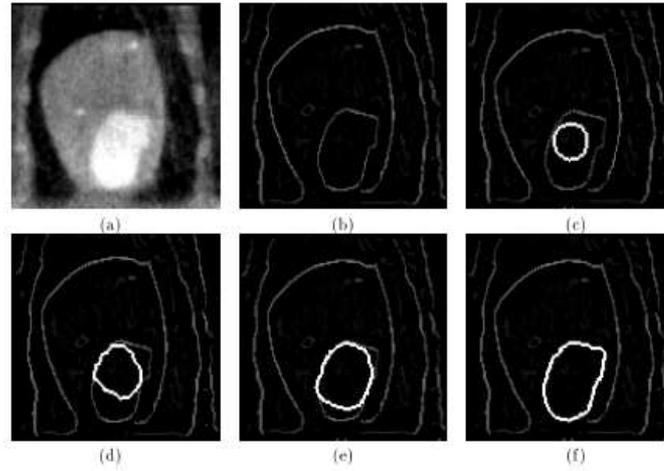


Figura 3.13: *Snake*. a) Imagem original; b) Imagem após um algoritmo de detecção de arestas c) *snake* original; d)-f) deformação da *snake* em direcção ao contorno. Adaptado de (McInerney e Terzopoulos, 1995).

- energia externa (terceiro termo).

$$E(C(s)) = \alpha \int_0^1 \left| \frac{\partial C(s)}{\partial s} \right|^2 ds + \beta \int_0^1 \left| \frac{\partial^2 C(s)}{\partial s^2} \right|^2 ds - \lambda \int_0^1 |\nabla I(C(s))|^2 ds \quad (3.25)$$

onde C é a curva, α , β e λ são constantes reais positivas, e s é a variável da parametrização.

A primeira e a segunda derivada determinam o comportamento da curva (a sua elasticidade e rigidez) e o terceiro termo atrai a curva ao contorno desejado.

Ao minimizar a energia dada pela Equação 3.25, está-se a tentar alocar a curva aos pontos onde $|\nabla I(C(s))|$ é máximo. Cohen e Cohen (1993) aconselham a leitura de (Fua e Leclerc, 1988) para uma discussão da relação entre a minimização da energia e a localização de contornos.

Na prática, a energia interna é minimizada segundo a equação diferencial de Euler-Lagrange (Dzyubachyk et al., 2007), seguida duma aproximação das diferenças ascendentes para substituir as derivadas temporais. A equação obtida, Equação 3.26, é resolvida de forma iterativa até ser atingida a condição de paragem escolhida (Silva, 2005).

$$V_{n+1} = (Id - \tau A)V_n + \tau F \quad (3.26)$$

onde V são as coordenadas dos pontos da *snake*, Id é a matriz identidade, τ é uma constante, A é uma matriz onde estão representadas as forças internas da *snake*, e F é o conjunto das forças externas.

Para mais pormenores sobre a derivação da Equação 3.26 e respectiva implementação, consulte (Kass et al., 1987; Cohen e Cohen, 1993).

Existem algumas outras terminologias para este mesmo método. Em (Soquet et al., 1998) ele é denominado *Elastic method* e, em (Demolin et al., 1998), é denominado *Elastic matching*.

Este método, embora seja tenha ampla aplicação, tem o problema de não poder ser usado directamente quando a zona a segmentar muda a sua topologia, ou seja, quando a curva se divide, desaparece, ou quando várias curvas se juntam - (McInerney e Terzopoulos, 2000) apresenta uma excepção. Para além deste problema, é ainda necessário que a inicialização da curva seja muito próxima do objecto a segmentar, porque se utiliza somente informação local ao longo do contorno (Shah, 2007). Uma outra desvantagem, referida em (Nascimento e Marques, 2005), é a detecção de regiões que não se encontram localizadas no limite do objecto, mas são causadas por outros objectos ou por arestas interiores. A sua extrapolação para o universo 3D também não é linear, embora existam já algumas implementações como (Cohen e Cohen, 1993; DeCarlo e Metaxas, 1998; Nascimento e Marques, 1999).

Modelos geométricos deformáveis

Os modelos geométricos deformáveis (*Geodesic Active Contours*) são alternativas geométricas às *Snakes* que não têm limitações no que diz respeito às mudanças de topologia durante a evolução da curva. Esta característica é normalmente considerada uma vantagem. Existem, no entanto, alguns casos em que ela se torna uma desvantagem. Imagine-se por exemplo, que se sabe à partida a topologia do objecto a segmentar. No nosso campo de aplicação em específico, esta característica é desejável, dado que o tracto vocal tanto pode ser um objecto conexo, como se pode encontrar dividido em duas ou até três partes.

Nos modelos geométricos deformáveis, a curva é direccionada para as características da imagem de forma óptima, sendo que cada ponto é movimentado segundo a direcção normal à curva, Equação 3.27 (Rendon et al., 2006).

$$\frac{dC}{dt} = (g(C)k - \nabla g \cdot \vec{N} - cg(C))\vec{N} \quad (3.27)$$

onde

\vec{N} é a normal à curva

k é a curvatura local

c é uma constante

g é a função de paragem

As vantagens em relação às *Snakes* são que, para além de as mudanças de topologia serem automaticamente resolvidas, a inicialização da curva não necessita de ser próxima do objecto a segmentar, Figura 3.14. Em (Han et al., 2003) é mesmo afirmado que os modelos geométricos deformáveis clássicos são insensíveis à inicialização. De forma contrária, em

(Li et al., 2003) é referido que a obtenção de bons resultados depende da localização das curvas iniciais e que estas devem estar localizadas aproximadamente de forma simétrica em relação ao contorno do objecto. Com a finalidade de contribuir para o esclarecimento deste problema, são estudadas neste trabalho diversas formas de inicialização das curvas.

Como desvantagem, McInerney e Terzopoulos (2000) apontam a dificuldade em adicionar mecanismos de controlo adicionais, como energias de deformação interna e interactividade com o utilizador, devido à sua formulação implícita.

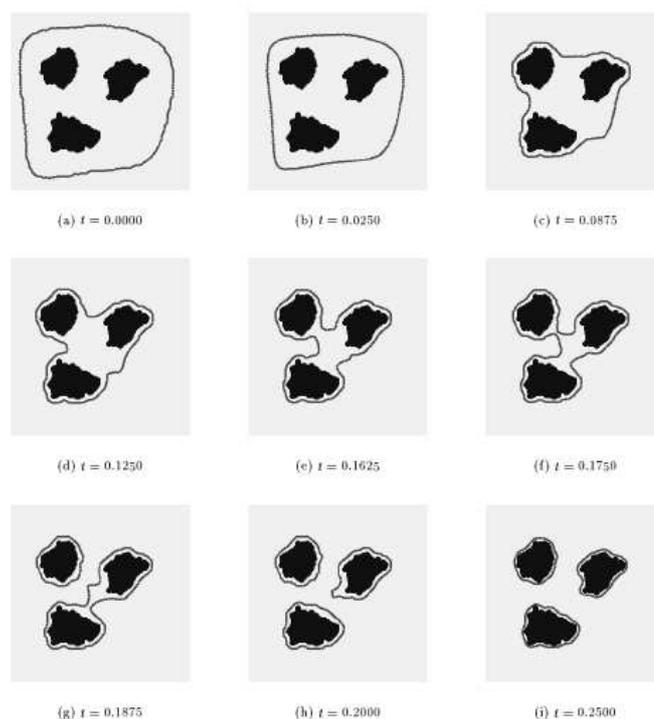


Figura 3.14: Ajuste e divisão de um contorno inicial a três regiões de interesse. Adaptado de (Malladi et al., 1995).

O facto de lidar de forma automática com as mudanças topológicas, é devido à sua implementação ser baseada em *Level Sets*, Figura 3.15. O *Level Sets* foi introduzido por Osher e Sethian (1988) e baseia-se no trabalho (Sethian, 1985), publicado três anos antes (Sethian, 1997). Estes assentam na resolução iterativa duma equação diferencial do tipo Hamilton-Jacobi. Optou-se por dedicar toda a Secção seguinte ao *Level Sets*, de forma a se poder examinar esta técnica com maior pormenor, uma vez que apresenta particular interesse no desenvolvimento da presente investigação.

Level Sets

A implementação dos modelos geométricos deformáveis pode ser baseada em *Level Sets*. Neste método, utiliza-se uma função ϕ , de uma dimensão acima da curva C . A curva C está contida

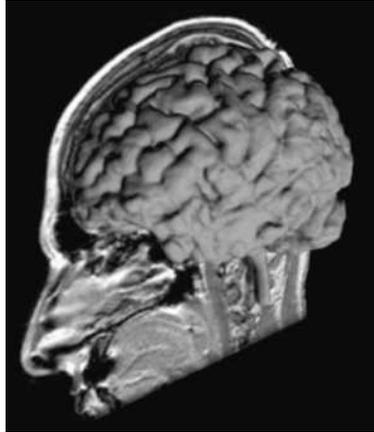


Figura 3.15: Segmentação 3D do cérebro através da utilização de *Level Sets*. Adaptado de (Johnson e Weinstein, 2006).

na função ϕ , sendo o seu nível zero (daqui o nome *Level Sets*), Figura 3.16.

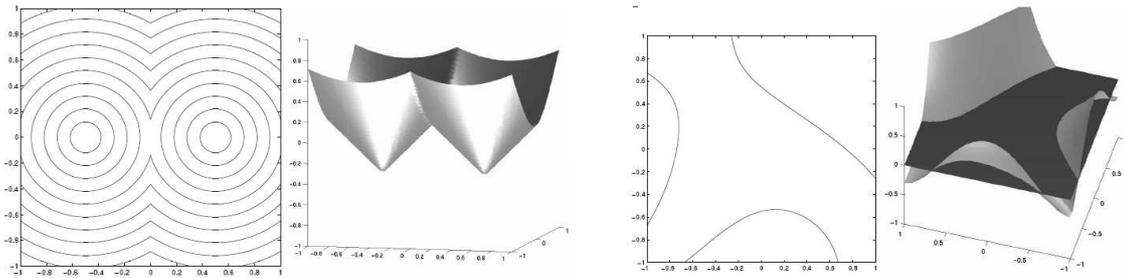


Figura 3.16: Exemplos de funções ϕ e respectivos níveis zero. Adaptado de (Tsai e Osher, 2004).

Tipicamente, ϕ é um mapa de distâncias de C . Na prática, ϕ é uma malha rectangular de valores reais, e tem a mesma dimensão que a imagem (Rendon et al., 2006).

Para mover cada ponto de C na direcção normal à curva, tem-se a Equação 3.28

$$\frac{dC}{dt} = F_v \vec{N} \quad (3.28)$$

onde F_v é a velocidade com que a curva se propaga. A função ϕ tem que ser ajustada segundo a Equação 3.29

$$\frac{d\phi}{dt} = F_v |\nabla\phi| \quad (3.29)$$

Sendo assim, o modelo geométrico deformável, é dado pela Equação 3.30

$$\frac{d\phi}{dt} = \left(\text{div} \left(g(x, y) \frac{\nabla\phi}{|\nabla\phi|} \right) \right) |\nabla\phi| \quad (3.30)$$

Adicionando o termo constante correspondente à velocidade (*balloon force*), tem-se a Equação

3.31

$$\frac{d\phi}{dt} = \left(cg(x, y) + \operatorname{div} \left(g(x, y) \frac{\nabla\phi}{|\nabla\phi|} \right) \right) |\nabla\phi| \quad (3.31)$$

A função ϕ degenera após cada iteração e tem, por isso, que ser recalculada.

Uma vantagem do *Level Sets* visível nas equações 3.30 e 3.31 é que não são controlados por coeficientes de elasticidade, tal como nos métodos baseados em contornos paramétricos. Não é necessário ajustar as tangentes à curva nem calcular as normais a cada vértice. No *Level Sets*, as normais estão embebidas no sistema, através da divergência do campo do fluxo (Suri et al., 2001).

Funções de velocidade

Por simplificação de notação, nesta Secção, os *Level Sets* irão ser considerados como uma superfície. No entanto, os métodos descritos, são igualmente aplicáveis à segmentação de volumes, com o correspondente *Level Sets* sendo uma função num espaço a quatro dimensões. Esta revisão limita-se a funções de velocidade no domínio de imagens e volumes em tons de cinzento. O estudo das cores foi aqui excluído, por não ter interesse na área de aplicação em questão: o estudo de imagens de RM.

A função de velocidade mais simples é uma função discreta dada por:

$$F_v(x, y) = \begin{cases} 1, & \text{se } I(x, y) \in [T - \epsilon, T + \epsilon] \\ -1, & \text{caso contrário} \end{cases} \quad (3.32)$$

onde T é a mediana dos valores de intensidade pertencentes ao interior da região a segmentar e ϵ controla o intervalo de valores próximos de T que devem ser considerados no interior da região. No entanto, para que se obtenham resultados óptimos, é necessário escolher uma função de velocidade apropriada para a aplicação em questão (Lefohn et al., 2003b). Usualmente, estas são derivadas em função de: (1) curvatura (Equação 3.33), (2) magnitude do gradiente da superfície (Equação 3.34), e (3) de uma função de Limiarização no domínio da imagem (Equação 3.35).

1. Função de velocidade dependente da curvatura:

$$F_v = \frac{1}{2} \nabla \frac{\nabla\phi}{|\nabla\phi|} \quad (3.33)$$

2. A função de velocidade dependente da magnitude do gradiente da superfície, denominada em (Lefohn et al., 2003b) como *rescaling speed term*, é calculada da seguinte forma:

$$F_v = \phi G_\phi - \phi |\nabla\phi| \quad (3.34)$$

onde G_ϕ é a magnitude do gradiente “alvo” (por gradiente “alvo” entende-se o valor de gradiente que se deseja obter, ou seja, gradientes diferentes de zero na vizinhança do *Level Set* de ordem zero e gradientes iguais a zero tanto fora como dentro deste), e $|G_\phi|$ é a magnitude do gradiente na direcção da superfície do *Level Sets*. O parâmetro alvo, G_ϕ , pode ser definido de acordo com a precisão numérica com que se calcula o *Level Sets*. Se se tomar $G_\phi = 1$, esta função transforma-se na distância (não há movimento da superfície).

3. Lefohn e Kniss, no seu trabalho (Lefohn et al., 2003a), apresentam ainda uma função de velocidade muito simples, que depende somente da intensidade da imagem, definida por:

$$F_v = \epsilon - |I - T| \quad (3.35)$$

Nesta função, T controla o brilho da região a segmentar e ϵ controla o intervalo de valores próximos de T que devem ser considerados no interior da região a segmentar. Deste modo, uma superfície situada em pixeis com valores de intensidade pertencentes ao intervalo $T \pm \epsilon$ irá expandir de modo a incluir estes pixeis. De forma contrária, uma superfície situada em valores de intensidade fora do intervalo irá contrair de forma a excluir os respectivos pixeis. Esta função de velocidade, tal como pode ser visto pela Figura 3.17, é gradual. Sendo assim, o efeito da função de velocidade diminui à medida que a superfície se aproxima dos limites de regiões com valores de cinzento dentro do intervalo $T \pm \epsilon$.

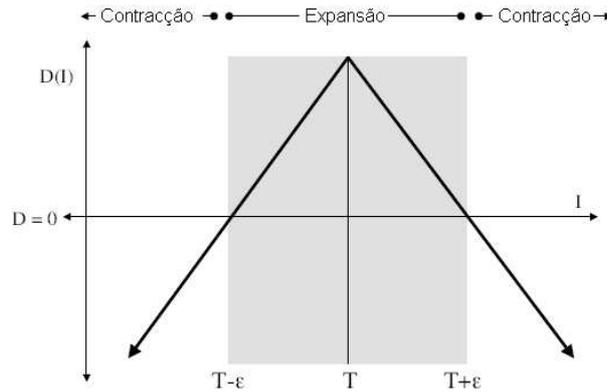


Figura 3.17: Função de velocidade baseada na intensidade da imagem proposta em (Lefohn et al., 2003a).

Suri et al. (2001), integra três termos de velocidade: função de velocidade da curvatura, V_k , função de velocidade regional, V_p , e função de velocidade do gradiente, V_{ext} :

$$V_k = \epsilon k, \epsilon = \frac{\alpha}{\gamma} \quad (3.36)$$

$$V_p = \frac{\omega_p(X)}{\gamma} \vec{N} \quad (3.37)$$

$$V_{ext} = \frac{F_{ext}}{\gamma} \quad (3.38)$$

onde k é a curvatura, γ é o coeficiente de amortecimento, α e β são constantes elásticas, e \vec{N} é a normal unitária da curva.

A função de velocidade regional, V_p , é matematicamente expressa como a combinação interior-exterior da área da região da superfície em evolução. P indica a região e pertence a $]0, 1[$.

A equação de evolução da superfície resultante é:

$$\frac{\partial \phi}{\partial t} = (\epsilon k + V_p) |\nabla \phi| - V_{ext} \nabla \phi \quad (3.39)$$

Uma aproximação um pouco diferente é dada em (Malladi e Sethian, 1995). Aqui consideram-se máximos e mínimos de fluxos:

$$F_v = \begin{cases} k, & \text{se } |\nabla I| < T \\ \text{min/max flow}, & \text{caso contrário} \end{cases} \quad (3.40)$$

onde T é um limiar. A função min/max flow é dada por:

$$\text{min/max flow} = \begin{cases} \text{Max}(k, 0), & \text{se } a(x,y) < G(x,y) \\ \text{Min}(k, 0), & \text{caso contrário} \end{cases} \quad (3.41)$$

onde $a(x, y)$ é a média das intensidades da imagem numa vizinhança (pequena) de (x, y) .

Sendo assim, pontos onde a magnitude do gradiente é maior do que T são preferidos e difusos através do fluxo máximo/mínimo. Os restantes pontos são difundidos através da curvatura. G foi considerado igual a 128 (média entre pixels pretos - 0 - e brancos - 255).

Uma função de velocidade que combina as propriedades da imagem com a curvatura é definida por:

$$F_v = h(I)(\nu - \epsilon k) \quad (3.42)$$

onde, ϵ é uma constante pertencente ao intervalo $]0, 1[$, e representa a importância dada ao termo de regularização (a curvatura k). A curvatura é definida por:

$$k = \text{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) = \frac{\phi_{xx} \phi_y^2 - 2\phi_x \phi_y \phi_{xy} + \phi_{yy} \phi_x^2}{(\phi_x^2 + \phi_y^2)^{3/2}} \quad (3.43)$$

O termo $h(I)$ pode ser calculado de diversas formas. Apresentam-se de seguida dois exemplos

comuns:

$$h(I) = e^{-|\nabla G_\sigma * I|} \quad (3.44)$$

$$h(I) = \frac{1}{1 + |\nabla G_\sigma * I|} \quad (3.45)$$

onde $\nabla G_\sigma * I$ representa a convolução da imagem com um filtro de Gaussiana com desvio padrão igual a σ .

Este critério permite que o modelo pare em zonas onde o gradiente da imagem é alto, através da redução a zero da função de velocidade. Sendo assim, esta função é aconselhada para zonas de interesse com bom contraste, não funcionando bem quando os limites da região a segmentar são mal definidos ou com interrupções.

Wu et al. (2005) redefinem os parâmetros da Equação 3.42 da seguinte forma:

1. Termo de suavização $\epsilon\xi$: O termo ϵk é substituído por $\epsilon\xi$, onde $\xi = \tan^{-1}(w)/(\pi/2)$ e $w = \nabla(\nabla\phi/|\nabla\phi|)$ é constante, mas necessita de ser ajustado de modo a que a superfície em evolução se ajuste à região a segmentar sem que seja atraída pelo ruído normalmente presente em imagens reais. As experiências conduzidas em (Wu et al., 2005) para a segmentação a nível cerebral, aconselham o valor 1,5 para esta constante;

2. Termo de propagação ν :

$$\nu = \begin{cases} +1, & \text{se } \alpha \geq 1 \\ \frac{2\alpha - \alpha_T - 1}{1 - \alpha_T}, & \text{se } \alpha_T < \alpha < 1 \\ -1, & \text{se } \alpha \leq \alpha_T \end{cases} \quad (3.46)$$

onde $\alpha = P/\bar{P}_b$ e α_T é um valor de limiar de α e pertence ao intervalo $[0, 1]$. Os autores aconselham $\alpha = 0,3$. $h(I)$ é uma função normalizada de probabilidade, $h(I) = \sigma\sqrt{2\pi}G(I)$, onde $G(I)$ é a distribuição Gaussiana calculada a partir da agregação de pixels na máscara inicial. P em $S(t = 0)$ é denominada $P_b(t = 0)$ e $P_b(t = t_{n+1})$ herda o valor de $P_b(t = t_n)$ no pixel mais próximo em $S(t = t_n)$, onde S é o *Level Set* de ordem 0. De modo a aumentar a robustez ao ruído, efectua-se uma convolução de $P_b(t = 0)$ com um *kernel*, obtendo-se $\bar{P}_b(t = 0)$;

3. Factor de paragem $h(I)$: $h(I)$ é uma distribuição Gaussiana modificada, $h(I) = \text{Prob}(I)^n$, e é utilizada para reproduzir a probabilidade da deformação da superfície, considerando o nível de intensidade da imagem.

Ciofalo et al. (2004), propõem uma aproximação também baseada na Equação 3.42, mas onde $h(I)$ é substituído por uma função de probabilidade, $-g(P_T)$. Nesta função, $(-g)$ é uma função decrescente e P_T é a probabilidade de transição entre o interior e o exterior da estrutura a segmentar.

Numa segunda fase, foi ainda integrado o gradiente: $h(I) = -g_f(P_T, \nabla I)$, onde

$$g_f(P_T, \nabla I) = g(f(P_T, \nabla I)) \quad (3.47)$$

P_T e ∇I são as entradas de um sistema de decisão *fuzzy* e f é a respectiva saída. Esta decisão *fuzzy* foi desenhada de forma a que quanto mais alta for a entrada, maior será a saída do sistema, $f(P_T, \nabla I)$, Tabela 3.1.

	MB P_T	B P_T	M P_T	A P_T	MA P_T
MB ∇I	MB_s	MB_s	B_s	B_s	M_s
B ∇I	MB_s	B_s	M_s	M_s	A_s
M ∇I	B_s	M_s	M_s	A_s	MA_s
A ∇I	M_s	A_s	A_s	MA_s	MA_s
MA ∇I	A_s	MA_s	MA_s	MA_s	MA_s

Tabela 3.1: Regras da decisão *fuzzy*: MB - muito baixo, B - baixo, M - médio, A - alto, MA - muito alto. Adaptado de (Ciofolo et al., 2004).

Lin et al. (2004a) substituem $h(I)$ por informação de região. Neste caso, $h(I)$ é calculado da seguinte forma:

$$h(I) = \begin{cases} g(I) - L, & \text{se } g(I) < \frac{U-I}{2} + L \\ U - g(I), & \text{caso contrário} \end{cases} \quad (3.48)$$

Com este novo termo e assumindo $\nu = 1$, a função de velocidade será:

$$F_v = \begin{cases} (g(I) - L)(1 - \epsilon k), & \text{se } g(I) < \frac{U-I}{2} + L \\ (U - g(I))(1 - \epsilon k), & \text{caso contrário} \end{cases} \quad (3.49)$$

Intensidades de g entre L e U levam a valores de $h(I)$ positivos (a superfície expande), enquanto que valores de g fora desse intervalo levam a valores de $h(I)$ negativos (a superfície contrai). Mais uma vez, este termo de velocidade é gradual, logo os efeitos de $h(I)$ diminuem à medida que o modelo se aproxima dos limites de regiões cujos valores de intensidade estejam no intervalo entre L e U .

Os mesmos autores, em (Lin et al., 2004b), propõem ainda a integração da função de densidade de probabilidade dentro e fora da estrutura a segmentar. Desta forma, o termo $h(I)$ tem o formato:

$$h(I, s) = e^{-k \cdot p(x_s | y_s)} \quad (3.50)$$

onde

$$\begin{cases} p(x_s|y_s) \propto p(x_s)p(y_s|x_s) \\ p(y_s|x_s) \propto \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_{in}} e^{-\frac{(I(x)-\mu_{in})^2}{2\sigma_{in}^2}}, & \text{dentro do objecto} \\ \frac{1}{\sqrt{2\pi}\sigma_{out}} e^{-\frac{(I(x)-\mu_{out})^2}{2\sigma_{out}^2}}, & \text{no exterior do objecto} \end{cases} \\ p(x_s) = \frac{1}{Z} e^{-\tilde{g}(y_s)} \end{cases} \quad (3.51)$$

onde $p(x_s)$ é a probabilidade no contorno, e $p(y_s|x_s)$ a probabilidade da região a segmentar. Sendo assim, as estatísticas da região são combinadas com uma medida baseada no gradiente. Esta função foi descrita em (Lin et al., 2004b) como sendo robusta em relação à presença de ruído e fronteiras mal definidas.

Baillard et al. (2000), definem a força de propagação num ponto (x, y) como:

$$F_v = SNG(I_{obj} - I_{bgr}) \cdot SNG(I(x, y) - \tilde{I}_{th}) \quad (3.52)$$

onde SNG é uma função que retorna o sinal (positivo ou negativo), I_{th} é um limiar, I_{bgr} é a média da intensidade do fundo na vizinhança do contorno (Figura 3.18), I_{obj} é a média da intensidade do objecto de interesse na vizinhança do contorno (Figura 3.18), e \tilde{I}_{th} é o valor estimado de I_{th} .

$$I_{th} = \begin{cases} \max \left\{ \frac{I_{obj} + I_{bgr}}{2}, I_{obj} + \sigma_{obj} \right\}, & \text{se } I_{obj} < I_{bgr} \\ \min \left\{ \frac{I_{obj} + I_{bgr}}{2}, I_{obj} - \sigma_{obj} \right\}, & \text{caso contrário} \end{cases} \quad (3.53)$$

Os parâmetros I_{bgr} , I_{obj} e I_{th} são estimados em cada iteração usando a última segmentação obtida e a respectiva vizinhança.

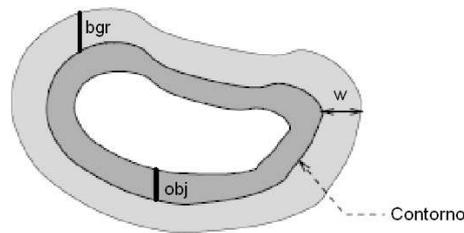


Figura 3.18: As intensidades I_{bgr} e I_{obj} caracterizam o objecto e o fundo, respectivamente. Adaptado de (Baillard et al., 2000).

Suri et al. (2001), publicaram uma revisão bastante completa dos métodos de segmentação através do uso de *Level Sets*.

A implementação do *Level Sets* é, forma geral, bastante lenta. Isto deve-se a dois motivos principais: (1) o número de cálculos por célula da malha é grande e (2) o passo temporal necessário para que haja estabilidade é pequeno (Rendon et al., 2006). Têm, por isso, vindo a ser desenvolvidos métodos para tornar a execução mais rápida. Um dos mais conhecidos é

designado *narrow band approach*. Neste processo, os cálculos são efectuados unicamente na vizinhança da curva (Adalsteinsson e Sethian, 1994). A maior desvantagem é a necessidade de actualizar a banda onde são feitos os cálculos, à medida que a curva se movimenta. Embora esta banda possa ser actualizada somente quando a curva se aproxima dos seus limites (o que já reduz os custos computacionais, comparativamente à actualização da banda em cada iteração) este método continua a ser considerado lento. Uma outra implementação rápida é o *fast marching method* (Sethian, 1996). Este método é bastante mais rápido do que o anterior, mas tem a desvantagem de só funcionar em funções monotonamente crescentes (ou decrescentes). Um outro método, *HERMES*, combina a rapidez do *fast marching* com a aplicabilidade geral da *narrow band* (Paragios e Deriche, 2000). Recentemente, Chopp (2007) propôs uma modificação ao *fast marching method*, que permite a resolução de uma equação de Hamilton-Jacobi estática, sem a restrição de monotocidade imposta à função de velocidade.

Neste trabalho optou-se por uma implementação que conjugasse a rapidez de execução com a facilidade de implementação, baseada em (Shi e Karl, 2005a,b). O algoritmo é descrito no Algoritmo 3. Uma vantagem, para além da rapidez de execução e facilidade de implementação, é que deixa de ser necessário reinicializar o *Level Sets*. Esta reinicialização é uma necessidade noutras implementações porque, à medida que a curva evolui, o *Level Sets* perde as suas características de função de distância, o que provoca instabilidade. Esta necessidade, aumenta ainda o custo computacional e reduz a flexibilidade topológica (Gelas et al., 2007).

Neste pseudo-código, ϕ representa o *Level Sets* (definido como sendo negativo no interior do *Level Set* de ordem zero e positivo no exterior), F a função de velocidade (ver a Secção seguinte), e L_{out} e L_{in} são duas listas de pontos, a primeira contendo os pontos da parte de fora do contorno e a segunda os pontos da parte de dentro:

$$L_{out} = \{x : \phi(x) > 0 \text{ e } \exists y \in N(x) \text{ tal que } \phi(y) < 0\} \quad (3.54)$$

$$L_{in} = \{x : \phi(x) < 0 \text{ e } \exists y \in N(x) \text{ tal que } \phi(y) > 0\} \quad (3.55)$$

por sua vez, $N(x)$ é a vizinhança de x :

$$N(x) = \left\{ y \in D : \sum_{k=1}^K |y_k - x_k| = 1 \right\} \forall x \in D \quad (3.56)$$

Uma simplificação final é feita:

$$\phi(x) = \begin{cases} 3, & \text{se } x \text{ é um ponto exterior} \\ 1, & \text{se } x \in L_{out} \\ -1, & \text{se } x \in L_{in} \\ -3, & \text{se } x \text{ é um ponto interior} \end{cases} \quad (3.57)$$

Algoritmo 3 - Level Sets

Passo 1: Inicialização de ϕ , F , L_{out} e L_{in}

Passo 2:

para $i = 1 : N_a$ fazer

Passo 2.1:

 para cada ponto $x \in L_{out}$ com $F(x) > 0$ fazer

$check_{in}(x)$

 fim para

 para cada ponto $x \in L_{in}$ fazer

 se para todos os $y \in N(x)$, $\phi < 0$ então

 retire x de L_{in} e defina $\phi(x) = -3$

 fim se

 fim para

 para cada ponto $x \in L_{in}$ com $F(x) < 0$ fazer

$check_{out}(x)$

 fim para

 para cada ponto $x \in L_{out}$ fazer

 se para todos os $y \in N(x)$, $\phi > 0$ então

 retire x de L_{out} e defina $\phi(x) = 3$

 fim se

 fim para

Passo 2.2: Verifique a condição de paragem. Se é satisfeita, vá para o **Passo 3**. Senão, continue este ciclo.

fim para

Passo 3:

para $i = 1 : N_g$ fazer

Passo 3.1:

 para todos os pontos $x \in L_{out}$ calcule $G \otimes \phi(x)$ fazer

 se $G \otimes \phi(x) < 0$ então

$check_{in}(x)$

 fim se

 fim para

 para cada ponto $x \in L_{in}$ fazer

 se para todos os $y \in N(x)$, $\phi(y) < 0$ então

 retire x de L_{in} e defina $\phi(x) = -3$

 fim se

 fim para

 para cada ponto $x \in L_{in}$ calcule $G \otimes \phi(x)$ fazer

 se $G \otimes \phi(x) > 0$ então

$check_{out}(x)$

 fim se

 fim para

 para cada ponto $x \in L_{out}$ fazer

 se para todos os $y \in N(x)$, $\phi(y) > 0$ então

 retire x de L_{out} e defina $\phi(x) = 3$

 fim se

 fim para

fim para

Passo 4: Se a condição de paragem no primeiro ciclo é satisfeita, termine o algoritmo, senão, volte ao **Passo 2**

onde os pontos no interior da curva, mas não pertencentes a L_{in} , são denominados pontos interiores e os pontos no exterior da curva, mas não pertencentes a L_{out} , são denominados pontos exteriores.

A rotina $check_{in}$ (Algoritmo 4) retira o ponto x da lista L_{out} e acrescenta-o à lista L_{in} .

Algoritmo 4 - $check_{in}$

Passo 1: retirar x de L_{out} e adicioná-lo a L_{in}

Atribua $\phi(x) = -1$ e calcule a função de velocidade $F(x)$

Passo 2: Para todo o $y \in N(x)$ tal que $\phi(y) = 3$

Adicione y a L_{out} , atribua $\phi(y) = 1$ e calcule a função de velocidade $F(y)$

A rotina (Algoritmo $check_{out}$) 5 retira o ponto x da lista L_{in} e acrescenta-o à lista L_{out} .

Algoritmo 5 - $check_{out}$

Passo 1: retirar x de L_{in} e adicioná-lo a L_{out}

Atribua $\phi(x) = 1$ e calcule a função de velocidade $F(x)$

Passo 2: Para todo o $y \in N(x)$ tal que $\phi(y) = -3$

Adicione y a L_{in} , atribua $\phi(y) = -1$ e calcule a função de velocidade $F(y)$

Neste algoritmo, somente é utilizado o sinal da função de velocidade, logo F_v é inteiro e toma os valores 1, 0 ou -1.

A rotina termina se alguma das seguintes condições é satisfeita:

1. A velocidade de todos os pontos da vizinhança satisfazem as condições:

$$F(x) \leq 0 \quad \forall x \in L_{out}$$

$$F(x) \geq 0 \quad \forall x \in L_{in}$$

2. O número de iterações máximo é alcançado

Outras hipóteses de critérios de paragem podem ser:

- controlo das diferenças entre contornos (Pluempitiwiriyaewej et al., 2005)
Se a diferença entre o contorno actual e o anterior for inferior a um certo limiar pré-definido, pára-se o algoritmo. Para medidas de comparação entre contornos, consulte a Secção 3.6.
- controlo do NSDE (*Normalized Step Difference Energy*) (Lee e Seo, 2006)
Calcula-se o NSDE, Equação 3.58, em todos os passos. Se o NSDE se tornar muito pequeno, pára-se o algoritmo.

$$NSDE = \frac{||\phi - \alpha||^2}{||\phi||^2} \quad (3.58)$$

$\alpha > 0,5$ é suficientemente grande para evitar oscilações.

O **Passo 3** corresponde a uma suavização gaussiana dos pontos do contorno e o parâmetro N_g é o tamanho do respectivo filtro e controla a eliminação de pequenos buracos no resultado

final. Se, por exemplo, se souber *a priori* que as regiões de interesse têm todas um raio maior ou igual a r , deve-se escolher $N_g = 2r$. N_a costuma ser maior do que N_g . Na segmentação de uma imagem de RM do cérebro, Shi e Karl (2005b) utilizam $N_g = 3$ e $N_a = 20$. No caso da segmentação de volumes, os mesmos autores utilizam $N_g = 3$ e $N_a = 50$.

No final deste procedimento, obtêm-se duas listas de pontos: L_{in} e L_{out} . O *Level Sets* 0 correspondente ao contorno é calculado através de interpolação linear (Rotina `contourc` do Matlab, versão 7.0.1).

Inicialização

Como já foi notado, não existe um consenso sobre a influência da inicialização das curvas na segmentação final. Neste trabalho foram tentadas quatro formas diferentes de proceder à inicialização:

1. Colocação de um ponto. Como, por exemplo, no caso do *Region Growing*;
2. Colocação de uma circunferência. Como, por exemplo, em (Tsai e Osher, 2004) (no exemplo do “The Chan-Vese algorithm”);
3. Colocação de várias pequenas circunferências. Como, por exemplo, em (Tsai e Osher, 2004) (no exemplo da “Segmentation of multiple phases”);
4. Linha divisória do espaço da imagem em duas partes. Como, por exemplo, em (Lee e Seo, 2006);

3.4.8 *Reaction Diffusion Active Contours*

Rendon et al. (2006) inspiraram-se nos modelos de contornos activos (Secção 3.4.7), no *Level Sets* (Secção 3.4.7), e num outro método proposto por Merriman et al. (1992). Neste último método, a propagação de interfaces é baseada na convolução com um *kernel gaussiano*, seguida de uma operação de Limiarização. Tal como no *Level Sets*, as mudanças topológicas são detectadas automaticamente. Como desvantagem, a resolução é limitada ao pixel.

Os *Reaction Diffusion Active Contours* consistem então numa combinação linear dum operador de média com um operador de máximo-mínimo. Este método possui algumas propriedades do *Level Sets*, uma vez que lida automaticamente com mudanças de topologia e não necessita que a inicialização da curva seja perto dos objectos a segmentar.

O operador média é um filtro que trabalha sobre 4 vizinhos e que tem como objectivo a suavização da função. O operador de máximo-mínimo é representado por det , onde $det = 1$ para se seleccionar o operador de máximo e $det = -1$ para se seleccionar o operador de mínimo. É esta operação que faz com que a curva se mova para dentro (caso $det = 1$) ou para fora (caso $det = -1$).

A superfície é inicializada como

$$\phi(x, y) \begin{cases} -1, & \text{se } (x, y) \text{ se encontram no interior da curva} \\ 1, & \text{caso contrário} \end{cases} \quad (3.59)$$

e o processo de evolução é feito em dois passos:

1. Difusão e erosão (ou dilatação) da curva usando a Equação 3.60 com $\alpha = 0,5$ e $det = 1$;
2. Afinamento da curva usando a Equação 3.60 com $\alpha = 1$ e $det = I(x, y)$ (erosão fora da curva e dilatação dentro);

A Equação 3.60 é:

$$\phi(x, y) = \frac{\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4}{4} + \alpha \left(\frac{det}{4} (|\gamma_1 - \gamma_2| + |\gamma_3 - \gamma_4|) + \frac{det}{4} |\gamma_1 + \gamma_2 - \gamma_3 - \gamma_4 + det|\gamma_1 - \gamma_2| - det|\gamma_3 - \gamma_4|| \right) \quad (3.60)$$

onde $\gamma_1 = I(x + 1, y)$, $\gamma_2 = I(x - 1, y)$, $\gamma_3 = I(x, y + 1)$ e $\gamma_4 = I(x, y - 1)$. Caso $det = 0$, esta equação reduz-se a um processo de difusão. Se $\alpha = 1$ e $det = 1$ ou $det = -1$, a equação reduz-se a um operador de máximo ou mínimo. Para outros valores, ambas as operações (difusão e erosão ou dilatação) são efectuadas.

Substituindo det por uma função contínua com valores entre -1 e 1, obtém-se um algoritmo de segmentação. A função det pode ser um detector de arestas, um detector de regiões, ou um detector de arestas e regiões. Em (Rendon et al., 2006) demonstra-se o método através de uma aplicação de detecção de arestas baseada nos níveis de cinzento presentes na imagem, Figura 3.19. Nesta Figura foi utilizado $\alpha = 0,5$ e $det(x, y) = (I(x, y) - 128)/128$.

Este método é extensível a 3D, utilizando uma vizinhança de 6 voxéis, em vez dos 4 pixéis (representados por $\gamma_1, \gamma_2, \gamma_3$ e γ_4). Rendon et al. (2006) provam ainda que o operador por eles proposto é estável, causal e que não gera ruído. Outra grande vantagem é a facilidade de implementação.

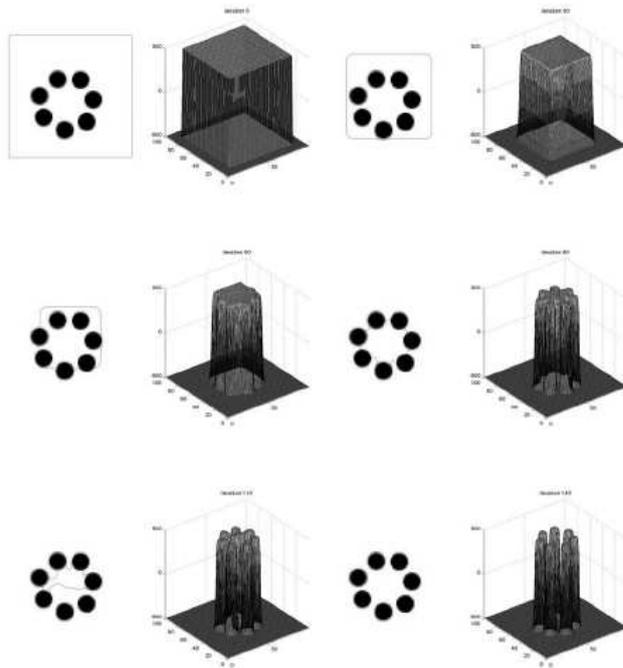


Figura 3.19: Detecção de arestas pelo método de *Reaction Diffusion*. Adaptado de (Rendon et al., 2006).

3.5 Segmentação automática

Ao avaliar uma técnica de segmentação é necessário ter em conta o contexto na qual foi utilizada. O grau de aplicação é directamente relacionado com as imagens às quais a técnica é aplicada. Que seja do nosso conhecimento não existe, até hoje, um método de segmentação totalmente automático e universal.

Na área do processamento de imagens de RM para utilização em linguística, dentro do conjunto dos métodos classificados como “automáticos”, destaca-se a utilização de um limiar fixo e global, determinado previamente; a detecção de mudanças bruscas na intensidade do sinal; e o método referido em (Narayanan et al., 2004). Neste último, é mencionado um método automático baseado em *Kalman Snakes* e *optical flow*, mas não são disponibilizados detalhes sobre a sua implementação nem volta a ser referido em nenhum dos trabalhos mais recentes. Um estudo bibliográfico revelou que este seria um método para sequências de imagens com variação temporal. Como o aproveitamento da resolução temporal não pertence ao âmbito principal deste trabalho, este método não foi explorado.

3.6 Métricas para comparação entre contornos

Os contornos gerados podem ser comparados através de diferentes métricas, entre as quais se destacam:

1. Índice de Pratt

O Índice de Pratt (IP), proposto em (Pratt, 1978), é dado pela expressão:

$$IP = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + \alpha d_i^2} \quad (3.61)$$

onde N é o número de pares de pontos correspondentes, d_i são as distâncias entre dois pontos correspondentes, e α está relacionado com o tamanho do contorno. Em (Silva, 2005) foi escolhido $\alpha = 1/9$, de forma a que $IP = 0,5$ quando todas as distâncias forem iguais a 3 pixels.

O IP varia no intervalo entre 0 e 1, em que 1 equivale à comparação de dois contornos iguais. Este índice corresponde a uma medida relativa do comportamento global das distâncias entre contornos.

2. Similaridade de áreas

A Similaridade de áreas (Pluempitiwiriyaewj et al., 2005), ou Coeficiente de Dice (Wu et al., 2005), é dada pela expressão:

$$S_{area} = \frac{2n(A_1 \wedge A_2)}{n(A_1) + n(A_2)} \quad (3.62)$$

onde \wedge é o operador “e” e $n(A)$ é o cardinal de A . $S_{area} > 0,7$ indica uma concordância excelente entre duas regiões.

3. Similaridade de formas

Embora a métrica do Ponto 2, seja um bom indicador de semelhança entre os tamanhos de duas regiões, não é um bom indicativo em relação aos detalhes de forma dos dois contornos. Em (Pluempitiwiriyaewj et al., 2005) é proposto o seguinte método:

(a) Criação de imagens binárias dos contornos: E_1 e E_2

$$D(x, y) = \begin{cases} -\min_{i,j \in C} \sqrt{(x-i)^2 + (y-j)^2}, & \text{se } (x, y) \in \Omega_1 \\ \min_{i,j \in C} \sqrt{(x-i)^2 + (y-j)^2}, & \text{se } (x, y) \in \Omega_2 \end{cases} \quad (3.63)$$

onde (x, y) representam os pixels da imagem; $(i, j) \in C$ representam os pixels no contorno C ; Ω_1 e Ω_2 são conjuntos de pixels dentro e fora do contorno C , respectivamente.

(b)

$$\eta_i(x, y) = \tan^{-1} \frac{\nabla_y D_i(x, y)}{\nabla_x D_i(x, y)} \quad (3.64)$$

para $i=1,2$

onde $\nabla_x D_i$ e $\nabla_y D_i$ representam os componentes x e y do gradiente do mapa de distâncias D_i , respectivamente.

(c)

$$S_{fase} = \frac{|\eta_1 - \eta_2 - \pi|}{\pi} \quad (3.65)$$

$S_{fase} \in [0, 1]$. Se $S_{fase} = 1$, os contornos têm a mesma fase, se $S_{fase} = 0$, existe uma diferença máxima de π na fase.

(d) Similaridade de formas (Pluempitiwiriyaewej et al., 2005)

$$S_{forma} = \frac{1}{n(C_2)} \sum_{(x,y) \in C_2} \Gamma_1(x,y) S_{fase}(x,y) \quad (3.66)$$

onde C_2 é o conjunto de pixeis no contorno automático, $n(C_2)$ é o cardinal de C_2 , e $\Gamma_1(x,y) \in [0, 1]$ e é derivado a partir de D_1

$$\Gamma_1(x,y) = e^{-\frac{D_1^2(x,y)}{\sigma^2}} \quad (3.67)$$

onde σ^2 é uma constante positiva.

As medidas de semelhança de forma originam normalmente valores menores do que as medidas de semelhança de áreas. Este facto explica-se devido à maior sensibilidade nas medidas de semelhança de forma a discrepâncias locais, do que nas medidas de áreas, que são mais globais (Pluempitiwiriyaewej et al., 2005).

Por questões de eficiência computacional, optou-se pelas métricas IP e S_{area} para comparar os contornos gerados pelos diferentes métodos de segmentação.

3.7 Segmentação em estudos de produção de fala

Na literatura, o estudo dos fonemas das várias línguas através de imagens de Ressonância Magnética tem vindo a ser feito através da medição de tamanhos (Alwan et al., 1997), distâncias, (Gick et al., 2000), volumes (Narayanan et al., 1995), formantes (Whalen et al., 1999), cálculo de funções de área (Takemoto et al., 2003), detecção de eventos durante a articulação, e da produção de modelos 2D (Perkell, 1996) ou 3D (Engwall, 2004), ou simples reconstruções 3D (Rua e Freitas, 2006; Rua, 2007) de várias estruturas, tais como a língua (Stone et al., 2000), ou o tracto vocal (Takemoto et al., 2004). Para que qualquer uma destas aproximações represente de forma fidedigna as imagens recolhidas, é necessário não só que seja aplicado um processo de segmentação, como também que ele seja o mais exacto e rápido possível, sem que se perca demasiado tempo neste procedimento (o que pode até inviabilizar o estudo).

Em relação ao tipo de resultados obtidos, notou-se uma certa confusão entre os conceitos de Modelação e Visualização. Por exemplo, em (Rua e Freitas, 2006) e (Mohammad, 1999), é referida a criação de um modelo 3D sendo que os resultados apresentados são apenas uma visualização 3D. A distinção prende-se no facto que enquanto na visualização se apresentam

os dados, ou reconstruções dos mesmos, na criação de modelos extraem-se parâmetros que podem depois ser manipulados de modo a simular diferentes situações.

A segmentação é então feita manualmente (Secção 3.3) ou recorrendo a métodos semi-automáticos. Neste último caso os mais utilizados são a Limiarização (Secções 3.4.1, 3.4.2 e 3.4.3), *Region Growing* (Secção 3.4.5) e métodos que utilizam contornos activos (Secção 3.4.7). Existem ainda alguns artigos que referem a segmentação automática seguida de uma verificação manual, uma segmentação interactiva (baseada na deformação de *templates*), segmentações baseadas em morfologia matemática, e detecção de mudanças bruscas na intensidade do sinal ao longo duma grelha.

Infelizmente, os esforços relacionados com a segmentação não têm vindo a ser feitos de forma evolutiva. Desde o primeiro estudo, em 1991, até ao presente, os métodos de segmentação utilizados dependem da equipa de investigação e seus recursos. A hipótese aqui defendida é que, se a equipa possui muitos recursos humanos, opta pela segmentação manual, continuando a ser considerada como a mais fidedigna. Este motivo leva a que se encontre pelo menos um artigo do corrente ano que utiliza este tipo de segmentação (Takano e Honda, 2007). Caso a equipa de trabalho seja mais interdisciplinar, são tentados métodos de segmentação mais automáticos. No entanto, houve só um esforço, em 1998, de se proceder a uma avaliação comparativa dos métodos de segmentação (Soquet et al., 1998). Por ser um estudo feito à quase dez anos atrás, os métodos testados resumem-se à segmentação manual, Limiarização, e *Snakes*. O presente estudo tenta colmatar esta falha, comparando métodos de segmentação mais actuais como o *Live Wire*, *Level Sets*, e *Reaction Diffusion*.

Apresenta-se de seguida uma classificação dos artigos relativos à produção de fala baseada nos métodos de segmentação utilizados e, dentro de cada classe, os artigos encontram-se ordenados por data. Sendo assim, a segmentação manual é utilizada em:

- Greenwood et al. (1992); Demolin et al. (1996); Mohammad (1999); Shadle et al. (1999); Yang (1999); Arnal et al. (2000); Badin et al. (2000); Engwall (2000); Stone et al. (2001); Soquet et al. (2002); Demolin et al. (2003); Avila-Garcia et al. (2004); Serrurier e Badin (2005a,b); Engwall et al. (2006); Takano e Honda (2007)

Dentro dos métodos semi-automáticos, há varias técnicas, listadas de seguida:

- Limiarização seguida de verificação manual
 - Baer et al. (1991); Narayanan et al. (1995, 1996); Alwan et al. (1997); Narayanan et al. (1997); Engwall et al. (2006)
- Limiarização com um limiar fixo
 - Crary et al. (1996); Badin et al. (1998a,b); Engwall e Badin (1999, 2000); Kröger et al. (2000); Engwall (2003a, 2004); Rua e Freitas (2006)

- Limiarização seguida de *Region Growing*
 - Story et al. (1996)
- *Region Growing*
 - Takemoto et al. (2003, 2004)
- Contornos Activos
 - Demolin et al. (1998); Bresch et al. (2006)
- Segmentação interactiva baseada em (Shechter et al., 1999). Esta segmentação baseia-se em volumes definidos *a priori*, que são deformados e interpolados no tempo.
 - Stone et al. (2000)
- Combinação de várias técnicas de morfologia matemática (*closing, opening, labelling, e watershed transform*). A inicialização depende de dois pontos que são utilizados para separar a cavidade nasal do tracto vocal.
 - Pritchard (2002)

Os métodos automáticos são também divididos em diversas classes:

- Limiarização com um nível ajustado adaptativamente à imagem a segmentar
 - Apostol et al. (1999, 2000)
- Detecção de mudanças bruscas na intensidade do sinal ao longo duma grelha pré-determinada
 - Whalen et al. (1999); Demolin et al. (2000); Gick et al. (2000, 2002)
- Combinação de *Kalman Snakes* e *optical flow*
 - Narayanan et al. (2004)

Vários são os softwares utilizados. Desde o software de visualização VRML (Mohammad, 1999; Shadle et al., 1999; Kröger et al., 2000), passando pelo *VIDATM* (Volumetric Image Display and Analysis) (Story et al., 1996) e pelo NIH-Image (Badin et al., 1998a,b; Whalen et al., 1999; Apostol et al., 1999, 2000; Badin et al., 2000; Gick et al., 2000, 2002; Rua e Freitas, 2006) e até por softwares criados especialmente para este tipo de tratamento de imagens como o *Bézier Curve Editor* (Engwall, 2000). O VRML possui a desvantagem de permitir somente a visualização dos dados, não sendo possível efectuar o processamento dos mesmos. As referências ao *VIDATM* consistem no trabalho de Hoffman e seus colegas (Hoffman et al., 1992) e na página on-line <http://everest.radiology.uiowa.edu>, no entanto, nenhuma destas fontes estava disponível aquando da realização deste trabalho. O NIH-Image é de domínio publico e permite a obtenção de pontos de articulação, medição de distâncias, segmentação por *threshod* ajustado à imagem, obtenção de contornos, reconstrução de imagens, etc. No

entanto, a manipulação dos contornos (de forma a, por exemplo, obter funções de área de forma automática) não é linear. O *Bézier Curve Editor* é um software muito específico que serve somente para o ajuste de curvas de Bézier à estrutura que se deseja segmentar. Neste trabalho, devido à sua versatilidade, optou-se pela utilização do Matlab. Esta ferramenta foi também utilizada nos estudos: (Apostol et al., 1999; Engwall e Badin, 1999, 2000; Apostol et al., 2000; Pritchard, 2002; Demolin et al., 2003; Engwall, 2003a, 2004; Avila-Garcia et al., 2004; Serrurier e Badin, 2005b).

3.8 Comparação das classificações adoptadas com outras presentes na literatura

No presente trabalho, os métodos de segmentação foram classificados em manual, semi-automático e automático. No entanto, esta não é a divisão mais comum. Na literatura, os métodos são divididos em “métodos orientados a contornos” e “métodos orientados a regiões” (Shah, 2007; Li e Yezzi, 2007). A diferença prende-se com a sua filosofia de base. Enquanto que os primeiros tentam determinar possíveis contornos, os segundos tentam dividir a imagem em regiões. Dos métodos sobre os quais este trabalho se debruça, o *Live Wire* pertence à classe dos métodos baseados em limites e *Region Growing*, *Level Set* e *Reaction Diffusion* são orientados a regiões.

Tanto os métodos de atenuação de ruído como os de segmentação foram apresentados separadamente. No entanto, existem alguns pontos de ligação entre alguns dos métodos apresentados. Em especial entre o Filtro de Difusão Anisotrópica e o *Level Sets*, e entre o *Live Wire* e as *Snakes*. Estas semelhanças serão descritas de seguida.

Comparação entre o Filtro de Difusão Anisotrópica e o *Level Sets*

A Equação de difusão 3.3 pode ser geometricamente interpretada como: os iso-contornos da imagem movem-se com velocidade d . Através deste movimento dos iso-contornos, normais a eles próprios, a suavização é feita de forma perpendicular às arestas, e por isso não alterando a sua definição (Malladi e Sethian, 1995). Esta noção é facilmente comparável com a noção de *Level Sets*, tal como pode ser confirmado pela semelhança entre as equações de difusão 3.4 e 3.5 e as equações 3.44 e 3.45. Ambos os métodos são ainda baseados em equações às derivadas parciais.

Comparação entre *Live Wire* e *Snakes*

A técnica *Live Wire* pode ser vista como pertencente à classe dos contornos activos (*Snakes*) e foi mesmo incorporada nas *united Snakes* (Chodorowski et al., 2005). No entanto, uns anos antes, Mortensen e Barrett (1995; 1997) apontavam algumas similaridades e diferenças entre *Live Wire* e *Snakes*, que se passam a descrever. Entre as similaridades encontram-se:

1. a magnitude do gradiente da função de custo no *Live Wire* é semelhante ao funcional de energia de aresta usado nas *Snakes*;
2. ambos os métodos empregam um termo de suavização de forma a minimizar os efeitos do ruído nos contornos;
3. os contornos do *Live Wire* e *Snakes* são atraídos para arestas fortes;
4. ambas as técnicas tentam descobrir contornos globalmente óptimos;
5. tanto o *Live Wire* como as *Snakes* necessitam de interacção do utilizador.

Como diferenças pode-se apontar:

1. a função de custo binária do Laplaciano não tem vindo a ser usada nos contornos activos;
2. o termo de suavização nos contornos activos é interno (baseado na posição dos pontos do contorno e na sua geometria) enquanto que no *Live Wire* o termo de suavização é calculado a partir das direcções externas do gradiente da imagem;
3. as *Snakes* são tipicamente atraídas para arestas somente na vizinhança de um “vale” na energia do gradiente, enquanto que o *Live Wire* pode ser atraído para arestas fortes a partir de qualquer distância;
4. as *Snakes* são globalmente óptimas ao longo de todo o contorno, enquanto que os contornos produzidos com o *Live Wire* são óptimos por zonas (ótimo entre pontos somente).
5. finalmente, as *Snakes* “refinam” um contorno inicial, enquanto que o *Live Wire* selecciona de forma interactiva um segmento óptimo do contorno a partir de caminhos de custo mínimo entre os caminhos possíveis.

Em (Mortensen e Barrett, 1995; Barrett e Mortensen, 1997) é afirmado que a característica dos *Live Wire* de serem óptimos por zonas cria uma harmonia entre a optimização global e o controlo local. Para além disso, permite o *boundary cooling* e treino *on-the-fly* intra-objecto.

Em “United Snakes”, Liang et al. (2006) defendem que o *Live Wire* é complementar às *Snakes*. O método consiste em aumentar a eficiência da difícil inicialização das *Snakes*, tornando-a num processo interactivo onde o utilizador pode construir os contornos iniciais de forma rápida e com pouco esforço, tirando partido do *Live Wire*. Estes autores afirmam ainda que as diversas técnicas desenvolvidas para o cálculo das energias nas *Snakes* podem ser usadas para gerar mapas de custo locais no *Live Wire*. Inversamente, os mapas de custo locais do *Live Wire* podem ser utilizados como energias nas *Snakes*. As *Snakes* podem também usufruir do treino (ou seja, do mapa de custo “treinado”) do *Live Wire*. O *Live Wire*, por ser baseado em imagem e não em modelos, como as *Snakes*, poderia usufruir destas impondo suavidade nos contornos gerados, e preenchendo possíveis falhas nos limites do objecto. A combinação do *Live Wire* com as *Snakes* oferece ainda um novo mecanismo para o tratamento de grandes

imagens, utilizando o *Live Wire* em réplicas de baixa resolução. Para mais detalhes destes e outros tipos de semelhanças entre *Live Wire* e *Snakes*, consulte (Liang et al., 2006).

3.9 Considerações finais

O presente Capítulo teve como objectivo contextualizar e fundamentar a tarefa principal desta investigação: o processamento de imagens de RM. Apresentaram-se de forma pormenorizada diferentes técnicas de segmentação de imagens, tendo sido este o método utilizado para o processamento das imagens neste estudo, como se verá no Capítulo seguinte.

Começou-se por descrever a fase prévia à segmentação (pré-processamento). Seguiu-se com a descrição de diferentes métodos de segmentação, adoptando para estes a classificação de manual, semi-automática e automática. As três classes foram apresentadas de forma comparada analisando para cada uma as respectivas vantagens e limitações. Na Secção seguinte foram também abordadas as métricas para comparação entre contornos. O Capítulo conclui-se com uma revisão da utilização da segmentação em estudos da produção da fala. Tal revisão levou à observação que, apesar de a segmentação ser amplamente utilizada nesta área, a sua aplicação limita-se às técnicas manuais ou semi-automáticas mais básicas (como a Limiarização).

Esta última parte do Capítulo introduz o âmbito específico deste estudo. O seguinte Capítulo servirá para apresentar o trabalho experimental realizado onde se aplicam, pela primeira vez, técnicas mais avançadas de segmentação às imagens de RM.

Capítulo 4

Trabalho experimental

O processo de segmentação de imagem, tal como foi visto no Capítulo anterior, pode ser realizado através da aplicação de um variado leque de técnicas. Com este Capítulo pretendem-se fundamentar as opções metodológicas adoptadas para o presente estudo, descrevendo-se o protocolo seguido para a implementação de cada uma das técnicas escolhidas e, ainda antes, o processo de aquisição de dados.

Escolheu-se estruturar o capítulo de acordo com a sequência temporal das tarefas executadas. Começa-se por apresentar o *corpus* escolhido para a análise (Secção 4.1.1). A seguir, descreve-se o processo de aquisição de imagens, começando pela caracterização do instrumento seleccionado para a recolha de dados (Secção 4.1.2). Apresentar-se-à então o conjunto de dados adquiridos: imagens 2D, imagens em Tempo Real e aquisições volumétricas.

A seguir, na Secção 4.2.1, serão ilustradas as modalidades de pré-processamento das imagens adquiridas, isto é, das técnicas de melhoramento qualitativo das imagens a segmentar; e, finalmente, na Secção 4.2.2, passar-se-á a ilustrar os métodos de segmentação utilizados para as imagens 2D, imagens de Tempo Real, e volumes.

4.1 Base de Dados de Ressonância Magnética para o Português Europeu

Esta Secção está dividida em duas partes, uma primeira onde se explicita os três tipos de *Corpus* adquiridos e uma segunda onde se caracteriza o tipo de aquisição efectuada para cada um dos *Corpus* descritos.

4.1.1 *Corpus* de Ressonância Magnética

Tendo este trabalho o objectivo final de executar a segmentação de imagens do tracto vocal no acto da fala, tornou-se inevitável pôr entre as primeiras tarefas a selecção dum conjunto de fones (estáticos) e palavras (dinâmicas) a analisar, isto é, um *corpus*. Define-se assim como *corpus* um conjunto de sons da fala (Ribeiro, 2003).

Escolheu-se um *corpus* para cada uma das modalidades de aquisição de imagens previamente definida: 2D, Tempo Real, e 3D. Os *Corpora* adquiridos serão apresentados de seguida.

Corpus 2D

A Tabela 4.1 resume o *Corpus* 2D.

Tabela 4.1: *Corpus* 2D

Fonema			Observações
Vogais	Orais	[a] [ø] [E] [O] [i] [e] [6̃] [o] [u]	isolada; 1 repetição
	Nasais	[ĩ] [ẽ] [6̃] [õ] [ũ]	isolada; 1 repetição
Consoantes	Nasais	[m] [n] [ɲ]	isolada; 1 repetição
	Fricativas	[f] [v]	no início de palavra; 1 repetição
		[s] [ʃ]	em contexto VCV, V=[i u a];
		[z] [ʒ]	3 repetições
	Oclusivas	[p] [t] [k]	em contexto VCV, V=[i u a];
Oclusivas	[b] [d] [g]	3 repetições	
Laterais	[l] [ʎ]	isolada; 1 repetição	

Corpus Tempo Real

As palavras deste *corpus* foram escolhidas com base no *corpus EMMA* (Teixeira e Vaz, 2001). Elas correspondem a pares de palavras diferindo apenas na vogal. Foram ainda incluídas algumas palavras ([karu]/[kaRu]) que permitem o estudo das vibrantes (dado que estas não são passíveis de ser sustidas, não podem ser estudadas através de aquisições 2D ou 3D estáticas). A Tabela 4.2 apresenta este *Corpus* de Tempo Real, adquirido no plano sagital.

Tabela 4.2: *Corpus* Tempo Real

Sequências de vogais e pares de palavras	Transcrição fonética
Sequência /aeiou/	-
Sequência /6̃ ẽ ĩ õ ũ/	-
manto/mato	[m6̃tu] / [matu]
minto/mito	[mĩtu] / [mitu]
canto/cato	[K6̃tu] / [katu]
carro/caro	[kaRu] / [karu]
quatro/coalho	[kuatru] / [kuaLo]

Corpus 3D

O *Corpus 3D*, adquirido em volume, está resumido na Tabela 4.3.

Tabela 4.3: *Corpus 3D*

Fonema			Observações
Vogais	Orais	[a] [u] [O]	isolada; 1 repetição
		[i] [e] [6] [o] [u]	isolada; 1 repetição
	Nasais	[i~] [e~] [6~] [o~] [u~]	isolada; 1 repetição
Consoantes	Nasais	[m] [n] [J]	isolada; 1 repetição
	Fricativas	[f]	em contexto VCV, V=[i u a]; 3 repetições
	Fricativas	[s] [S]	
	Laterais	[l] [L]	isolada; 1 repetição

4.1.2 Aquisição das imagens

Após a análise das vantagens e desvantagens que o conjunto de métodos de aquisição de dados apresenta (ver Capítulo 2, Secção 2.3), optou-se pela utilização da RM para adquirir os dados necessários, isto é, as imagens e volumes do tracto vocal.

Esta Secção apresenta o processo de aquisição de imagens, e as opções técnicas efectuadas para tal fim.

Instrumento seleccionado para a recolha de dados

As imagens e os volumes utilizados para esta investigação, foram adquiridos utilizando um equipamento de 1,5 Tesla (Magnetom Symphony, Maestro Class, Siemens, Erlanger, Alemanha, Figura 4.1) equipado com gradientes Quantum (máxima amplitude = 30 mT/m; *rise time* = 240 μ s; *Slew rate* = 125 T/m/s; FOV = 50 cm). Foram empregues antenas de crânio e de pescoço, em simultâneo, em todas as aquisições.

Imagens 2D

As imagens 2D (alguns exemplos foram já apresentados nas Figuras 2.2, 2.3, 2.4, 2.5, e 2.6) foram adquiridas utilizando uma sequência TSE (ver Secção 2.3.1) com ponderação em T_1 com $TA = 5,6$ segundos, $ETL = 15$, $TR = 400$, $TE = 8,3$, $FA = 180^\circ$, $FOV = 200$ mm e $ST = 5$ mm. As aquisições foram executadas no plano sagital, possuindo, as imagens obtidas, um tamanho de 256x256, com um espaçamento entre pixels de 0,78 mm em ambas as direcções.

Conjuntos de imagens de Tempo Real

Para as aquisições em Tempo Real, utilizou-se uma sequência TrueFISP (fortemente ponderada em T_2) com $TA = 0,2$ segundos/frame, $ETL = 1$, $TR = 3,56$ ms, $TE = 1,5$ ms, $FA = 60^\circ$, $FOV = 240$ mm e $ST = 5$ mm. As aquisições foram feitas no plano sagital,



Figura 4.1: Siemens Magnetom Symphony.

possuindo, cada *frame* obtido, um tamanho de 256×256 , com um espaçamento entre pixels de 0,94 em ambas as direções. A aquisição permite obter uma resolução temporal de 5 *frames* por segundo. Um exemplo de alguns *frames* retirados de uma das sequências adquiridas é apresentado na Figura 4.2.

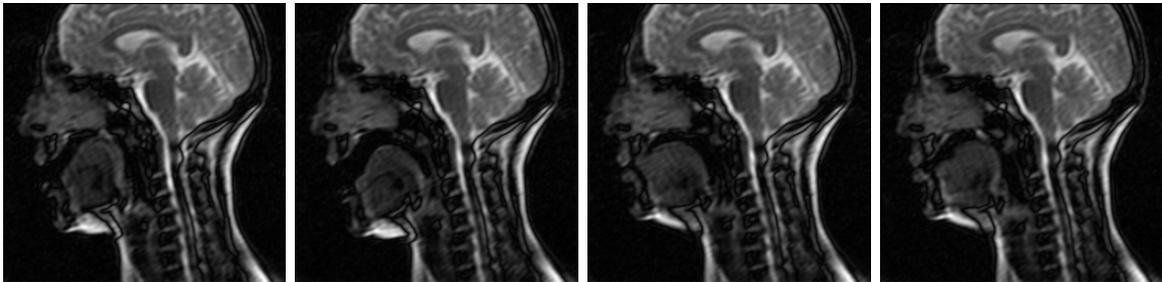


Figura 4.2: Frames [k], [a], [t] e [u] na produção da palavra *catu* [katu].

Estas imagens não foram exploradas como sendo uma sequência dinâmica, mas sim de forma independente. Serviram assim para testar os métodos de segmentação em imagens com qualidade inferior às imagens 2D adquiridas com a sequência TSE.

Aquisições volumétricas

Os volumes aqui estudados foram adquiridos com uma sequência de Eco de gradiente rápido FLASH 3D VIBE no plano axial com ponderação em T_1 , $TA = 18$ segundos, $ETL = 1$, $TE = 2,44$ ms, $FA = 10^\circ$, $FOV = 270$ mm e $ST = 2$ mm. Cada volume obtido possui um tamanho de $512 \times 416 \times 60$ pixels, sendo que cada voxel mede 0,53 mm em largura e altura e 2 mm em profundidade. A título exemplificativo, na Figura 4.3, apresentam-se alguns cortes de um dos volumes adquiridos.

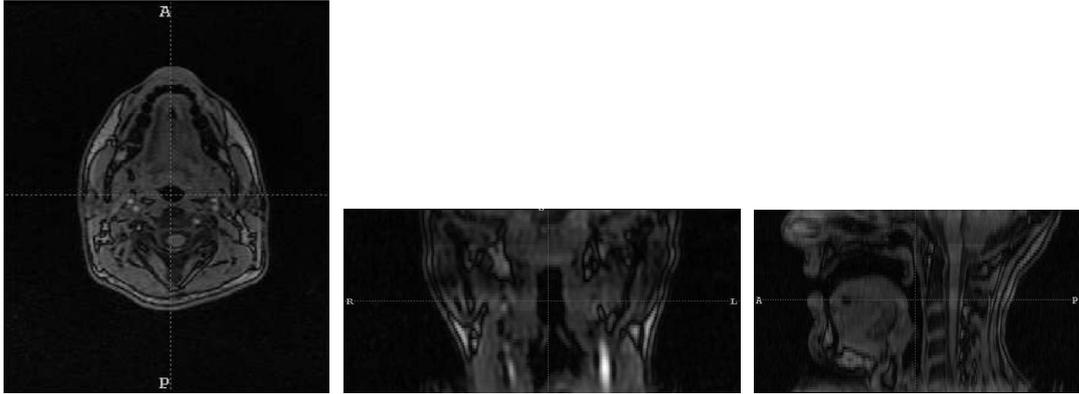


Figura 4.3: Visualização de três planos (axial, coronal e sagital) para o volume adquirido para o som [a].

4.2 Métodos de processamento de imagem escolhidos

Nesta segunda parte do Capítulo foca-se a atenção nos métodos de processamento de imagem escolhidos. Começa-se por referir ao pré-processamento: remoção do ar exterior, atenuação de ruído, interpolação, e sub-amostragem. A Secção continua com a listagem dos parâmetros utilizados nos métodos de segmentação escolhidos: *Region Growing*, *Level Sets*, *Reaction Diffusion*, e *Live Wire*.

4.2.1 Pré-Processamento

Antes de se proceder à segmentação, é necessário efectuar algum melhoramento às imagens a segmentar. Esta manipulação depende da tipologia da informação com que se está a trabalhar (2D, Tempo Real, ou 3D). Os pré-processamentos utilizados para cada tipo de informação são descritos de seguida. Antes de se proceder, convém referir que os níveis de cinzento de todas as imagens e volumes foram normalizados para pertencerem ao intervalo [0,1].

Remoção do ar exterior em imagens 2D e Tempo Real

O tracto vocal e o meio envolvente, no qual o informante se encontra, são preenchidos pela mesma substância gasosa: o ar. Nas imagens de RM, e por esta razão, o tracto vocal e o ar externo, apresentam níveis de cinzentos semelhantes. Tornou-se por isso necessária a implementação de um método que delimitasse o tracto vocal. Adoptou-se para tal fim pelo método descrito em (Behrends et al., 2003), e já relatado no presente trabalho na Secção 3.2.1.

Atenuação do ruído em imagens de Tempo Real

Sobretudo no que diz respeito às aquisições de imagem de Tempo Real, a baixa resolução das imagens (ruído) representou um dos principais problemas. Com o objectivo de minimizar

os efeitos do ruído, optou-se pela implementação de alguns métodos diferentes: filtro de mediana, filtro de Gaussiana, Difusão Anisotrópica, e *Wavelets*.

A execução do filtro de mediana foi feita com a função `medfilt2` da *toolbox* de processamento de imagem do Matlab, da versão 7.0.1 (Lim, 1990). O tamanho da vizinhança onde foi calculada a mediana é 3x3.

Com a Difusão Anisotrópica fizeram-se dois tipos de experiências:

- $d(x, y, t) = 1$ com um número de iterações igual a 1 (filtro de Gaussiana);
- função de difusão 3.4 com uma constante $K = 2.0$, e um número de iterações igual a 3;

O passo de iteração, dado que a vizinhança era de 4 pixels (Secção 3.2.2), foi 1/5.

Tal como em (Silva, 2005), para a suavização por *Wavelets*, foi escolhida como função de base a função Haar, tendo-se efectuado uma decomposição até ao nível dois. Esta estratégia foi implementada através da função `swt2` (transformada *Wavelet* 2D discreta e estacionária) existente no Matlab versão 7.0.1 (Coifman e Donoho, 1995; Nason e Silverman, 1995; Pesquet et al., 1996).

Interpolação dos volumes

Os volumes adquiridos não possuem um voxel isotrópico, ou seja, as dimensões do voxel não são iguais em todas as direcções. Como já foi visto, na Secção 4.1.2, cada volume possui 512x416x60 voxels, correspondendo cada voxel a 0,5x0,5x2 mm. Devido ao impacto positivo que um voxel isotrópico provoca na visualização, decidiu-se efectuar interpolação linear na terceira dimensão, de modo a que todos os voxels de “trabalho” possuíssem uma resolução de 0,5x0,5x0,5 mm. Tenta-se desta forma aproveitar a informação relevante dos pixels observados para inferir os dados intermédios (Albornoz, 2007). Convém no entanto referir que não se pode gerar informação que não foi adquirida. Este método de transformação dos voxels em voxels isotrópicos tem, portanto, um erro associado. Normalmente, esse erro corresponde ao “esborratamento” da imagem (Tsai et al., 2001). Uma das tarefas futuras é, assim, estudar métodos de interpolação mais fidedignos e apropriados para a posterior segmentação dos dados (onde a presença de arestas bem definidas é fundamental).

Sub-amostragem dos volumes

Por questões de eficácia computacional e para prova de conceitos optou-se, por vezes, trabalhar com volumes sub-amostrados. Os volumes sub-amostrados possuem uma resolução de 2x2x2 mm, onde cada voxel do novo volume nas duas primeiras dimensões é a média, na respectiva dimensão, de 4 voxels do volume inicial. A terceira dimensão, como já possuía a resolução pretendida, foi deixada intocada.

4.2.2 Segmentação de Imagens e Volumes

A segmentação de imagens e volumes adquiridos através de RM constitui o principal foco de atenção deste trabalho. Nesta Secção explicitam-se os métodos utilizados para este fim e respectivos parâmetros aplicados. Começar-se-á por apresentar os métodos de segmentação aplicados às imagens 2D, continuando com o tratamento das imagens de Tempo Real e, por fim, dos volumes.

É importante realçar que não será utilizada a informação temporal das sequências de imagens adquiridas em Tempo Real, aplicando-se os mesmos métodos utilizados para a segmentação das imagens bidimensionais. Tal estudo servirá de experiência no tratamento de imagens 2D de menor qualidade, como são as imagens RM de Tempo Real (a Secção 2.5 apresenta um estudo da qualidade dos tipos de informação aqui trabalhados).

Segmentação de imagens 2D

Para executar a segmentação das imagens adquiridas em 2D, foram utilizados os métodos *Region Growing*, *Level Sets* e *Reaction Diffusion*. As especificações relativas à aplicação de cada um destes métodos neste estudo serão de seguida explicitadas.

Region Growing

A escolha das sementes utilizadas neste método é feita de forma manual, sendo que têm que pertencer ao objecto de interesse (tracto vocal). Cada segmento do tracto vocal terá que ser inicializado individualmente. Ou seja, se o tracto vocal se apresenta como um elemento conexo, é somente necessário colocar uma semente. Caso ele se encontre dividido, é necessário colocar o número de sementes correspondentes ao número de partes em que ele se encontra dividido. Tendo-se o cuidado de todos os segmentos possuírem pelo menos uma semente. O limiar empiricamente escolhido, e validado pelos histogramas da Figura 4.4, para a segmentação das imagens 2D é 0,1. Isto significa que se a diferença entre a média das intensidades dos pixels já classificados como pertencentes à região e o pixel que se está a testar for menor do que 0,1, esse pixel é incluído na região, caso contrário, é excluído.

Level Sets

Para o *Level Sets* foram tentadas quatro inicializações diferentes. Na primeira, e tal como no caso do *Region Growing*, escolheu-se um ponto no interior do tracto vocal; na segunda uma linha horizontal a passar pelo centro da imagem; na terceira um círculo centrado no centro da imagem (ponto (128,128)) e de raio 85; finalmente, na última inicialização foi testado um conjunto de 100 círculos de raio 10 igualmente espaçados por toda a imagem. No Capítulo dos resultados, as Figuras 5.8, 5.9, 5.10 e 5.11 ilustram as inicializações aqui descritas.

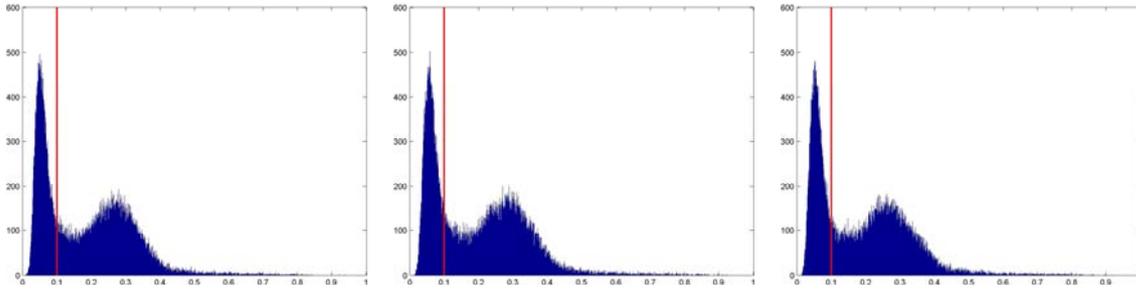


Figura 4.4: Histogramas das imagens RM 2D das vogais extremas: [i], [u] e [a].

A primeira função de velocidade implementada foi uma função do tipo 3.32 onde ϵ e T foram experimentalmente (através de uma análise visual das imagens e de um processo de “tentativa/erro”) determinados como sendo $\epsilon = 0,09$ e $T = 0,2997$. A segunda função de velocidade, na sua formulação original, inclui a curvatura e é dada pela Equação 3.40. Neste trabalho substituiu-se a curvatura $k = \text{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right)$ pelo Laplaciano $(\nabla^2 \phi)$ (Tai et al., 2007). O valor de T foi empiricamente escolhido como 0,04. Esta nova formulação da Equação 3.40 foi combinada linearmente com a primeira função de velocidade implementada. A terceira função de velocidade estudada foi a proposta por Baillard et al. (2000), e já apresentada na Equação 3.52.

A primeira função de velocidade não depende da superfície a evoluir. Pode por isso ser calculada uma única vez no início da rotina. As duas funções de velocidade restantes têm que ser actualizadas a cada iteração, tornando assim a evolução mais lenta.

A condição de paragem é composta por três factores: (1) número de iterações superior a 200, (2) é verificada a condição 1 (página 56), (3) o IP (Equação 3.61) da comparação do contorno actual com o anterior é superior a 0,9998 e a Similaridade de Áreas (Equação 3.62) entre o contorno actual e o anterior é menor do que 1,0005. A primeira destas três condições a ser satisfeita leva ao fim da evolução.

Um problema identificado nestas segmentações foi a sobre-segmentação, Figura 5.12. Ou seja, é segmentado não só o tracto vocal, como outras estruturas presentes na imagem. Este problema já tinha sido identificado em (Cates et al., 2004) e (Silva, 2005). O primeiro sugere que se aumente o peso da curvatura na função de velocidade. O segundo propõe rejeitar as regiões cujo volume seja inferior ao volume de uma esfera com um diâmetro fixo.

O método de Cates et al. (2004) foi experimentado, mas não leva a resultados satisfatórios. Neste trabalho, seguiu-se uma filosofia semelhante à de Silva (2005). Para restringir a área a segmentar à zona do tracto vocal, começou-se por multiplicar a superfície ϕ pela máscara representada na Figura 4.5. Esta máscara foi determinada empiricamente. De seguida recorreu-se às funções do matlab `bwlabel` e `regionprops`. A função `bwlabel` divide

a imagem nos seus componentes conexos (Haralick e Shapiro, 1992). É um processamento morfológico não trivial em que se determinam alguns descritores de regiões conexas na imagem. Com base nesses descritores, procede-se à classificação das áreas de interesse. A função `regionprops` disponibiliza algumas propriedades das regiões numa imagem, entre elas, a área de cada região. Foram assim seleccionadas as três regiões de maior área. A segunda região seria eliminada caso a sua área fosse inferior a 1670 pixels. A terceira região teria que ter uma área superior a 900 pixels (sendo que só é testada caso se tenha optado pela inclusão da segunda área).



Figura 4.5: Máscara pela qual se multiplica a superfície ϕ .

A Figura 4.6 exemplifica o método descrito no parágrafo anterior. São apresentadas três segmentações onde o que se apresenta a preto foi classificado como não pertencente ao tracto vocal e a parte branca foi classificada como pertencente ao tracto vocal. A combinação das rotinas `bwlabel` e `regionprops` permitiu identificar as três regiões conexas de maior área (contornadas a vermelho, verde e amarelo). As suas áreas são, respectivamente, 6554, 616 e 383 para o fone [a]; 3548, 2545 e 542 para o fone [o]; e 2328, 1753 e 1244 para o fone [l]. A leitura do parágrafo anterior leva a concluir que para o fone [a] é escolhida somente a região contornada a vermelho, no caso do fone [o] considera-se que ele é um objecto não conexo com duas partes distintas: a contornada a vermelho e a contornada a verde; e no do do [l] as três regiões contornadas são aceites como pertencentes ao tracto vocal.

Reaction Diffusion

O comportamento deste método de segmentação depende essencialmente de uma constante α . Caso $\alpha = 0$, a curva possui uma grande rigidez, no outro extremo, $\alpha = 1$, a rigidez é reduzida. Valores intermédios de α levam a valores intermédios de rigidez. Este método é ainda bastante robusto à escolha da inicialização.

Escolheu-se então como inicialização um círculo centrado no centro da imagem (ponto (128,128)) e com raio $r = 85$. A função $det(x,y)$ implementada, é idêntica à função de velocidade 3.40 adaptada e já utilizada no *Level Sets*.

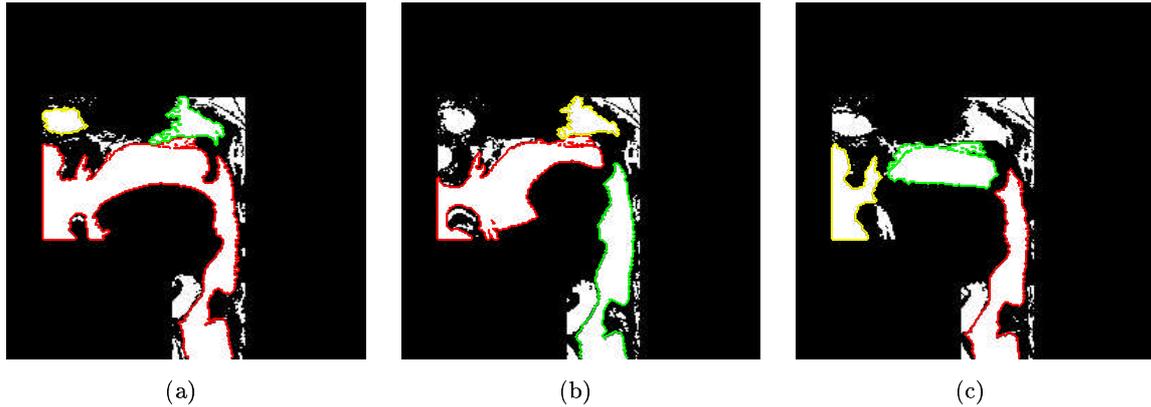


Figura 4.6: Eliminação da sobre-segmentação: (a) resultados relativos ao fone [a], (b) resultados relativos ao fone [o], e (c) resultados relativos ao fone [l]. Todas as segmentações aqui representadas foram obtidas com *Level Sets* 2D que classificaram as zonas representadas a preto como não pertencentes ao tracto vocal e as zonas a branco como interiores ao tracto vocal. A maior região conexa apresenta-se contornada a vermelho, a segunda maior região conexa apresenta-se contornada a verde e a terceira maior região conexa apresenta-se contornada a amarelo.

Os critérios de paragem utilizados são: (1) número de iterações superior a 300, (2) o IP (Equação 3.61) da comparação do contorno actual com o anterior é superior ou igual a 0,9990 e a Similaridade de Áreas (Equação 3.62) entre o contorno actual e o anterior é menor ou igual a 1,6898. A primeira destas duas condições a ser satisfeita leva ao fim da evolução.

O pós-processamento é o descrito para o *Level Sets*.

Segmentação de imagens de Tempo Real

A primeira segmentação das imagens de Tempo Real foi feita de forma manual por uma Técnica de Radiologia, investigadora do projecto HERON, Paula Martins. Para este efeito foi utilizado o editor de imagem do Windows, Paint.

Os contornos obtidos com a segmentação manual, devido ao grande controlo do utilizador, foram considerados como a “ground truth” (Mendonça et al., 2004) e serviram por isso como termo de comparação para os outros métodos testados: *Region Growing*, *Level Sets*, e *Reaction Diffusion*.

Com a excepção da segmentação manual, as imagens de Tempo Real foram segmentadas não só no seu estado original como também após a aplicação dos métodos de atenuação de ruído descritos na Secção 4.2.1.

Na implementação do método de *Region Growing*, o limiar empiricamente escolhido coincide com o caso 2D e é igual a 0,10.

Para o *Level Sets*, a função de velocidade escolhida foi a que obteve os melhores resultados nas imagens 2D e corresponde à Função 3.40 adaptada. Também os parâmetros aplicados foram os mesmos.

No *Reaction Diffusion*, a função $det(x, y)$ escolhida é a mesma que para o caso das imagens 2D. Os seus parâmetros são também os mesmos.

Devido à sobre-segmentação verificada no *Level Sets* e no *Reaction Diffusion* foi necessária uma fase de pós processamento. Tal como o efectuado no caso 2D, multiplicou-se a superfície por uma máscara. Devido às diferentes características deste tracto vocal em relação ao analisado nas imagens 2D (o informante das imagens 2D é do sexo masculino e o informante das imagens de Tempo Real é do sexo feminino), foi adoptada uma máscara rectangular com os limites [25 135 71 144]. Foram depois seleccionadas as três regiões de maior área: a primeira região é sempre incluída, enquanto que a segunda região é eliminada caso a sua área seja inferior a 255 pixels, e a terceira região tem que ter uma área superior a 407 pixels (sendo que só foi testada caso se tenha optado pela inclusão da segunda área).

Segmentação de volumes

Após se ter efectuado a segmentação de imagens de qualidades variáveis, o passo natural seguinte é a segmentação de volumes. Nesta Secção, são explicitados os métodos de segmentação aplicados aos volumes de RM adquiridos. É importante notar que foram utilizadas duas filosofias opostas para a obtenção de segmentações 3D. A primeira não se trata de uma segmentação 3D, mas sim de um “empilhamento” de segmentações 2D. A segunda é uma verdadeira segmentação 3D, ou seja, utiliza toda a informação disponível no volume a segmentar.

A primeira filosofia utiliza *Live Wire* 2D e *Level Sets* 2D e é aqui denominada por Segmentação $2\frac{1}{2}$ D.

Devido a limitações temporais, para o caso 3D, não foram tentadas todas as técnicas descritas para o caso 2D. O método de segmentação 3D foi escolhido tendo em conta os resultados 2D. Eliminou-se o *Region Growing* por se prever que a sua implementação levasse a um tempo de execução demasiado longo (recorde-se que não se fez neste trabalho nenhuma tentativa de optimização das rotinas criadas). O *Reaction Diffusion* foi excluído por ter produzido, em 2D, resultados inferiores aos obtidos com o *Level Sets*. Escolheu-se assim este último como implementação duma segmentação verdadeiramente 3D.

Segmentação $2\frac{1}{2}D$

Nesta Secção descreve-se um método de segmentação que, embora seja aplicado aos volumes de RM, não pode ser considerado uma verdadeira segmentação em volume. O método é baseado na segmentação “corte a corte” e é por isso denominado por segmentação $2\frac{1}{2}D$. A grande desvantagem deste tipo de segmentações “corte a corte” é, tal como apontado por El-Zehiry et al. (2005), o efeito *interslice*. Ou seja, as segmentações produzidas não são suaves, mas apresentam mudanças abruptas entre cada corte utilizado.

O primeiro passo efectuado foi a interpolação do volume, já descrita na Secção 4.2.1. De seguida, é seleccionada a imagem médio-sagital do volume, tendo sido segmentado o tracto vocal (através do *Live Wire*) - a amarelo na Figura 4.7. Ao contorno gerado é sobreposta uma grelha fixa - a verde na Figura 4.7(a). Calcula-se os pontos de intersecção entre a grelha fixa e o contorno, e é obtido o seu valor médio. Esta é uma primeira aproximação para a linha central - a azul na Figura 4.7(a). A linha central é sobre-amostrada e suavizada - a azul na Figura 4.7(b). A partir desta linha central, são calculados cerca de 45 planos perpendiculares à linha central, representados a vermelho na Figura 4.7(c).

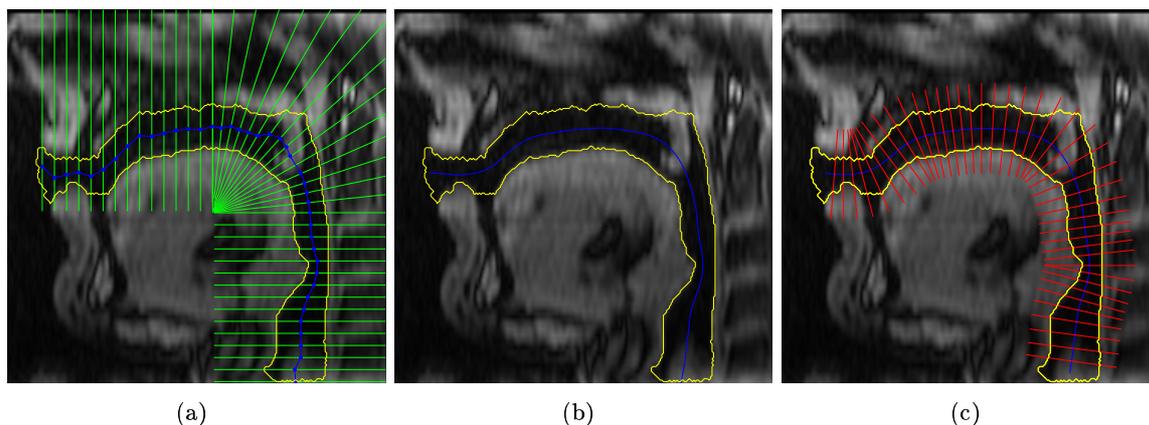


Figura 4.7: Cálculo da grelha adaptativa para o fone [a]. Em (a) representa-se a grelha fixa a verde e uma aproximação inicial para a linha central a azul; em (b) ilustra-se a linha central final; e em (c) a grelha adaptativa gerada.

Os mesmos planos foram então segmentados, implementando duas diferentes técnicas e obtendo-se por fim duas diferentes visualizações. As duas técnicas 2D de segmentação escolhidas foram:

1. Segmentação de cada plano por *Live Wire*: para a implementação da rotina *Live Wire*, optou-se por utilizar uma ferramenta já existente constituída por dois ficheiros: um em código objecto (sem código fonte), `simple1lw.dll` e outro implementado em Matlab, `lwcontour.m`. O *download* desta ferramenta foi efectuado através da pagina <http://www.cs.sfu.ca/~hamarneh/software/livewire/index.html>.

2. Segmentação de cada plano por *Level Set 2D*: a inicialização consiste num círculo de raio 10 e centro no ponto de intersecção entre o plano a ser segmentado e a linha central. A função de velocidade utilizada inclui o Laplaciano e é dada por uma transformação da Equação 3.40. O valor de T foi empiricamente escolhido como 0,04. Esta nova formulação de 3.40 foi combinada linearmente com a função de velocidade 3.32, com $\epsilon = 0,1$ e $T = 0$. A condição de paragem utilizada é a já descrita para o caso 2D. O problema da sobre-segmentação, de forma idêntica ao apresentado para o 2D, foi reduzido recorrendo à função do Matlab `bwlabel`. Esta função tem como saída a região conectada a um ponto de entrada. Escolheu-se uma vizinhança igual a 4 pixels. Inicialmente, pensou-se escolher como ponto de entrada o ponto de intersecção entre o plano a ser segmentado e a linha central. No entanto, existem alguns casos em que este ponto não se encontra no interior do tracto vocal, ver região a verde na Figura 4.8. A solução encontrada foi adoptar este ponto caso a função ϕ nesse ponto seja igual ou maior a 1 e pedir ao utilizador para inserir um ponto pertencente ao tracto vocal, caso contrário.

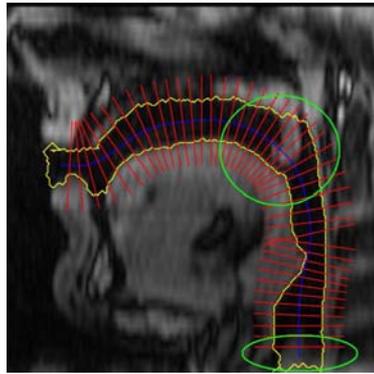


Figura 4.8: Imagem médio-sagital e respectiva grelha adaptativa gerada para o fone [a]. As regiões a verde representam zonas onde os pontos pertencentes à linha central não podem ser escolhidos como sementes (ver texto).

Depois da segmentação dos 45 planos, cada contorno obtido é projectado para o espaço 3D e o resultado final pode ser observado através de uma representação tridimensional do conjunto das segmentações 2D obtidas. A Figura 4.9 mostra alguns dos 45 cortes do volume adquirido para o fone [a] segmentados pelo *Level Sets 2D*. Para a visualização final consultar as Figuras 5.25 e 5.27 do Capítulo dos Resultados.

Segmentação 3D por *Level Sets*

O procedimento aplicado ao 3D é semelhante ao 2D, aumentando apenas uma dimensão. Para a escolha dos vizinhos, após várias tentativas (onde também foram testadas as tradicionais 6, 16 e 26 conectividades), decidiu-se por manter os 8 vizinhos 2D e acrescentar somente o voxel imediatamente à frente e o voxel imediatamente atrás, para a terceira dimensão. O critério de paragem escolhido foi baseado em dois indicadores: (1) a diferença entre o número de voxels classificados como pertencentes à superfície na iteração actual e na anterior é igual

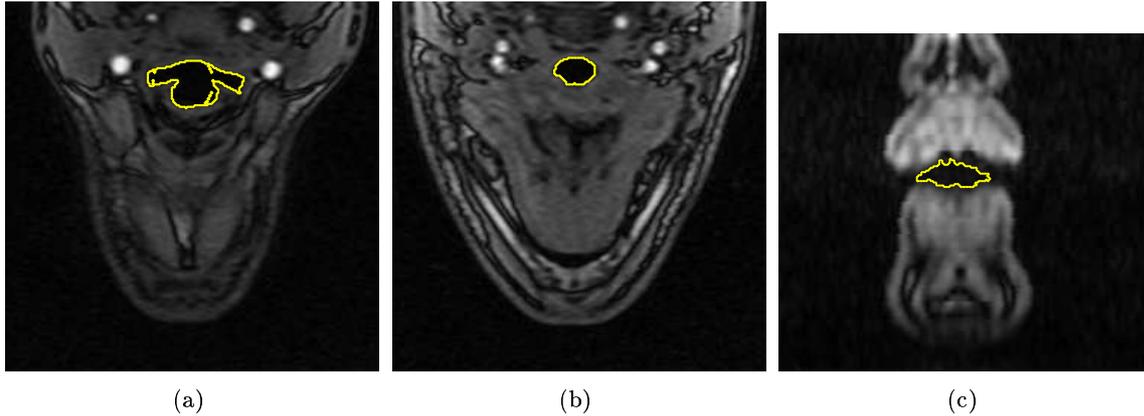


Figura 4.9: Exemplo de três planos (Plano 5, 14 e 45) perpendiculares à linha central do tracto vocal para o fone [a]. As segmentações a amarelo foram obtidas com o *Level Sets* 2D.

a zero e (2) número de iterações maior que 500. O primeiro critério a ser verificado faz com que a rotina termine.

Começou-se por escolher um sub-volume e efectuar sub-amostragem (Secção 4.2.1), de modo a aliviar o peso computacional. De um volume inicial de $512 \times 416 \times 64$ passa-se a trabalhar com um volume de dimensão $48 \times 32 \times 64$. A inicialização do *Level Set* consiste numa esfera centrada no centro do volume, ou seja, no ponto $(24, 16, 32)$ e de raio 16 voxels.

A função de velocidade utilizada inclui o Laplaciano e é dada por uma transformação da Equação 3.40. O valor de T foi empiricamente escolhido como 0,04. Esta nova formulação de 3.40 foi combinada linearmente com a função de velocidade 3.32, com $\epsilon = 0, 1$ e $T = 0$.

O pós-processamento é baseado na ideia já aplicada para o 2D e descrita na Secção 4.2.2. No entanto, tiveram que ser efectuadas algumas modificações, por não existir no Matlab uma rotina equivalente à `regionprops`, para um universo 3D. Criou-se então uma nova rotina que, a partir de um ponto seleccionado pelo utilizador, criasse um vector com todos os pontos a ele ligados com conectividade 26.

4.3 Considerações finais

Os métodos implementados neste trabalho foram escolhidos de forma a se ter um variado leque de graus de interacção com o utilizador. Enquanto que a implementação do método manual, ou ajudado pelo *Live Wire*, implica muita interacção com o utilizador, o *Level Set* necessita da escolha da função de velocidade apropriada (e definição dos respectivos parâmetros) e de uma inicialização para o contorno; por fim, no *Region Growing* são definidos somente um valor de limiar e uma semente. O *Reaction Diffusion* foi escolhido por ser um método de segmentação recente e que possui algumas das características do *Level Sets*, especialmente a facilidade de extensão a 3D.

No que diz respeito à metodologia de segmentação das imagens adquiridas em Tempo Real, ao processar esta informação, optou-se por não ter em consideração a informação temporal. Após as experiências efectuadas, pode-se afirmar que o estudo serviu para testar métodos de segmentação 2D em imagens de baixa qualidade.

A segmentação de volumes é feita com base em duas filosofias distintas. Uma inovadora segmentação $2\frac{1}{2}$ D e a esperada segmentação 3D.

Capítulo 5

Resultados

O quinto Capítulo deste trabalho é dedicado aos resultados obtidos ao longo do estudo. Pretende-se com ele responder a uma série de perguntas, das quais apresentamos de seguida uma selecção:

- É necessário/vantajoso efectuar pré-processamento?
- Qual das técnicas de pré-processamento leva a melhores resultados?
- Para segmentar as imagens 2D, basta uma técnica simples?
- Como se comparam em termos de desempenho as técnicas mais simples de segmentação 2D com outras mais complexas?
- As técnicas aplicada às imagens 2D estáticas podem ser aplicadas às imagens adquiridas em tempo real?
- Como segmentar informação volumétrica?

Sendo assim, a primeira parte do Capítulo foca-se nas técnicas de pré-processamento aplicadas, em especial a eliminação do ar exterior ao tracto vocal e o melhoramento da qualidade de imagens. A segunda parte trata da segmentação de imagens 2D, de imagens com menor qualidade (adquiridas em Tempo Real), e da segmentação de volumes. Termina-se com uma discussão qualitativa das técnicas de segmentação implementadas e umas breves conclusões.

5.1 Pré-Processamento

A **redução do ruído** é um dos fins de algumas técnicas de pré-processamento. Para a optimização das imagens utilizadas para este estudo, foram comparadas diferentes técnicas. A Figura 5.1 apresenta uma comparação visual entre os vários métodos de atenuação de ruído testados. Na Figura 5.2 apresenta-se o efeito da atenuação de ruído em algumas linhas seleccionadas das imagens de Tempo Real.

Para a **avaliação do efeito do pré-processamento** nas imagens, optou-se pelas seguintes métricas: Diferenças absolutas (DA) e PSNR, utilizadas, por exemplo, em (Penedo et al.,

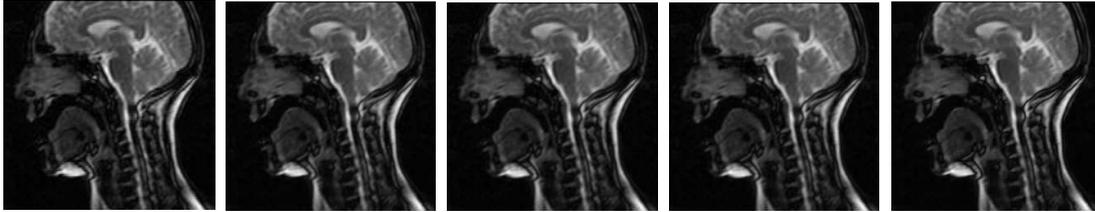


Figura 5.1: Atenuação de ruído: Imagem Original na primeira coluna, Filtro de Mediana na segunda coluna, Filtro de Gaussiana na terceira coluna, Difusão Anisotrópica na quarta coluna, e *Wavelet Denoising* na quinta coluna.

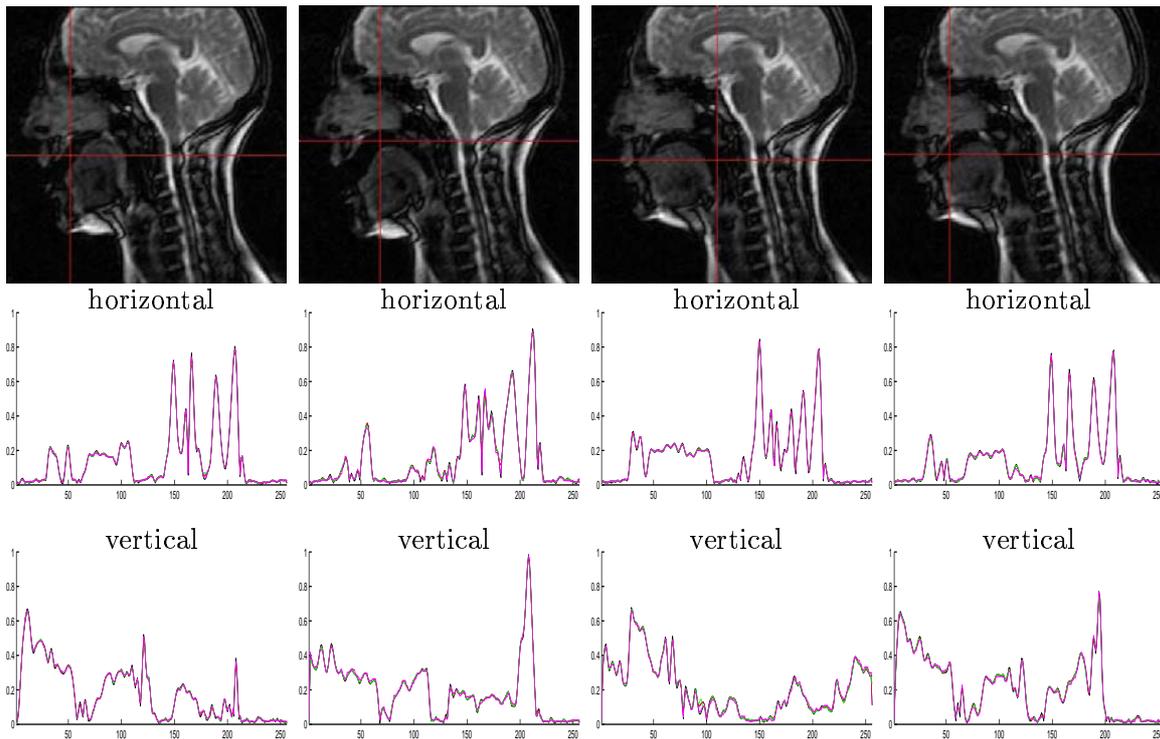


Figura 5.2: Efeito da atenuação de ruído em duas linhas de cada imagem: [k] (esquerda), [a] (esquerda-centro), [t] (centro-direita), e [u] (direita). Imagem Original a preto, Filtro de Mediana a verde, Filtro de Gaussiana a azul, Difusão Anisotrópica a vermelho, e *Wavelet Denoising* a roxo.

2006; Oliveira, 2007). Estes valores são calculados tendo como comparação as imagens originais, ou seja, sem pré-processamento. Na Tabela 5.1 encontra-se uma esquematização dos resultados obtidos. Foi feito um teste estatístico t (através da rotina do Matlab `ttest`) cuja hipótese é verificar se os dados provêm de uma distribuição com média zero. Esta hipótese foi rejeitada com um nível de significância de 5%.

A Tabela 5.1 indica que o melhor método de atenuação de ruído é o *Wavelet Denoising*, por ser o que possui o melhor PSNR (embora a variância seja, em geral, um pouco superior à dos outros métodos). O DA apresenta valores superiores aos dos outros métodos, o que significa

Tabela 5.1: Avaliação do pré-processamento para as imagem de Tempo Real na produção de [k], [a], [t] e [u]. DA = Diferenças absolutas, PSNR = *Peak Signal to Noise Ratio*.

	DA		PSNR	
	Média	Variância	Média	Variância
Filtro Gaussiano	431	49	85	0,01
Difusão Anisotrópica	431	49	85	0,01
Filtro de Mediana	365	39	86	0,04
<i>Wavelet Denoising</i>	559	72	87	0,02

que esta foi a imagem que mais alterações sofreu em relação à imagem original.

Contudo, para se poder avaliar a importância dos efeitos do pré-processamento na qualidade final dos contornos gerados, torna-se necessário estudar os resultados finais da segmentação. Por tal razão, e para uma coerência diacrónica, remete-se este assunto para a Secção 5.3.

5.2 Segmentação de imagens (2D)

Nesta Secção, serão apresentados os resultados relativos à segmentação de **imagens 2D**, dedicando-se a última parte à avaliação dos mesmos.

5.2.1 *Region Growing*

Os resultados relativos à segmentação de imagens 2D pelo *Region Growing* foram já parcialmente apresentados em (Martins, 2007; Martins et al., 2007a). Por uma questão de complementaridade, serão aqui repetidos. Começa-se por explicar o procedimento utilizado de modo a se calcular a média dos contornos, e de seguida apresenta-se alguns testes de avaliação do impacto da localização da semente no contorno final. Refere-se ainda uma métrica que poderá vir a ser útil na classificação dos fones nas diversas classes descritas na Secção 2.2.3.

Média dos contornos

De modo a gerar um **contorno de referência**, começou-se por gerar um contorno com uma semente escolhida manualmente. De seguida, identificaram-se todos os pontos pertencentes ao interior do tracto vocal. Destes, 100 foram escolhidos aleatoriamente, a partir dos quais foram gerados 100 contornos. O contorno de referência, é representado pela média dos 100 contornos gerados em cada imagem.

O cálculo da **média dos contornos** não é um processo trivial. De seguida ilustra-se o processo utilizado, efectuando a média entre dois contornos. Note-se que na realidade, em vez de dois, são utilizados 100 contornos. Nas Figuras 5.3(a) e 5.3(b) são apresentados dois contornos no plano bidimensional e as respectivas coordenadas x e y. As coordenadas x e y foram re-amostradas de modo a que todos os contornos possuam o mesmo número de pontos. A sobreposição das coordenadas x e y na Figura 5.3(c) demonstra a necessidade de fazer um ajuste espacial dos contornos. Tal ajuste, neste caso em particular, foi feito através da

inversão e ajuste dos contornos pelo seu máximo (Figura 5.3(d)). A necessidade de cada uma destas operações é detectada automaticamente pelo algoritmo da forma descrita no parágrafo seguinte.

Todos os contornos são ajustados pelo seu mínimo. De seguida, elimina-se alguns contornos considerados como *outliers*. Isto é feito através do cálculo do Erro Médio Quadrático de cada contorno em relação à média de todos os contornos. Caso este erro seja superior a $3,04 \times 10^6$ (valor escolhido empiricamente), o contorno é eliminado. Note-se que o total de contornos eliminados em cada caso nunca ultrapassou os 18%. Volta-se a calcular o Erro Médio Quadrático de cada contorno em relação à média dos contornos (desta vez sem os contornos eliminados no passo anterior). Através do erro, alguns dos contornos são invertidos. Ou seja, se o erro for ainda maior do que $2,95 \times 10^6$, o contorno em questão é invertido (o ponto final passa a ser o ponto inicial). Após esta operação, os contornos são todos ajustados pelo seu máximo. No fim, é calculada a média entre todos os contornos restantes e ajustados, Figura 5.3(e). Na Figura 5.3(f) são apresentados os dois contornos originais e a sua média.

Alguns **exemplos** dos contornos médios para as vogais cardinais são apresentados na Figura 5.4.

A influência da semente

A influência da semente no contorno final foi avaliada através do cálculo do Índice de Pratt (IP) entre cada um dos contornos gerados e a respectiva média dos contornos. A Figura 5.5 apresenta um gráfico “de bigodes de gato” dos IP obtidos. As caixas apresentam linhas nos quartis inferior, mediano e superior. Os bigodes estendem-se entre os extremos dos dados. Os *outliers* são representados pelo sinal “+”. Para todos os casos testados, os IP situavam-se entre 0,84 e 1. Pode-se desta forma concluir que o método é robusto em relação à posição da semente, isto é, a diferença de posicionamento das sementes na fase de inicialização, não interfere significativamente com o resultado final da segmentação.

Os contornos que apresentam **menor robustez** à colocação da semente ([e], [i~] e [m]) foram localmente analisados de forma a investigar as áreas do tracto vocal que são mais difíceis de segmentar. Na Figura 5.6, o eixo horizontal de cada elipse representa a variância, na direcção horizontal, no ponto correspondente ao centro da elipse. O eixo vertical representa a variância na direcção vertical. De modo a não sobrecarregar a imagem, estão apenas representadas as elipses que, num conjunto de 20 pontos sucessivos, apresentam o maior valor de variância na direcção horizontal somada à direcção vertical. Desta figura conclui-se que as zonas com maior variabilidade localizam-se na parte inferior da faringe. Tal situação não acarreta grande problema, sendo que nos estudos articulatórios a zona de interesse corresponde somente ao tracto vocal (zona entre os lábios e a glote) e que fica acima da zona que apresenta problemas.

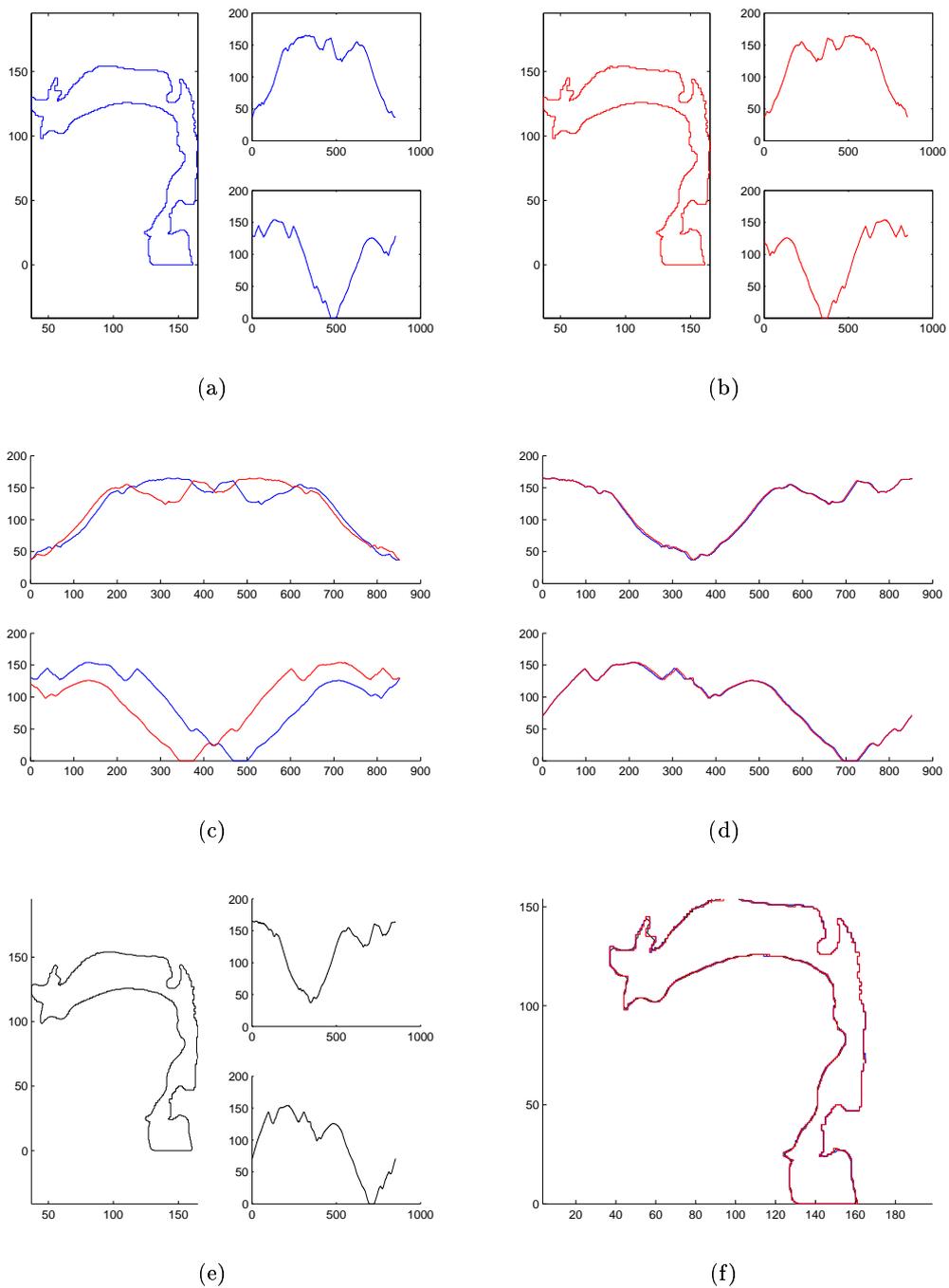


Figura 5.3: Cálculo da média dos contornos. Contorno A a azul, contorno B a vermelho, média dos contornos a preto. Ver texto para mais informações.

Outro estudo com importância no campo da articulação fonética é a implementação de uma **medida que consiga caracterizar cada som** e, ao mesmo tempo, quantificar a influência do contexto na produção dos sons (coarticulação). Para ir ao encontro desta exigência, foram nesta investigação elaboradas uma série de comparações, nomeadamente entre os IP obtidos para os contornos de uma só imagem (Intra) e os IP obtidos para os contornos de imagens diferentes (Inter), apresentados na Figura 5.7. Foi realizado o teste estatístico t com a hipótese

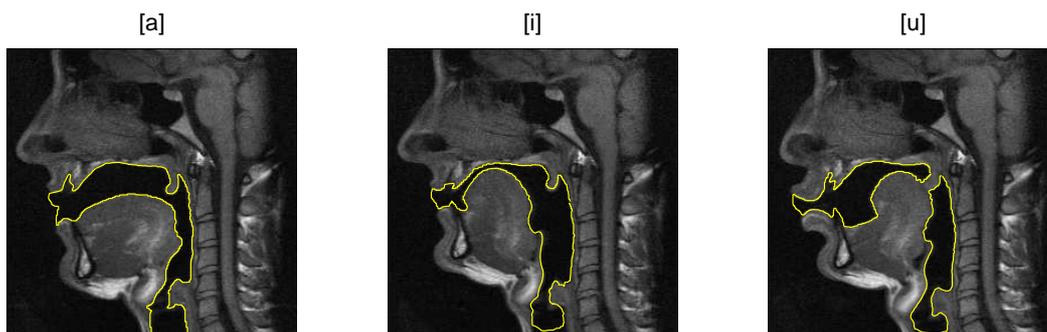


Figura 5.4: Média de 100 contornos gerados com o *Region Growing* 2D: [a] (esquerda), [i] (centro) [u] (direita).

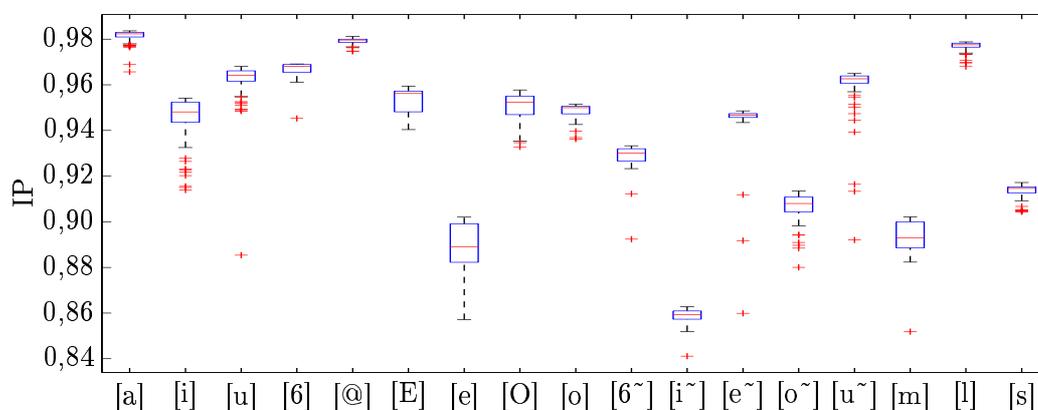


Figura 5.5: Robustez do *Region Growing* 2D em relação à posição da semente.

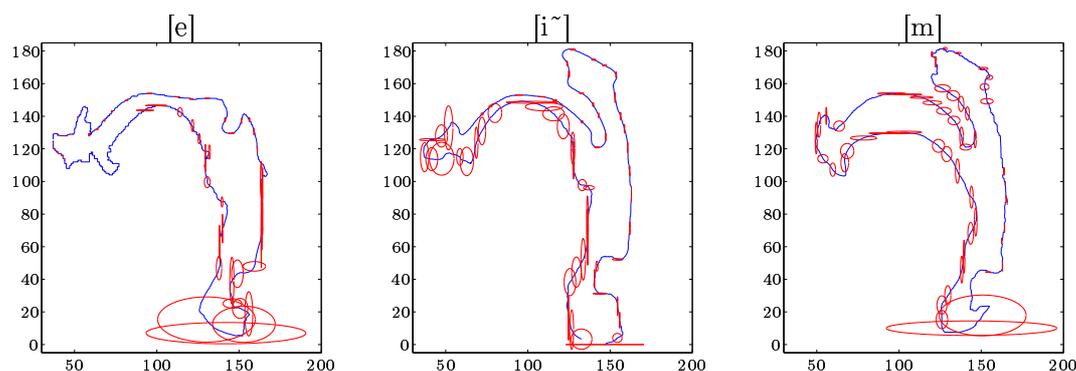


Figura 5.6: Estudo local das zonas com maior variabilidade entre contornos.

de que os dois IPs (Intra e Inter) provêm de distribuições com a mesma média. Esta hipótese foi rejeitada com um nível de significância de 5%, significando que os IPs são estatisticamente diferentes no caso da Intra-variabilidade e da Inter-variabilidade. Conclui-se assim que o IP poderá ser considerado como uma métrica de distinção entre sons e deverá ser futuramente utilizada no estudo da coarticulação.

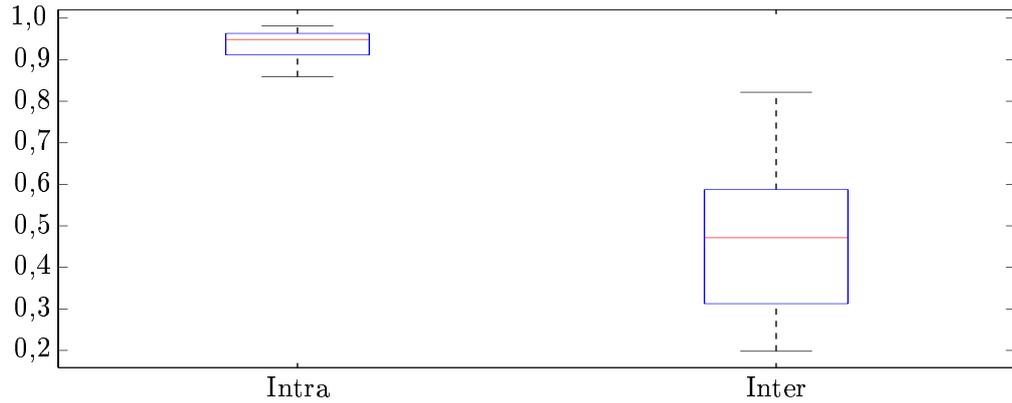


Figura 5.7: Avaliação do IP como caracterizador dos sons.

5.2.2 *Level Sets*

Para a implementação do método *Level Sets*, começou-se, numa primeira fase, por testar diferentes **inicializações**, nomeadamente: utilizando um ponto no interior do tracto vocal, como ilustrado na Figura 5.8; uma linha (Figura 5.9); um círculo (Figura 5.10); e 100 círculos (Figura 5.11).

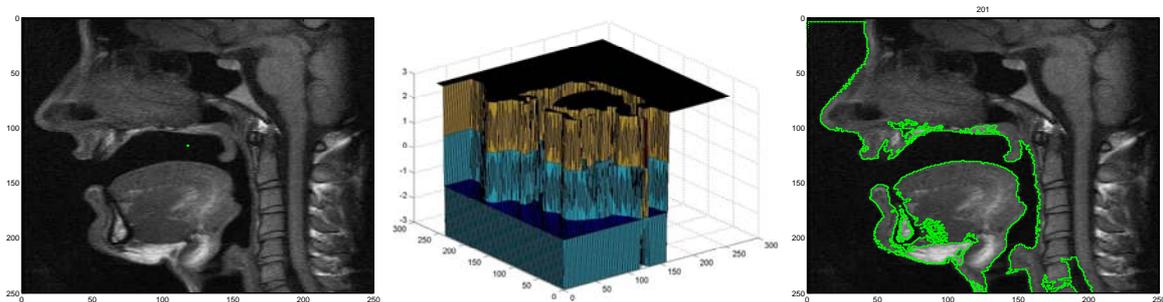


Figura 5.8: Inicialização com um ponto no interior do tracto vocal (esquerda), *Level Set* (centro) e segmentação após 200 iterações (direita).

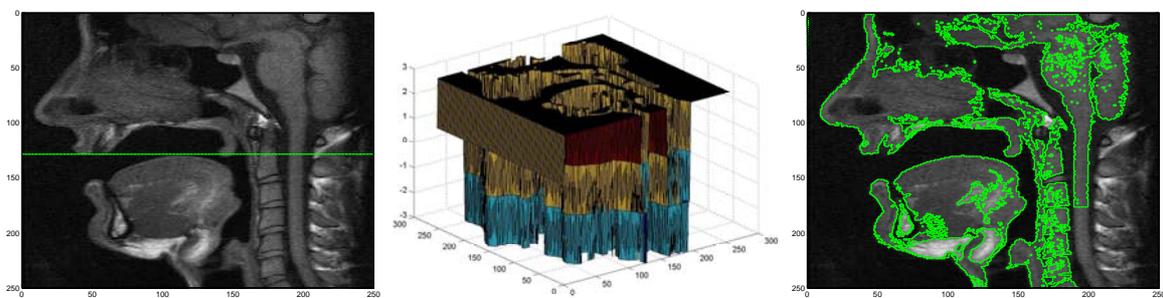


Figura 5.9: Inicialização com uma linha (esquerda), *Level Set* (centro) e segmentação após 190 iterações (direita).

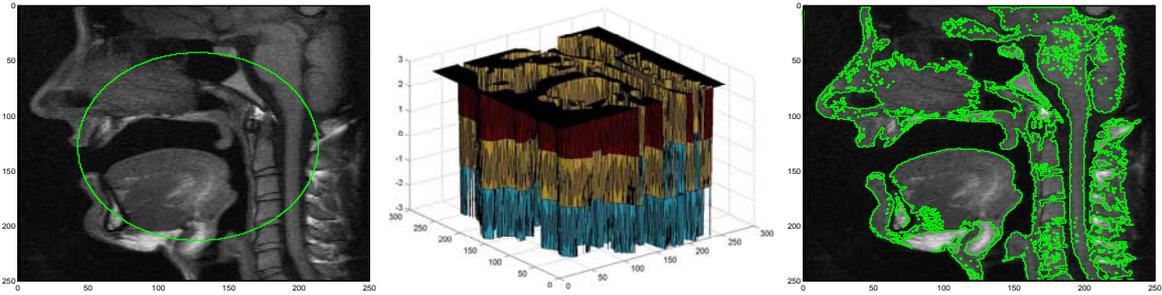


Figura 5.10: Inicialização com um círculo (esquerda), *Level Set* (centro) e segmentação após 95 iterações (direita).

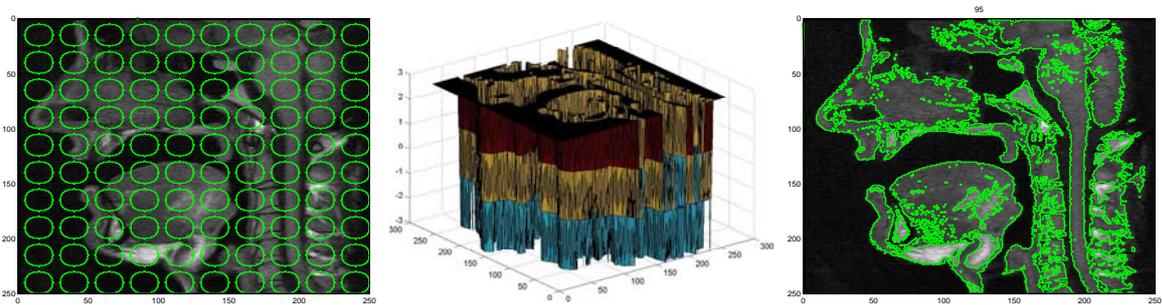


Figura 5.11: Inicialização com 100 círculos (esquerda), *Level Set* (centro) e segmentação após 95 iterações (direita).

Através da análise das imagens anteriores, a primeira inicialização poderá parecer, à primeira vista, a mais adequada, por segmentar o tracto vocal e não se prender a outras estruturas. No entanto, é importante realçar que nesta segmentação o critério de paragem (descrito na Secção 4.2.2) não foi observado, tendo a rotina terminado por chegar ao número máximo de iterações permitidas. Pode-se assim supor que, caso se tivesse deixado a rotina evoluir mais, o resultado da segmentação teria sido idêntico aos obtidos através das outras inicializações. A inicialização com uma linha leva a uma segmentação semelhante às outras duas, mas com um maior número de iterações. As duas inicializações restantes têm um comportamento idêntico. Neste trabalho optou-se pela **inicialização com um círculo** por ter uma evolução mais fácil de seguir visualmente quando comparada com a evolução dos 100 círculos.

Com a inicialização escolhida, testaram-se as três **funções de velocidade** descritas no Capítulo 4.

Tal como visto na Secção 4.2.2, as duas primeiras funções de velocidade não dependem da superfície a evoluir. Podem por isso ser calculadas uma única vez no início da rotina. A terceira função de velocidade tem que ser actualizada a cada iteração, tornando assim a evolução mais lenta. No entanto, o aumento no tempo de processamento é compensado pela convergência num menor número de iterações. A função de velocidade escolhida é assim a função modificada da Equação 3.40 onde a curvatura é substituída pelo Laplaciano.

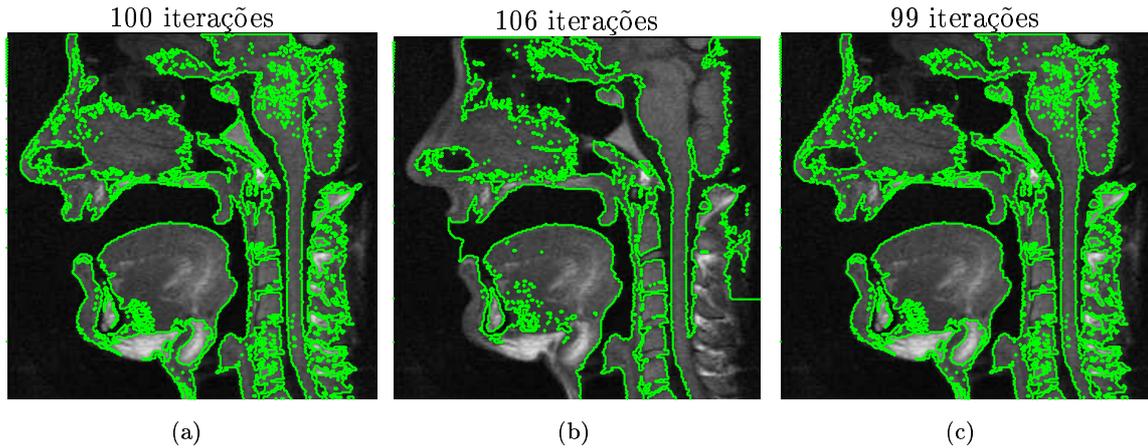


Figura 5.12: Segmentação por *Level Sets* com diferentes funções de velocidade. (a) *Threshold* - função do tipo 3.32; (b) Triangular - função do tipo 3.52; (c) Laplaciano - função do tipo 3.40 onde a curvatura foi substituída pelo Laplaciano. O número de iterações é dado no título de cada imagem.

Foi ainda tentada a evolução com a função de velocidade baseada em fluxos de máximos e mínimos, Equação 3.40. No entanto, esta necessita que o contorno inicial seja o mais próximo possível do contorno que se pretende, neste caso, o contorno do tracto vocal, e foi por isso abandonada.

Tal como já explicado na Secção 4.2.2, foram aplicadas algumas técnicas de **pós-processamento** de modo a reduzir o efeito de sobre-segmentação visível nas imagens da Figura 5.12. A Figura 5.13 apresenta as segmentações finais para fones onde o tracto vocal é conexo, e dividido em dois ou três segmentos.

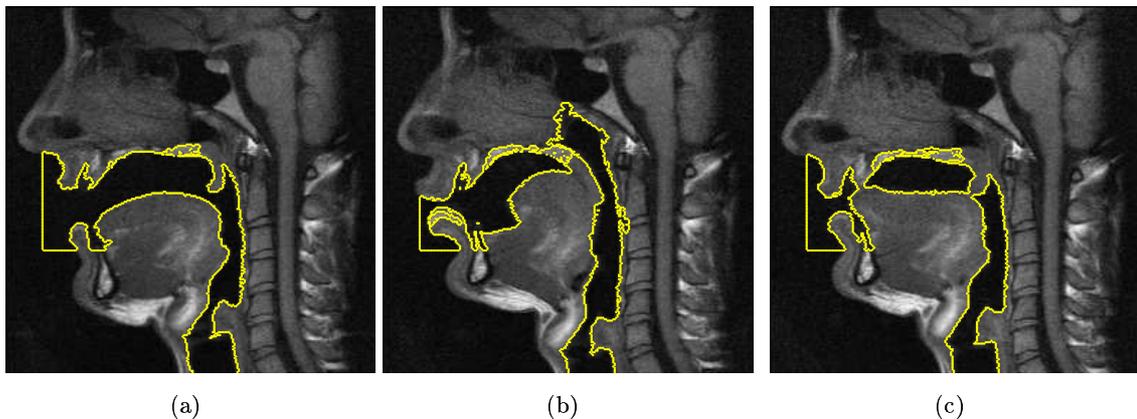


Figura 5.13: Segmentação por *Level Sets* após o pós-processamento: (a) fone [a], (b) fone [u~], e (c) fone [ɪ].

5.2.3 *Reaction Diffusion*

Este método depende do valor de α . Experimentou-se, por isso, alguns valores diferentes para este parâmetro. Os resultados estão representados nas imagens da Figura 5.14.

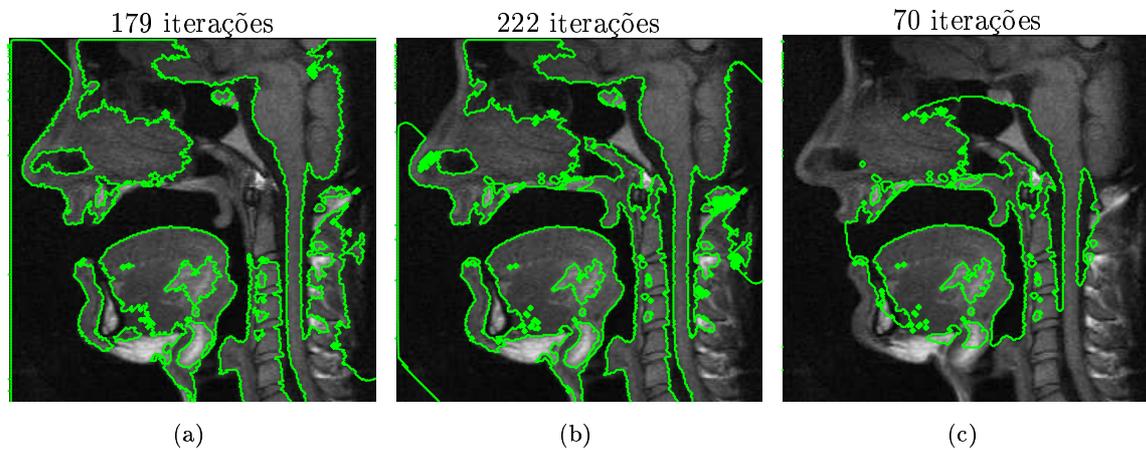


Figura 5.14: Segmentação por *Reaction Diffusion*. (a) $\alpha = 0,0$ e 179 iterações; (b) $\alpha = 0,5$ e 222 iterações; (c) $\alpha = 1,0$ e 70 iterações.

Estas e outras experiências levaram à escolha de $\alpha = 0,8$. Escolhido o valor de α , procedeu-se ao pós-processamento. Os contornos, para comparação com a Figura 5.13, encontram-se na Figura 5.15.

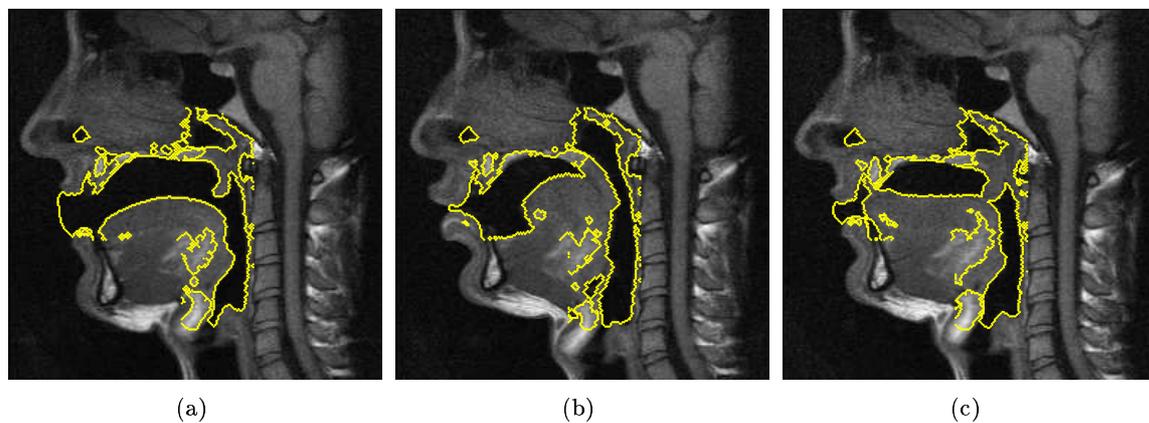


Figura 5.15: Segmentação por *Reaction Diffusion* após o pós-processamento: (a) [a], (b) [u~], e (c) [1].

5.2.4 Comparação entre as três técnicas de segmentação 2D

As segmentações por cada um dos métodos das 64 imagens 2D adquiridas foram **comparadas** entre si. A Figura 5.16 mostra as métricas IP e S_{area} (SA) para as comparações: entre *Reaction Diffusion* e *Region Growing* (RDvsRG), entre *Level Sets* e *Reaction Diffusion*

(LSvsRD), e entre *Region Growing* e *Level Sets* (RGvsLS).

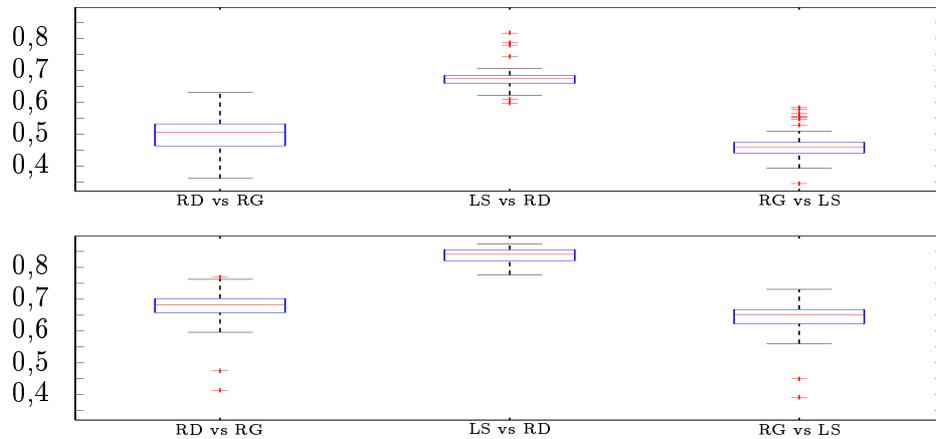


Figura 5.16: Comparação entre *Reaction Diffusion* (RD), *Level Sets* (LS) e *Region Growing* (RG), utilizando o Índice de Pratt (IP) (cima) e a similaridade de áreas (SA) (baixo).

A Figura 5.17 apresenta a comparação entre alguns dos contornos obtidos com *Region Growing* e com *Level Sets*. Na Figura 5.17(a), apresentam-se os contornos que obtiveram uma melhor correspondência, pelo IP, obtidos através da análise comparada dos resultados por implementação do *Region Growing* e do *Level Sets*. Ao centro, a Figura 5.17(b) refere-se ao melhor valor de S_{area} . A Figura 5.17(c) refere-se simultaneamente ao pior IP e ao pior S_{area} . Chegou-se à conclusão que a incongruência nos resultados comparados é devida a uma falha na segmentação por *Region Growing*. Provavelmente, o limiar geral escolhido para segmentar todas as imagens 2D não se adequa a esta imagem.

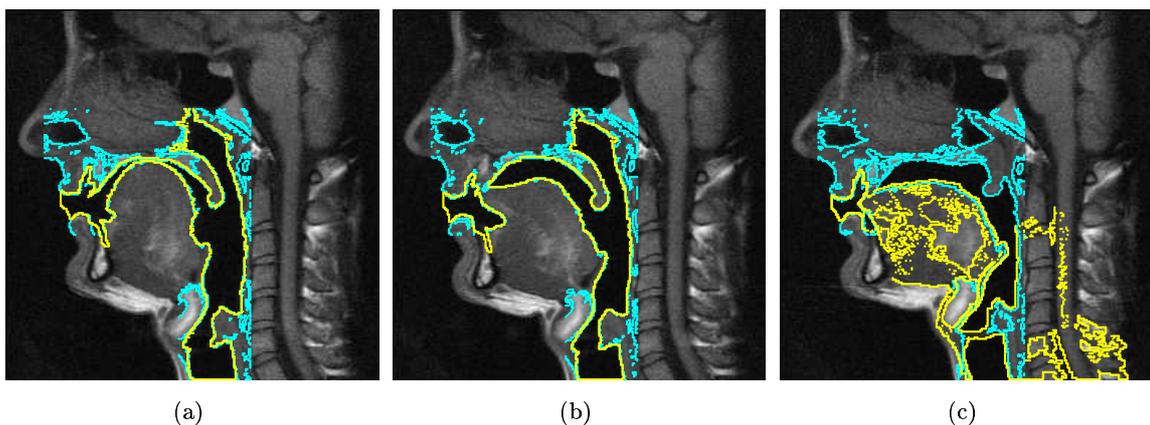


Figura 5.17: Comparação entre segmentações com *Region Growing* e *Level Sets*. A amarelo apresentam-se os resultados obtidos com o *Region Growing* e a azul com o *Level Sets*. Os fonos representados são: (a) [i~] em [pi~t6], e (b) [n] em [k6n6], e (c) [t] em [ata].

A Figura 5.18 apresenta a comparação entre alguns dos contornos obtidos com *Level Sets*

e *Reaction Diffusion*. A Figura 5.18(a) refere-se ao melhor IP obtido entre os resultados do *Level Sets* e do *Reaction Diffusion*. A Figura 5.18(b) refere-se ao melhor S_{area} obtido entre os resultados do *Level Sets* e do *Reaction Diffusion*. A Figura 5.18(c) refere-se ao pior S_{area} obtido entre os resultados do *Level Sets* e do *Reaction Diffusion*. Como já adivinhado pelos gráficos da Figura 5.16 estes dois métodos de segmentação levam sempre a resultados muito semelhantes entre si. Mesmo nos piores casos como o da Figura 5.18(c).

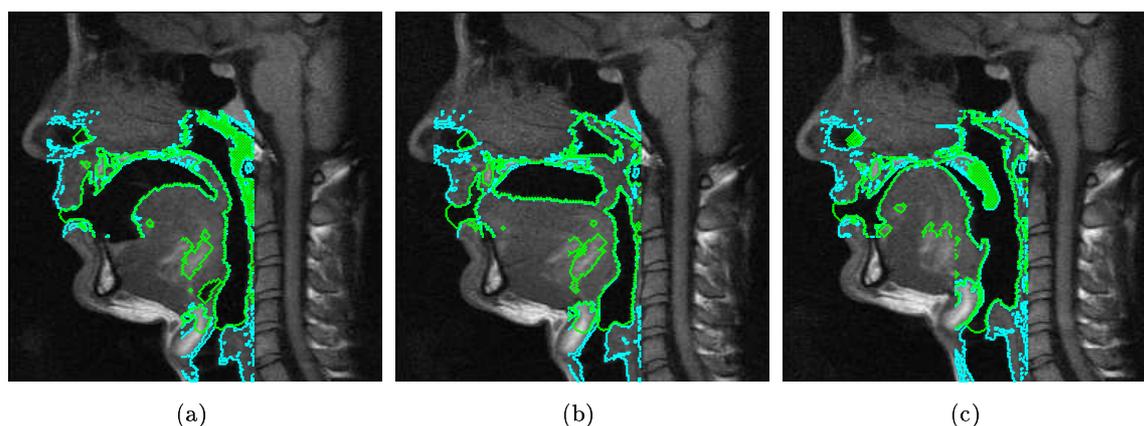


Figura 5.18: Comparação de segmentações entre *Level Sets* e *Reaction Diffusion*. A azul apresentam-se os resultados obtidos com o *Level Sets* e a verde com o *Reaction Diffusion*. Os fonos representados são: (a) [õ] em [põt̃], (b) [l] em [mal̃] e (c) [ĩ] em [mĩtu].

A Figura 5.19 apresenta a comparação entre alguns dos contornos obtidos com *Reaction Diffusion* e *Region Growing*. A Figura 5.19(a) refere-se ao melhor IP obtido entre os resultados do *Reaction Diffusion* e do *Region Growing*. A Figura 5.19(b) refere-se ao melhor S_{area} obtido entre os resultados do *Reaction Diffusion* e do *Region Growing*. A Figura 5.19(c) refere-se ao pior IP obtido entre os resultados do *Reaction Diffusion* e do *Region Growing*. Note-se que esta segmentação por *Region Growing* é a já apresentada na Figura 5.17(c), sendo que o mau desempenho traduzido pela métrica IP é devido a uma falha na segmentação por *Region Growing*. Mais uma vez, o limiar geral escolhido para segmentar todas as imagens 2D não se adequa a esta imagem.

5.3 Segmentação de imagens de Tempo Real

No que se refere às imagens adquiridas em **Tempo Real**, é de se realçar que estas apresentaram uma dificuldade acrescida: a qualidade inferior, em relação às imagens adquiridas em 2D. O objectivo desta Secção é apresentar os resultados das experiências feitas aplicando os métodos acima descritos neste tipo de imagens.

Se nas imagens 2D não se sentiu necessidade de recorrer a métodos de **pré-processamento** que facilitassem a segmentação, neste caso, pelo contrário, a qualidade original das imagens

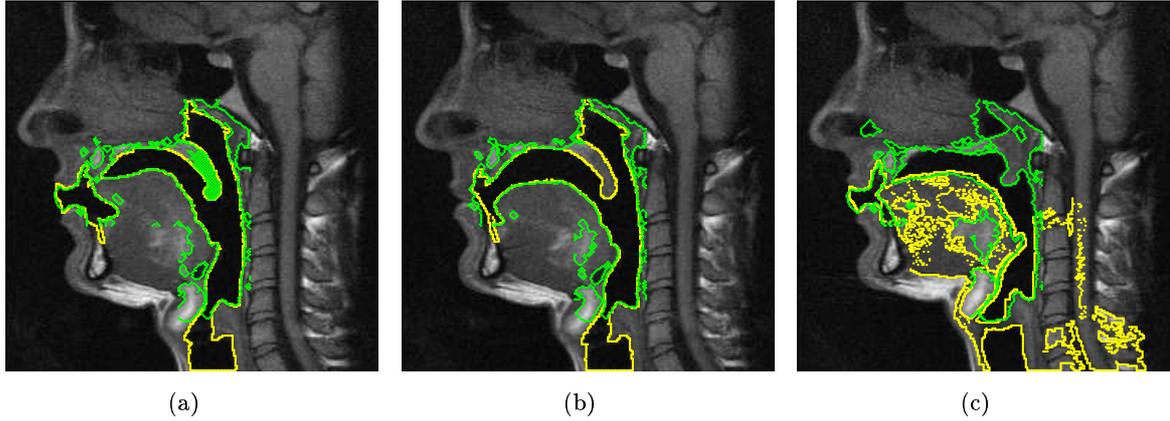


Figura 5.19: Comparação de segmentações entre *Reaction Diffusion* e *Region Growing*. A verde apresentam-se os resultados obtidos com o *Reaction Diffusion* e a amarelo com o *Region Growing*. Os fonos representados são: (a) [n] em [k6n6], (b) [m] em [k6m6], e (c) [t] em [ata].

levou a que se procedesse a algumas tentativas de melhoramento da qualidade da imagem (segundo as técnicas apresentadas na Secção 5.1). Os métodos de segmentação foram por fim aplicados às duas tipologias de imagens: imagens originais, e imagens pré-processadas.

5.3.1 Segmentação manual

O processo de **segmentação manual** requer muita intervenção por parte do utilizador. Por impossibilidade temporal, não foi possível segmentar todas as imagens adquiridas em Tempo Real de forma manual. Segmentaram-se apenas as quatro imagens já apresentadas na Figura 4.2. As segmentações manuais, apresentadas na Figura 5.20, irão ser consideradas como a “ground truth” aquando da análise dos restantes métodos de segmentação testados.

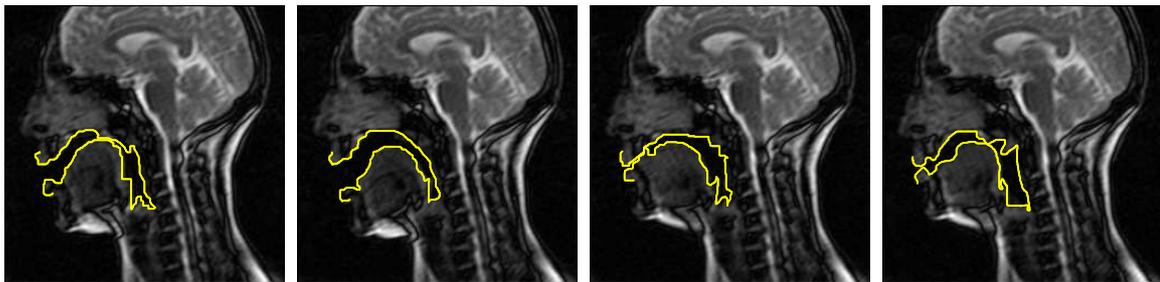


Figura 5.20: Segmentação manual dos frames [k], [a], [t] e [u].

5.3.2 *Region Growing*

Os resultados da aplicação do *Region Growing* a imagens com baixa qualidade são ilustrados na Figura 5.21. Com uma simples avaliação visual é já notória a diferença entre estes

resultados e os obtidos com a segmentação manual. Neste caso, os contornos apresentam-se menos suaves e incluem algumas regiões (como os maxilares e parte da cavidade nasal) não incluídas pela segmentação manual.

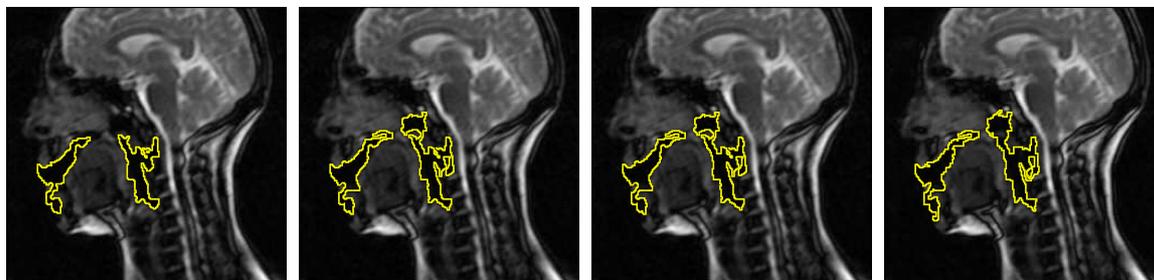


Figura 5.21: Segmentação por *Region Growing* 2D após a atenuação de ruído na imagem correspondente ao fone [k]: Filtro de Mediana na primeira coluna, Filtro de Gaussiana na segunda coluna, Filtro de Difusão Anisotrópica na terceira coluna, *Wavelet Denoising* na quarta coluna.

5.3.3 *Level Sets*

De modo a não sobrecarregar o Capítulo, não se ilustram todas as segmentações obtidas com a aplicação do *Level Sets* às imagens adquiridas em Tempo Real. Refere-se para a Secção 5.3.5 para a comparação entre estes resultados e os obtidos através da segmentação manual.

5.3.4 *Reaction Diffusion*

Este estudo baseou-se também na aplicação de outra técnica de segmentação semi-automática para as imagens de Tempo Real, denominada *Reaction Diffusion*. Os resultados obtidos através da aplicação desta técnica foram analisados comparativamente, em relação aos resultados obtidos através das outras técnicas de segmentação, Secção 5.3.5.

5.3.5 Comparação entre as quatro técnicas de segmentação para as imagens de Tempo Real

A Figura 5.22 encontra-se estruturada de forma a que se consiga interpretar de forma intuitiva as comparações entre os resultados obtidos com a implementação do método manual e do *Region Growing* (azul), do método manual e do *Level Sets* (vermelho), do método manual e do *Reaction Diffusion* (verde), e do *Reaction Diffusion* e do *Level Sets* (verde). Ainda na Figura 5.22, encontram-se os resultados obtidos através da execução dos métodos de segmentação após os diversos tipos de atenuação de ruído: sem atenuação de ruído (IO), após o Filtro de Mediana (FM), após o Filtro de Gaussiana (FG), após o Filtro de Difusão Anisotrópica (DA), e após o *wavelet Denoising* (WD). Por fim, a mesma figura, diferencia cada imagem segmentada pelos símbolos: pontos para representar o fone [k], triângulos para o fone [a],

cruzes para o [t] e asteriscos para o [u].

Destes gráficos podem ser tiradas várias conclusões. Focando a atenção no IP, pode-se começar por referir que os métodos mais semelhantes entre si são o *Reaction Diffusion* e o *Level Set*. O método que produz resultados mais próximos dos resultados obtidos com a segmentação manual é o *Region Growing*, embora a dispersão seja mais elevada do que o desejável. Focando a atenção no S_{area} , pode-se referir que os métodos mais semelhantes entre si são também o *Reaction Diffusion* e o *Level Set*. O método que produz resultados mais próximos dos resultados obtidos com a segmentação manual é ainda o *Region Growing*.

Fazendo agora uma análise compartimentada, e começando pela comparação entre o método manual e os *Level Set*, a análise dos IP e S_{area} da Figura 5.22 demonstra que o melhor método de atenuação de ruído é o **Filtro de Gaussiana**. Com o objectivo de ilustrar estes resultados apresenta-se, na Figura 5.23, os resultados da segmentação após a aplicação deste filtro.

Passando à comparação entre o método manual e o *Reaction Diffusion*, através da análise dos IP e dos S_{area} obtidos com a aplicação de diferentes técnicas de atenuação de ruído, chegou-se à seguinte conclusão: em imagens com uma qualidade reduzida, tais como as imagens de Tempo Real, o melhor método de atenuação de ruído é o **Filtro de Gaussiana**. Os resultados estão apresentados na Figura 5.24.

Para terminar, e focando os IP, concluí-se que o método de atenuação de ruído mais fiável, isto é, o que leva à criação de contornos mais semelhantes aos obtidos com a segmentação manual, é o *Wavelet Denoising*. Excepções para a segmentação do [a] e do [t], onde se obteve melhores resultados com a Imagem Original. Usando como métrica de comparação a S_{area} , os melhores resultados foram obtidos com o Filtro de Mediana.

Os resultados presentes nos gráficos da Figura 5.22 mostram claramente a **incompatibilidade** entre estas duas métricas. Em muitos casos, notou-se que os resultados das análises dos IP é incongruente com os da análise da S_{area} . Isto é, enquanto que segundo o primeiro, um certo método de pré-processamento parece ser o melhor, ao analisar a S_{area} , o mesmo método é o menos apropriado. Por fim, é interessante realçar que o Filtro de Gaussiana e o Filtro de Difusão Anisotrópica levam aos mesmos resultados.

5.4 Segmentação de volumes (3D)

No que diz respeito à **segmentação de volumes**, ou informação 3D, resolveu-se utilizar dois métodos distintos: segmentação $2\frac{1}{2}$ D (ou seja, segmentação de um conjunto de planos retirados do volume) e segmentação 3D (segmentação onde é utilizada toda a informação presente no volume). De seguida, são apresentados os resultados das segmentações obtidas para cada uma destas técnicas.

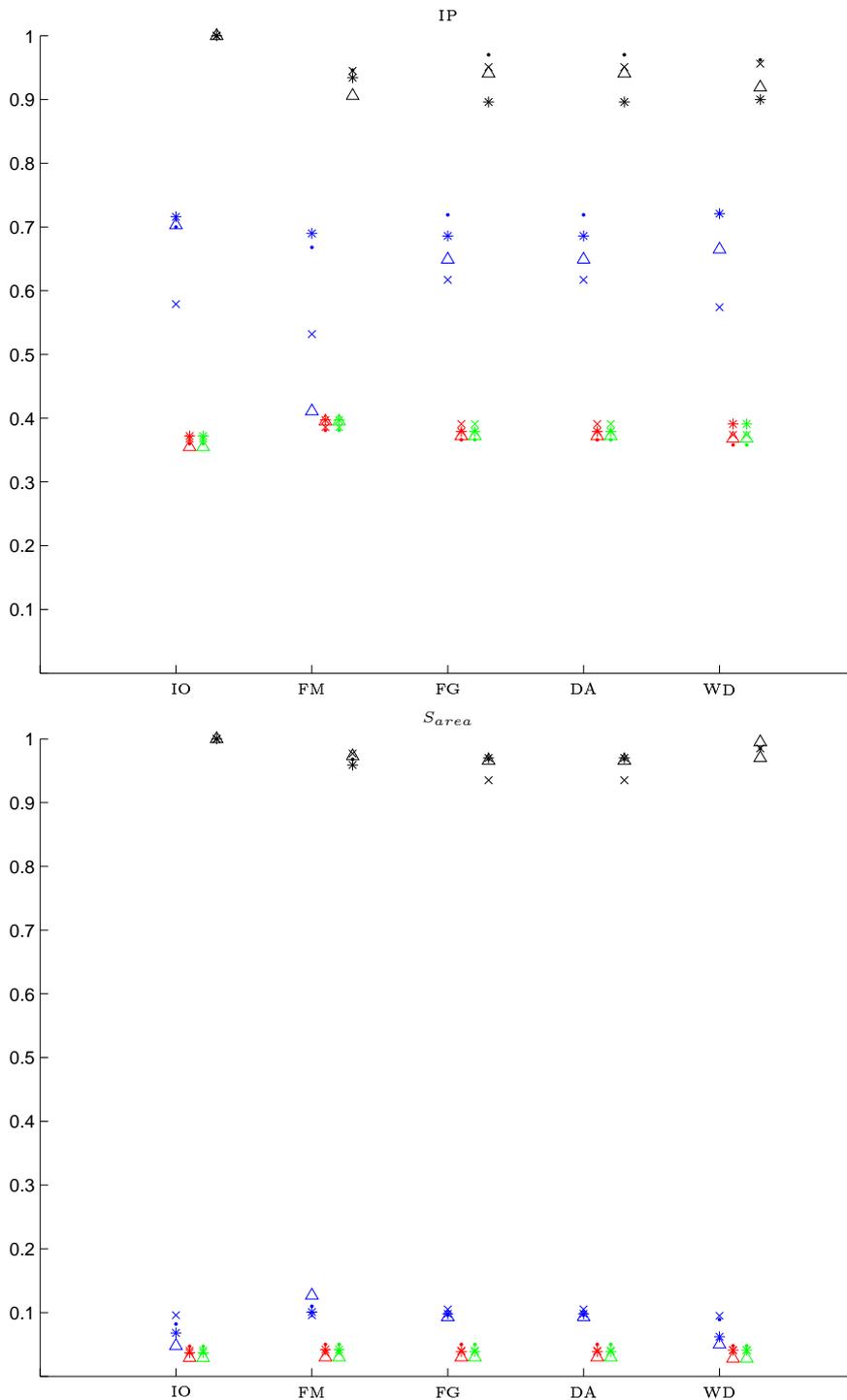


Figura 5.22: Comparação entre os vários resultados obtidos para as imagens em Tempo Real correspondentes à produção de [k] (pontos), [a] (triângulos), [t] (cruzes) e [u] (asteriscos). A azul apresenta-se a comparação entre o método manual e o *Region Growing*, a vermelho a comparação entre o método manual e o *Level Set*, a verde a comparação entre o método manual e o *Reaction Diffusion*, e a preto a comparação entre o *Reaction Diffusion* e o *Level Set*. IO = Imagem Original, FM = Filtro de Mediana, FG = Filtro de Gaussiana, DA = Difusão Anisotrópica, e WD = *wavelet Denoising*.

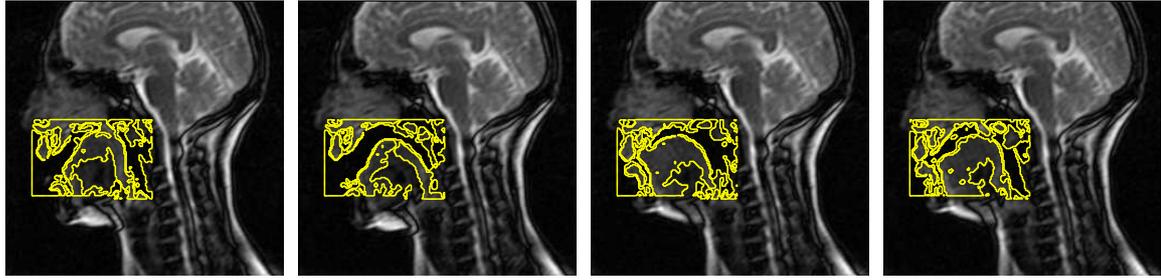


Figura 5.23: Segmentação por *Level Sets* após a aplicação do Filtro de Gaussiana às imagens de Tempo Real correspondentes à produção de [k], [a], [t] e [u].

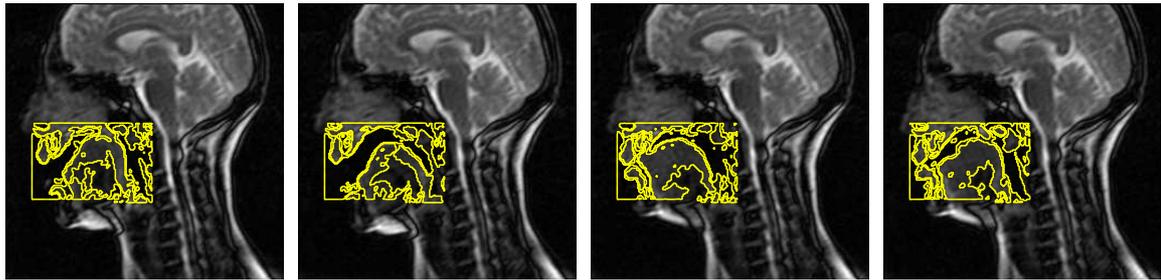


Figura 5.24: Segmentação por *Reaction Diffusion* após a aplicação do Filtro de Gaussiana às imagens de Tempo Real correspondentes à produção de [k], [a], [t] e [u].

5.4.1 Segmentação $2\frac{1}{2}D$

A **segmentação $2\frac{1}{2}D$** , segmentação de um conjunto de planos retirados do volume, pode-se definir por segmentação “corte a corte” (tal como apontado na Secção 4.2.2), e foi efectuada através da utilização de dois métodos: *Live Wire* e *Level Set*. Exemplos de segmentações com o *Live Wire* para as vogais cardinais [i], [u] e [a] são apresentadas na Figura 5.25.

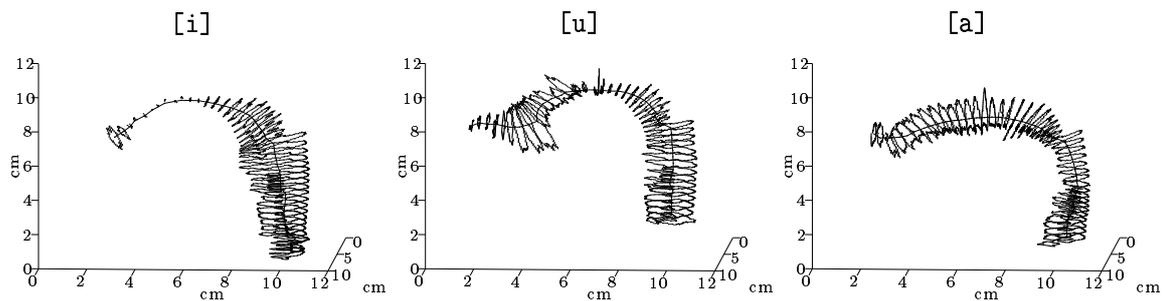


Figura 5.25: Segmentação $2\frac{1}{2}D$ por *Live Wire* das três vogais cardinais, [i], [u] e [a].

A aplicação de uma *mesh* aos resultados apresentados na Figura 5.25 resulta nos volumes da Figura 5.26.

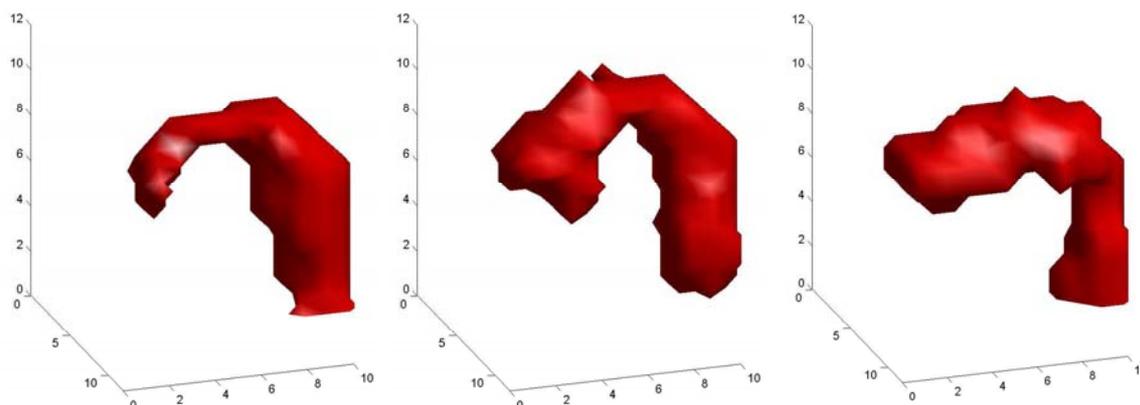


Figura 5.26: Aplicação de uma *mesh* à segmentação $2\frac{1}{2}$ D por *Live Wire* das três vogais cardinais, [i], [u] e [a].

A Figura 5.27 apresenta alguns **exemplos** de segmentações com o *Level Sets* para as vogais cardinais [i], [u] e [a].

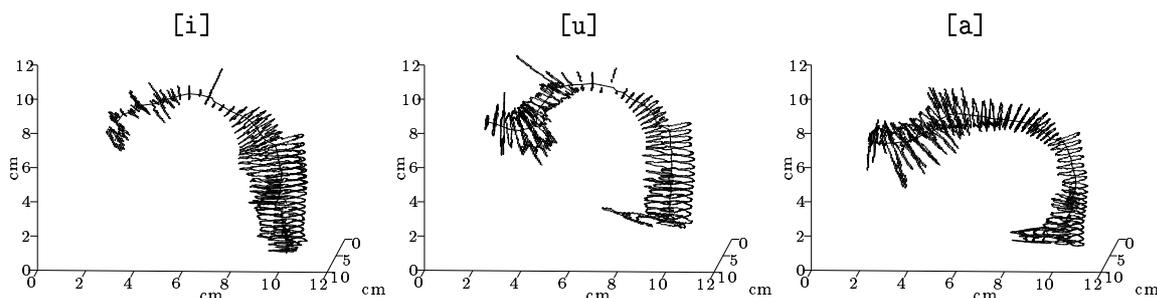


Figura 5.27: Segmentação $2\frac{1}{2}$ D por *Level Sets* das três vogais cardinais, [i], [u] e [a].

A Figura 5.28 apresenta os produtos da aplicação de uma *mesh* aos resultados apresentados na Figura 5.27.

À luz dos resultados obtidos, conclui-se que o *Live Wire*, tendo uma maior intervenção com o utilizador, tem a vantagem de ser mais fidedigno por se basear na experiência adquirida por um técnico. No entanto, esta mesma vantagem torna-se numa desvantagem, devido às variâncias intra e inter sujeitos. O segundo método, *Level Sets*, pode levar a resultados menos precisos, mas possui uma menor variância e apresenta, por isso, uma maior reprodutibilidade.

5.4.2 *Level Sets* 3D

Foi já referido que a segmentação de volumes foi executada segundo duas modalidades diferentes: segmentação $2\frac{1}{2}$ D e segmentação 3D. Para o primeiro caso, como foi exposto na Secção anterior, a segmentação do volume consistiu num conjunto de segmentações 2D.

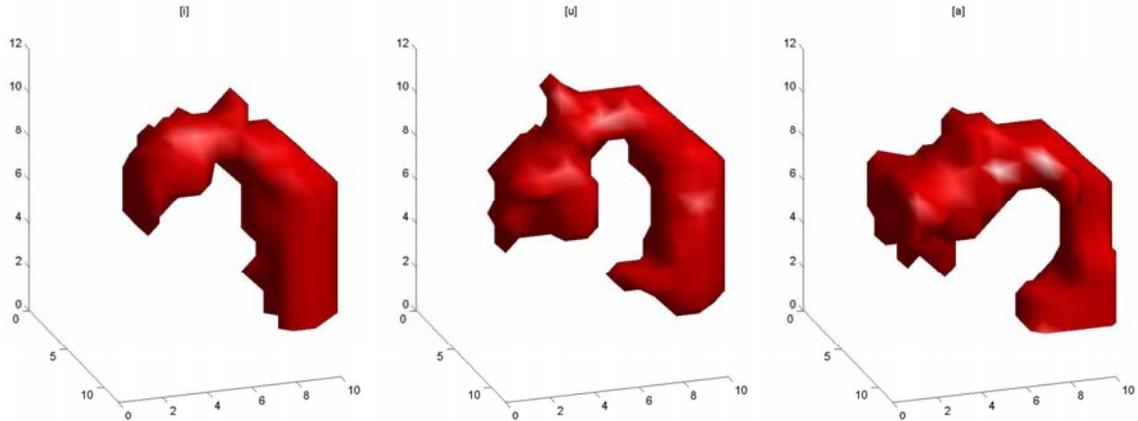


Figura 5.28: Aplicação de uma *mesh* à segmentação $2\frac{1}{2}D$ por *Level Sets* das três vogais cardinais, [i], [u] e [a].

Passaremos, nesta Secção, a apresentar os resultados obtidos com as **segmentações 3D**.

A segmentação 3D foi feita através de uma adaptação da técnica do *Level Sets* 2D para o universo 3D. Tal como seria de esperar pelos resultados obtidos em 2D, no caso 3D os volumes sofreram também o efeito de sobre-segmentação. Isto obrigou a efectuar um tratamento da imagem após a segmentação (pós-processamento) com o fim de eliminar os efeitos da sobre-segmentação. Este pós-processamento é semelhante ao utilizado no 2D e foi já descrito na Secção 4.2.2. Um exemplo de segmentação de um volume com e sem pós-processamento é dado na Figura 5.29.

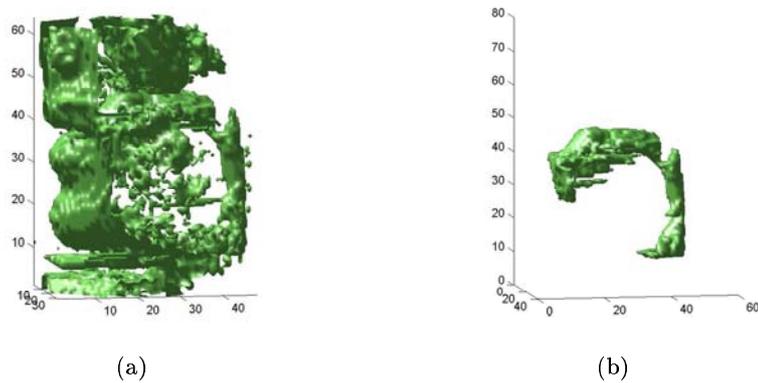


Figura 5.29: Efeitos do pós-processamento na segmentação 3D efectuada por *Level Sets*: (a) resultado da segmentação do volume antes da fase de pós-processamento, onde é bem visível o efeito da sobre-segmentação; (b) resultado da segmentação do volume após a fase de pós-processamento.

A Figura 5.30 apresenta os **resultados da segmentação 3D** por *Level Sets* para os fonemas [i], [u] e [a].

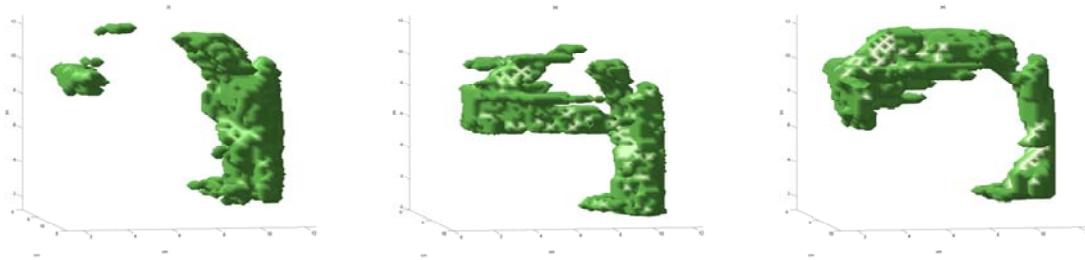


Figura 5.30: Visualização da segmentação 3D por *Level Sets* das três vogais cardinais, [i], [u] e [a].

Mesmo após a execução do pós-processamento, denotam-se ainda alguns **problemas** de sub-segmentação ([i] na Figura 5.30) e de sobre-segmentação ([u] na Figura 5.30). A zona que apresenta maiores falhas é a cavidade oral, sendo, pelo contrário, a zona da faringe sempre segmentada de forma satisfatória. A zona da úvula é sempre bem visível, contrariamente ao que acontece nas imagens das Figuras 5.26 e 5.28. Esta é uma observação importante, tendo-se trabalhado neste caso com volumes sub-amostrados (enquanto que no método de segmentação $2\frac{1}{2}$ D se trabalhou com volumes sobre-amostrados). No futuro, a otimização das rotinas criadas irá permitir que se trabalhe com a qualidade original dos volumes, prevendo-se que as segmentações obtidas desta forma sejam mais pormenorizadas.

5.4.3 Comparação entre as três técnicas de segmentação para os volumes

A **análise comparada** dos dois métodos de segmentação para os volumes assentou no estudo das suas funções de área, sendo este tipo de estudo o mais utilizado pelos investigadores na área da fala e em especial na síntese articulatória (Apostol et al., 2000, 1999; Badin et al., 1998a; Baer et al., 1991; Demolin et al., 1996; Engwall e Badin, 2000; Ericsson, 2005; Greenwood et al., 1992; Heinz e Stevens, 1964; Kröger et al., 2000; Narayanan et al., 1997; Shadle et al., 1999; Soquet et al., 2002; Story et al., 1996; Takemoto et al., 2003, 2004). As funções de área representam a área do tracto vocal em função da sua distância à glote. De seguida, na Figura 5.31, apresenta-se uma comparação das funções de área obtidas com cada um dos métodos para os fones [i], [u] e [a].

Na Figura 5.31 apresentam-se as **funções de área** obtidas com as segmentações $2\frac{1}{2}$ D (com *Live Wire* e com *Level Sets*) e 3D (com *Level Sets*).

A média dos **Erros Quadráticos** entre as duas segmentações $2\frac{1}{2}$ D é: 0,6 cm para o fone [i], 1,1 cm para o fone [u], e 1,0 cm para o fone [a], sendo que as segmentações com *Level Sets* produzem áreas inferiores às obtidas com *Live Wire*. A zona do tracto vocal em que os Erros Quadráticos são mais acentuados é a cavidade bucal. Este resultado é facilmente explicado pela forma como a aquisição do volume foi executada, dando maior resolução ao plano axial, em detrimento dos planos sagitais e coronais. Tal escolha levou à obtenção de

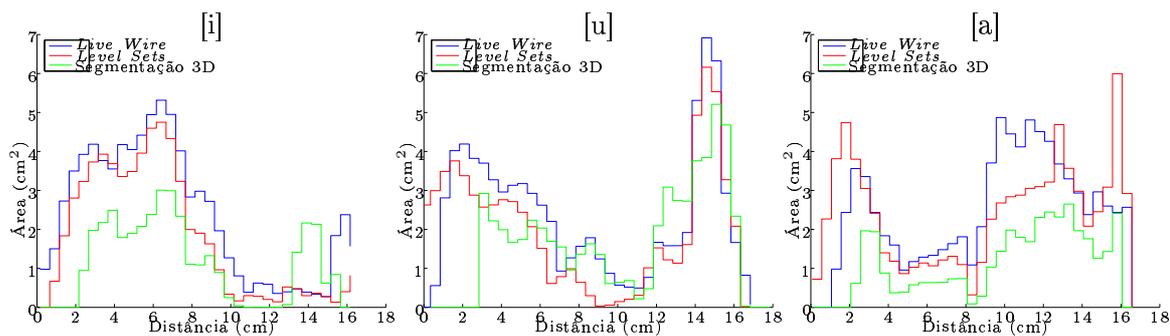


Figura 5.31: Comparação entre as funções de área obtidas com a segmentação $2\frac{1}{2}$ D - *Live Wire* (a azul) e *Level Sets* (a vermelho) - e com a segmentação 3D - a verde - para as três vogais cardinais, [i], [u] e [a].

uma maior resolução na zona correspondente à cavidade faríngea, sendo por isso esta mais facilmente segmentável.

Comparando as segmentações $2\frac{1}{2}$ D com a 3D, observa-se que esta última leva quase sempre a áreas menores. Um pormenor que fica para ser aperfeiçoado no futuro é a escolha do valor zero. Na literatura este valor corresponde normalmente à glote. Neste trabalho, por se ter encontrado dificuldades na determinação da glote, decidiu-se ajustar as funções de área de forma empírica, confiando no bom senso e na observação visual. Sendo assim, as funções de área correspondentes ao [i] e ao [u] foram ajustadas pelos seus máximos, e as funções de área correspondentes ao [a] foram ajustadas pelos seus mínimos.

5.5 Discussão

O trabalho efectuado permitiu a criação de uma plataforma de conhecimentos empíricos através da qual se pode avançar para uma **análise comparativa** dos métodos de segmentação e consequente avaliação da sua eficácia. Segundo Falcão et al. (2000) e Poon et al. (2007), tal avaliação pode ser feita tendo por base três indicadores:

- **Precisão:** reflecte a possibilidade de repetição do método, podendo ser avaliada através das variações nos resultados da segmentação devidas à subjectividade introduzida pelo utilizador;
- **Exactidão:** refere-se ao grau de semelhança entre os resultados da segmentação e a estrutura anatómica real;
- **Eficiência:** avalia-se com base na “usabilidade” do método, que se mede através de uma função do tempo total necessário para completar o processo de segmentação.

Este conjunto de indicadores foi adoptado para a avaliação dos três diferentes tipos de dados nos quais os métodos de segmentação foram implementados.

No que se refere às segmentações em ambiente **2D**, e em relação à Precisão, os métodos que sofrem menos de subjectividade introduzida pelo utilizador são o *Level Sets* e o *Reaction Diffusion*, por serem praticamente imunes à inicialização. A Exactidão dos métodos não pode ser medida por não se ter adoptado uma referência exacta do verdadeiro contorno do tracto vocal. A aproximação normalmente utilizada é a comparação dos contornos semi-automáticos com contornos manuais, assumindo estes últimos como mais prováveis. No entanto, devido à variabilidade intra e inter sujeito que caracteriza o processo de elaboração de contornos manuais, optou-se por não aprofundar este tipo de análise comparativa. Uma inspecção visual das segmentações obtidas, levou a considerar o método do *Region Growing*, como o mais apropriado para a segmentação de imagens bidimensionais. O *Region Growing*, cujo tempo de processamento é o menor, é considerado o mais Eficiente.

No que se refere às técnicas adoptadas para a segmentação de imagens de **Tempo Real**, os resultados obtidos pela avaliação dos métodos segundo a sua Precisão, Exactidão e Eficiência, é semelhante à das imagens 2D. Isto é, a segmentação manual executada para as imagens de Tempo Real, apresenta um baixo grau de Precisão, sofrendo bastante com a variabilidade intra e inter sujeito, mas um elevado grau de Exactidão. Os métodos mais precisos são o *Level Sets* e o *Reaction Diffusion*. Em relação à sua Eficiência, a segmentação manual requer grande intervenção por parte do utilizador, tornando-se, por isso, demorada, e foi avaliada de forma negativa. O *Region Growing* é o método mais Eficiente. Ao comparar todos os indicadores e incluído a (subjectiva) inspecção visual da autora, considera-se que o melhor método para segmentar as imagens de Tempo Real é o *Region Growing*, mas apenas com a condição de se diminuir previamente o ruído através da utilização do método *Wavelet Denoising*. Com base no sistema de avaliação de Falcão et al. (2000), adoptado para esta análise comparativa, pode-se afirmar então que o método é (simultaneamente) Preciso, Exacto e Eficiente de forma satisfatória.

No que se refere à avaliação dos métodos aplicados à segmentação de volumes, o *Level Sets* 3D são o método mais Preciso e Eficiente, sendo que o método mais exacto é a segmentação $2\frac{1}{2}$ D por *Live Wire*. Conclui-se que, no universo tridimensional, o método que se mostra mais indicado é a segmentação $2\frac{1}{2}$ D por *Live Wire*.

Os resultados obtidos, resumidos na Tabela 5.2, consideram-se satisfatórios, no que diz respeito à sua aplicação concreta, nomeadamente na sua funcionalidade e aplicabilidade na área da segmentação de imagens e volumes.

5.6 Considerações finais

O presente Capítulo teve como objectivo a apresentação de um conjunto de resultados obtidos através da aplicação de métodos de segmentação a imagens e volumes de RM. A primeira parte é dedicada aos resultados obtidos ao longo do estudo, e foi estruturada de forma a apresentar não apenas os resultados em si, mas também de forma a proporcionar,

Tabela 5.2: Avaliação das técnicas de segmentação.

	Precisão	Exactidão	Eficiência
Manual 2D	--	++++	---
<i>Region Growing 2D</i>	+	+++	-
<i>Level Sets 2D</i>	++	+	--
<i>Reaction Diffusion 2D</i>	++	+	--
Segmentação $2\frac{1}{2}$ D por <i>Live Wire</i>	-	++++	----
Segmentação $2\frac{1}{2}$ D por <i>Level Sets</i>	+	+++	---
<i>Level Sets 3D</i>	++	++	-

ao longo da exposição, uma análise comparada entre diferentes resultados obtidos através da implementação de diferentes métodos de segmentação. Isto em ambiente bidimensional e tridimensional. O estudo comparativo levou à conclusão que a valência das técnicas de segmentação a utilizar não se pode avaliar univocamente, isto é, em absoluto, mas depende do tipo de dados (2D ou 3D) com que se está a trabalhar. Ou seja, informações de diferentes naturezas necessitam de diferentes métodos para o seu processamento.

O Capítulo acaba com uma síntese do exposto anteriormente, composta por uma análise comparativa das metodologias utilizada, assente em indicadores predefinidos (nomeadamente Eficiência, Precisão e Exactidão). Concluiu-se que, segundo os parâmetros adoptados e para a segmentação de imagens 2D, aconselha-se a utilização do método *Region Growing*, bem como para a segmentação das imagens de Tempo Real, embora, neste caso, apenas após a aplicação do *Wavelet Denoising*; para a segmentação de volumes, considerou-se mais indicado o uso de uma segmentação $2\frac{1}{2}$ D recorrendo ao *Live Wire* 2D para a segmentação de cada plano criado neste processo.

Como contributo para futuras investigações, este Capítulo pode ser considerado o epílogo dos esforços e do trabalho efectuado. O próximo Capítulo apresentará as principais conclusões formuladas graças a este estudo.

Capítulo 6

Conclusões

O presente trabalho representa uma sub-tarefa de um projecto mais amplo: o projecto HERON, cujo objectivo é a criação de um simulador da fala. Tal propósito requer uma série de conhecimentos e investigações pluridisciplinares prévias, entre as quais o conhecimento de uma das estruturas anatómicas responsáveis pela produção da fala: o tracto vocal. Ao momento do início do projecto HERON, já se encontravam disponíveis informações acerca da produção do PE, nomeadamente dados EMMA. Contudo, estes possuem resolução espacial reduzida, o que representa uma limitação ao horizonte de conhecimentos necessários. Decidiu-se portanto complementar os dados já existentes através da aquisição de nova informação. Optou-se pela elaboração, processamento e análise de imagens de RM do tracto vocal.

A extracção da informação útil, através da segmentação de dados de RM do tracto vocal, representou o objectivo específico da presente investigação, que se foca na utilização das técnicas de segmentação em dados de RM, e cujos resultados foram nesta dissertação apresentados.

Neste último Capítulo, percorrem-se as diferentes fases do trabalho realizado, revisita-se a arquitectura geral da dissertação e focam-se as principais conclusões retiradas da investigação realizada, realçando o potencial inovador da abordagem efectuada.

6.1 Síntese do trabalho desenvolvido

Cronologicamente, as tarefas desempenhadas seguiram a seguinte sequência:

1. Revisão bibliográfica na área da produção de fala, com particular atenção aos artigos com dados adquiridos por RM;
2. Colaboração na fase de selecção do *corpus*;
3. Colaboração na fase de aquisição do *corpus*;
4. Revisão bibliográfica dos métodos de segmentação de imagens adquiridas por RM e utilizadas na área de estudo da produção de fala;

5. Revisão bibliográfica dos métodos de segmentação de imagens e volumes utilizados na área de processamento de sinal;
6. Tratamento de dados 2D:
 - (a) Implementação do método de *Region Growing*;
 - (b) Implementação do método de *Reaction Diffusion*;
 - (c) Implementação do método de *Level Sets*;
7. Tratamento dos dados em Tempo Real:
 - (a) Implementação de técnicas de remoção de ruído (pré-processamento):
 - i. Filtro de mediana;
 - ii. Difusão Anisotrópica;
 - iii. Filtro Gaussiano;
 - iv. *Wavelet denoising*;
 - (b) Experimentação dos métodos de segmentação já implementados para as imagens 2D após a aplicação de cada um dos métodos de atenuação de ruído implementados;
8. Tratamento dos dados 3D através de duas abordagens distintas:
 - (a) Segmentação $2\frac{1}{2}$ D:
 - i. Por *Live Wire*;
 - ii. Por *Level Set*;
 - (b) Segmentação 3D:
 - i. Por *Level Set*;
9. Elaboração da presente dissertação de mestrado que documenta o trabalho efectuado.

6.2 Principais resultados

Embora a segmentação de imagens não represente, por si só, um objectivo no projecto HERON, torna-se de extrema importância para o objectivo final do projecto enquanto meio adoptado para a extracção da informação pretendida. Daqui, a importância de concentrar esforços na sua implementação, até para proporcionar uma análise comparada das diferentes técnicas disponíveis para uma maior eficiência e rapidez no processo de tratamento de dados no futuro.

Ao optar pela segmentação, tiveram que se seleccionar, entre o variado leque de técnicas existentes, as que se pretendiam implementar. Foram então escolhidas técnicas de segmentação de imagens de RM até ao momento nunca utilizadas no estudo do tracto vocal: *Live Wire*, *Level Sets* e *Reaction Diffusion*. E também algumas já aplicadas à segmentação deste tipo de dados: manual e *Region Growing*. Ao tomar esta decisão, optou-se por uma

abordagem metodológica experimental, inovadora no estudo do tracto vocal. Implementou-se e fez-se um estudo e uma avaliação comparativa de algumas técnicas consideradas como o “estado da arte” no campo da segmentação. O método adoptado para a realização deste trabalho, nomeadamente as técnicas da segmentação implementadas, proporcionaram resultados bastante satisfatórios que permitem concluir a sua viabilidade para futuras análises inter-sujeito que levem à criação de bases de dados mais específicas, isto é, próprias para cada classe de sons.

O estudo comparativo das técnicas implementadas levou às seguintes conclusões: para as imagens 2D, o melhor método é o *Region Growing* porque leva a segmentações precisas, somente com a necessidade da escolha manual de uma semente; trabalhando com imagens de pior qualidade, o melhor método é o *Region Growing* após a aplicação do *Wavelet Denoising*; para a segmentação de volumes, o melhor método a aplicação de uma segmentação 2.5D com apoio em segmentações 2D através do *Live Wire* porque leva a segmentações bastante exactas.

O desenvolvimento do método 2.5D, suportou a aquisição volumétrica mais rápida experimentada no âmbito do projecto HERON (Martins, 2007). Sem este desenvolvimento, o método de aquisição não teria utilidade prática.

Este trabalho permitiu a obtenção de resultados (contornos, medidas de posições dos articuladores, VPOQ, funções de área) que permitiram estudos na área da fonética do PE como os apresentados na tese de mestrado da licenciada Paula Martins (Martins, 2007) e no artigo submetido à *Speech Communication* (Martins et al., 2007a).

É ainda de importância mencionar o estudo de um conjunto de técnicas de pré-processamento. Destas, foi seleccionado o *Wavelet Denoising* para o tratamento das imagens adquiridas em tempo real. Foram assim dados alguns primeiros passos no estudo das possibilidades e limitações dos vários métodos para o processamento (incluindo técnicas de atenuação de ruído e técnicas de segmentação) de imagens em tempo real relativas à produção do PE.

É de salientar, por fim, que o presente trabalho se insere num projecto de pesquisa bastante abrangente e pluridisciplinar, dentro do qual a tarefa descrita nesta dissertação foi focar e analisar exaustivamente o aspecto da utilização da segmentação de imagem aplicada a imagens RM do tracto vocal. Pensa-se ser oportuno realçar que desta investigação emergiram diferentes produtos, complementares entre eles. Entre outros assinala-se, como complemento ao presente trabalho, a investigação levada a cabo, contemporaneamente à presente, por Martins (2007), focando-se nos procedimentos de aquisição das imagens por RM. Martins (2007) fez ainda uma extensiva revisão de outros métodos de aquisição de dados, dos conceitos relacionados com a fonética e sons do Português Europeu (PE), e uma interpretação dos resultados obtidos (nomeadamente dos contornos) em termos fonéticos.

6.3 Trabalho futuro

O presente trabalho pode-se considerar como a “pedra miliar” (*milestone*) para o estudo das imagens RM do tracto vocal através de segmentação semi-automática. Antes de mais, no caso específico de estudos de fonética e do tracto vocal em geral, demonstrou-se a vantagem de complementar os dados EMMA com os resultados das segmentações de imagem de RM. Sendo assim, seria interessante alargar a base de dados de RM com a aquisição de imagens em mais informantes e com mais produções de cada fone.

Em relação às técnicas de atenuação de ruído, seria interessante experimentar outras técnicas que poderão vir a ser mais eficazes. Alguns exemplos a considerar podem ser: (Aubert e Vese, 1997; Schneider et al., 2000; Vese e Chan, 2001; Song, 2003). Dentro destes, é de notar que tanto em (Vese e Chan, 2001) como em (Song, 2003), são estudados métodos de atenuação de ruído baseados em *Level Sets*. Daqui, pode depreender-se que os métodos utilizados neste trabalho com a finalidade de segmentação de imagens, podem ser também usadas em outras áreas como a atenuação de ruído (e também *inpainting*). Para a avaliação do impacto dos métodos de remoção de ruído nas imagens, é também de considerar métricas baseadas no Sistema Visual Humano (SVH) (Peterson, 1990; Eskicioglu, 2000; Wang et al., 2004).

Por outro lado, no que diz respeito às técnicas existentes de segmentação de imagem, considera-se que um dos trabalhos mais importantes a desenvolver futuramente seria a implementação de técnicas que utilizem a resolução temporal das imagens de Tempo Real. Poder-se-ia, por exemplo, seguir a filosofia presente em (Nascimento e Marques, 2002) ou em (Cardoso et al., 2007).

Um outro factor importante é a automatização do processo de selecção dos parâmetros de cada método.

No que diz respeito ao tempo necessário para a implementação das técnicas de segmentação, é inegável a forte necessidade de otimizar as rotinas em termos temporais. Uma forma de aumentar a rapidez da segmentação seria estudar técnicas de inicialização que produzissem contornos iniciais mais próximos do corpo a segmentar.

Uma outra melhoria seria o aumento de resolução espacial de modo a tornar os voxels dos volumes em isotrópicos. Em vez da clássica interpolação aqui utilizada (que provoca algum esborratamento), poder-se-á utilizar, por exemplo, o método proposto em (Tsai et al., 2001), baseado no funcional de Mumford-Shah.

A experiência realizada, levou também a concluir que, no *Live Wire*, alguns possíveis desenvolvimentos futuros poderiam ser representados por:

1. Implementação de uma fase de treino (Barrett e Mortensen, 1997);
2. Estimação sub-pixel de contornos;
3. Extensão do domínio a sequências de imagens temporais (Tempo Real) ou espaciais (3D).

Uma sugestão para este terceiro ponto seria o cálculo dos custos locais através de *kernels* de convolução 3D (Mortensen e Barrett, 1998).

De forma mais geral, algumas metas a alcançar na área do processamento de imagens do tracto vocal, poderiam ser:

- Co-registo dos dentes, cuja exigência é sobretudo necessária no estudo das fricativas (utilizar, por exemplo, o método descrito em (Maciel e Costeira, 2003) ou o proposto em (Munim e Farag, 2007). Este ultimo não necessita de pontos correspondentes entre as formas a registar.);
- Detecção de contornos, ao longo do tempo, dos articuladores passivos e activos;
- Medição da abertura do tracto vocal;
- Estudo do tempo intergestual, isto é, do momento em que um gesto começa e acaba.

Saindo do campo da segmentação e passando à comparação entre contornos, seria interessante estudar as métricas aqui apresentadas (e outras) na quantificação da coarticulação.

6.4 Considerações finais

O presente estudo, idealizado no âmbito do projecto HERON para o prosseguimento dos fins nele propostos, teve uma função dupla. Se, por um lado, se proporcionaram as informações técnicas necessárias para dar seguimento ao projecto de elaboração de um sintetizador de fala, complementando as informações EMMA, por outro, esta investigação forneceu um valioso contributo metodológico à área do estudo do tracto vocal, através da implementação pioneira de técnicas de segmentação que, embora já existentes, foram pouco utilizadas para este fim. Os resultados apresentaram-se sob a forma de um avanço metodológico na área de investigação, pois demonstraram-se amplamente as vantagens ligadas à utilização das técnicas semi-automáticas na segmentação de imagens de RM do tracto vocal.

De forma geral, pode-se concluir que o grande contributo deste trabalho em que as metodologias e os protocolos adoptados se encontram fundamentados e apresentados, está na experimentação e na descrição comparada da funcionalidade e eficácia dos métodos de segmentação seleccionados. Os resultados das experiências conduzidas poderão servir de ponto de partida para estudos futuros sobre a segmentação de imagens de RM, sobretudo no que diz respeito ao processo de tomada de decisão na fase metodológica.

Bibliografía

- ABRANTES, A. J. e MARQUES, J. S. (1995a). “Exploiting the common structure of some edge linking algorithms”. In “International Conference on Image Processing (ICIP)”, volume 3, página 3624. IEEE Computer Societ, Washington DC, USA.
- (1995b). “Unified approach to snakes, elastic nets and Kohonen maps”. In “International Conference on Acoustics, Speech, and Signal Processing (ICASSP)”, volume 5, páginas 3427–3430. Detroit, MI, USA.
- (1996). “A class of constrained clustering algorithms for object boundary extraction”. *IEEE transactions on Image Processing*, 5(11):1507–1521.
- ADALSTEINSSON, D. e SETHIAN, J. A. (1994). “A fast level set method for propagating interfaces”. *Journal of computational physics*, 118(2):269–277.
- ADAMS, R. e BISCHOF, L. (1994). “Seeded region growing”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647.
- ALBORNOZ, E. M. (2007). “Inpainting and zooming using sparse representations”. doi: CursodeAnalisisyProcesamientoAvanzadodeSenales.
- ALWAN, A., NARAYANAN, S., e 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.
- APOSTOL, L., PERRIER, P., BACIU, M., SEGEBARTH, C., e BADIN, P. (2000). “Using the formant/cavity affiliation to study the inter-speaker variability: assessment from MRI data”. In “5th Seminar on Speech Production”, páginas 213–216. Kloster Seeon, Germany.
- APOSTOL, L., PERRIER, P., RAYBAUDI, M., e SEGEBARTH, C. (1999). “3D geometry of the vocal tract and inter-speaker variability”. In “XIVth International Congress of Phonetic Sciences (ICPhS)”, páginas 443–446. San Francisco, CA.
- ARNAL, A., BADIN, P., BROCK, G., CONNAN, P.-Y., FLORIG, E., PEREZ, N., PERRIER, P., SIMON, P., SOCK, R., VARIN, L., VAXELAIRE, B., e ZERLING, J.-P. (2000). “An X-ray database for French”. In “5th Seminar on Speech Production”, páginas 293–296. Kloster Seeon, Germany.

- ATKINS, M. S. e MACKIEWICH, B. T. (1998). “Fully automatic segmentation of the brain in MRI”. *IEEE transactions on medical imaging*, 17(1):98–107.
- AUBERT, G. e VESE, L. (1997). “A variational method in image recovery”. *SIAM Journal on Numerical Analysis*, 34(5):1948–1979.
- AVILA-GARCIA, M. S., CARTER, J. N., e DAMPER, R. I. (2004). “Extracting tongue shape dynamics from magnetic resonance image sequences (speech)”. In “International Conference on Signal Processing (ICSP)”, páginas 288–291. Istanbul, Turkey.
- BADIN, P., BAILLY, G., RAYBAUDI, M., e SEGEBARTH, C. (1998a). “A three-dimensional linear articulatory model based on MRI data”. In “5th International Conference on Spoken Language Processing (ICSLP)”, páginas 417–420. Sydney, Australia.
- BADIN, P., BOREL, P., BAILLY, G., REVÉRET, L., BACIUO, M., e SEGEBARTH, C. (2000). “Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images”. In “5th Seminar on Speech Production”, páginas 261–264. Kloster Seeon, Germany.
- BADIN, P., POUCHOY, L., BAILLY, G., RAYBAUDI, M., SEGEBARTH, C., LEBAS, J.-F., TIEDE, M., VATIKIOTIS-BATESON, E., e TOHKURA, Y. (1998b). “Un modele articuloire tridimensionnel du conduit vocal base sur des donnees IRM”. In “Journées d’Etude sur la Parole (JEP)”, páginas 283–286. Martigny, Valais, Suisse.
- BAER, T., GORE, J. C., GRACCO, L. C., e 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.
- BAILLARD, C., HELLIER, P., e BARILLOT, C. (2000). “Segmentation of 3D brain structures using level sets and dense registration”. In “IEEE Workshop on Mathematical Methods in Biomedical Image Analysis”, páginas 94–101. Hilton Head Island, SC.
- BARRETT, W. A. e MORTENSEN, E. N. (1997). “Interactive live-wire boundary extraction”. *Medical Image Analysis*, 1(4):331–341.
- BEHREND, J., HOOLE, P., LEINSINGER, G., TILLMANN, H. G., HAHN, K., REISER, M., e WISMÜLLER, A. (2003). “A segmentation and analysis method for MRI data of the human vocal tract”. In T. W. Meinzer, P. Hastreiter, U. Hoppe, H. Handels, A. Horsch, e Hans-Peter, editores, “CEUR Workshop”, volume 80, páginas 186–190. CEUR-WS.org, Melbourne, Florida, USA.
- BLINK, E. J. (2004). “An easy introduction: basic MRI physics for anyone who has not a degree in physics”.
- BRESCH, E., ADAMS, J., POUZET, A., LEE, S., BYRD, D., e NARAYANAN, S. (2006). “Semi-automatic processing of real-time MR image sequences for speech production studies”. In “7th Seminar on Speech Production”, páginas 427–434. Ubatuba, Brazil.

- BROWN, M. A. e SEMELKA, R. C. (2003). *MRI: basic principles and applications*. Wiley-Liss.
- CARBONE, I., MARTINS, P., SILVA, A., e TEIXEIRA, A. (2007a). “Volumetric MRI acquisition and processing”. In “154th Meeting of the Acoustical Society of America”, volume 122, página 3031. New Orleans, Louisiana.
- CARBONE, I., MARTINS, P., TEIXEIRA, A., e SILVA, A. (2006). *Base de dados RM dos sons do Português Europeu*. Relatório técnico, Universidade de Aveiro.
- (2007b). “Vocal tract segmentation and analysis over a European Portuguese MRI database”. *Revista periódica do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro (submetido)*.
- CARDOSO, J. S., CARDOSO, J. C. S., e CORTE-REAL, L. (2007). “Object-based spatial segmentation of video guided by depth and motion information”. In “Workshop on Motion and Video Computing (WMVC)”, página 7. Austin, Texas.
- CARROLL, T., SAKAIE, K., WIELOPOLSKI, P., e EDELMAN, R. (2006). “Advanced imaging techniques, including fast imaging”. In M. Z. . J. C. R. Edelman, J. Hesselink, editor, “Clinical magnetic resonance imaging”, volume 1. Philadelphia: Elsevier.
- CATES, J. E., LEFOHN, A. E., e WHITAKER, R. T. (2004). “GIST: an interactive, GPU-based level set segmentation tool for 3D medical images”. *Medical Image Analysis*, páginas 217–231.
- CHODOROWSKI, A., MATTSSON, U., LANGILLE, M., e HAMARNEH, G. (2005). “Color lesion boundary detection using live wire”. páginas 1589–1596. The International Society for Optical Engineering (SPIE), Bellingham, WA.
- CHOPP, D. L. (2007). “Another look at velocity extensions in the level set method”. *submitted to SIAM Journal of Scientific Computing*.
- CIOFOLO, C., BARILLOT, C., e HELLIER, P. (2004). “Combining fuzzy logic and level set methods for 3D MRI brain segmentation”. In “IEEE International Symposium on Biomedical Imaging: Macro to Nano”, volume 1, páginas 161–164. Arlington, USA.
- COHEN, L. D. e COHEN, I. (1993). “Finite-element methods for active contour models and balloons for 2-D and 3-D images”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147.
- COIFMAN, R. e DONOHO, D. (1995). “Translation invariant de-noising”. *Lecture Notes in Statistics*, 103:125–150.
- CRARY, M. A., GORHAM, M., BURTON, S., KOTZUR, I. M., e GAUER, J. (1996). “Dynamic magnetic resonance imaging in the study of vocal tract configuration”. *Journal of Voice*, 10(4):378–388.

- DANG, J. e HONDA, K. (1997). “Acoustic characteristics of the piriform fossa in models and humans”. *Journal of the Acoustical Society of America (JASA)*, 101(1):456–465.
- DANG, J., SHADLE, C. H., KAWANISHI, Y., HONDA, K., e SUZUKI, H. (1998). “An experimental study of the open end correction coefficient for side branches within an acoustic tube”. *Journal of the Acoustical Society of America (JASA)*, 104(2):1075–1084.
- DECARLO, D. e METAXAS, D. (1998). “Shape evolution with structural and topological changes using blending”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 20(11):1186–1205.
- DEMOLIN, D., DELVAUX, V., METENS, T., e SOQUET, A. (2003). “Determination of velum opening for French nasal vowels by magnetic resonance imaging”. *Journal of Voice*, 17(4):454–467.
- DEMOLIN, D., LECUIT, V., METENS, T., NAZARIAN, B., e SOQUET, A. (1998). “Magnetic resonance measurements of the velum port opening”. In “5th International Conference on Spoken Language Processing (ICSLP)”, páginas 425–429. Sydney, Australia.
- DEMOLIN, D., METENS, T., e SOQUET, A. (1996). “Three-dimensional measurement of the vocal tract by MRI”. In “4th International Conference on Spoken Language Processing (ICSLP)”, volume 1, página 272. Philadelphia.
- (2000). “Real time MRI and articulatory coordinations in vowels”. In “5th Seminar on Speech Production”, páginas 86–93. Kloster Seeon, Germany.
- DING, L. e GOSHTASBY, A. (2001). “On the Canny edge detector”. *Pattern Recognition*, 134(3):721–725.
- DZYUBACHYK, O., NIESSEN, W., e MEIJERING, E. (2007). “A variational model for level-set based cell tracking in time-lapse fluorescence microscopy images”. In “IEEE International Symposium on Biomedical Imaging: From Nano to Macro”, páginas 97–100. Arlington, VA, USA.
- EL-ZEHIRY, N., FAHMI, R., e FARAG, A. A. (2005). “Level set methods in medical imaging: an overview”. In “First international conference on Graphics, Vision, and Image Processing (GVIP)”, The Cairo, Egypt.
- ELSTER, A. D. e BURDETTE, J. H. (2001). *Questions and answers in magnetic resonance imaging*. Mosby.
- ENGWALL, O. (2000). “A 3D tongue model based on MRI data”. In “6th International Conference on Spoken Language Processing (ICSLP)”, volume III, páginas 901–904. Beijing, China.
- (2001). “Making the tongue model talk: merging MRI & EMA measurements”. In “7th European Conference on Speech Communication and Technology (Eurospeech)”, páginas 261–264. Scandinavia.

- (2003a). “Combining MRI, EMA and EPG measurements in a three-dimensional tongue model”. *Speech Communication*, 41(2-3):303–329.
- (2003b). “A revisit to the application of MRI to the analysis of speech production - testing our assumptions”. In “6th Seminar on Speech Production”, páginas 43–48. Sydney.
- (2004). “From real-time MRI to 3D tongue movements”. In “8th International Conference on Spoken Language Processing (ICSLP)”, páginas 1109–1112. Jeju Island, Korea.
- ENGWALL, O. e BADIN, P. (1999). *Collecting and analysing two and three dimensional MRI data for Swedish*. Quarterly progress and status report (stl-qpsr), Speech transmission Laboratory.
- (2000). “An MRI study of Swedish fricatives: coarticulatory effects”. In “5th Seminar on Speech Production”, páginas 297–300. Kloster Seeon, Germany.
- ENGWALL, O., DELVAUX, V., e METENS, T. (2006). “Interspeaker variation in the articulation of French nasal vowels”. In “7th Seminar on Speech Production”, páginas 3–10. Ubatuba, Brazil.
- ERICSDOTTER, C. (2005). *Articulatory-acoustic relationships in Swedish vowel sounds*. Tese de Doutorado, Stockholm University.
- ESKICIOGLU, A. (2000). “Quality measurement for monochrome compressed images in the past 25 years”. In “International Conference on Acoustics, Speech and Signal Processing (ICASSP)”, páginas 1907–1910. Istanbul, Turkey.
- FALCÃO, A. X., UDUPA, J. K., e MIYAZAWA, F. K. (2000). “An ultra-fast user-steered image segmentation paradigm: live wire on the fly”. *IEEE transactions on medical imaging*, 19(1):55–62.
- FROMKIN, V. e RODMAN, R. (1993). *Introdução à Linguagem*. Almedina, Coimbra.
- FUA, P. e LECLERC, Y. G. (1988). “Model driven edge detection”. In “DARPA Image Understanding Workshop”, páginas 45–56. Cambridge, Massachusetts.
- GALLO, G., ZINGALE, A., e ZINGALE, R. (1997). “Detection of MRI brain contour using nonlinear anisotropic diffusion filter”. In “18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society”, volume 18, páginas 1062–1064. Amsterdam, Netherlands.
- GELAS, A., BERNARD, O., FRIBOULET, D., e PROST, R. (2007). “Compactly supported radial basis functions based collocation method for level-set evolution in image segmentation”. *IEEE transactions on Image Processing*, 16(7):1873–1887.
- GERIG, G., KÜBLER, O., KIKINIS, R., e JOLESZ, F. A. (1992). “Nonlinear anisotropic filtering of MRI data”. *IEEE transactions on medical imaging*, 11(2):221–232.

- GICK, B., KANG, A. M., e WHALEN, D. H. (2000). “MRI and X-ray evidence for commonality in the dorsal articulations of English vowels and liquids”. In “5th Seminar on Speech Production”, páginas 69–72. Kloster Seeon, Germany.
- (2002). “MRI evidence for commonality in the post-oral articulations of English vowels and liquids”. *Journal of Phonetics*, 30(3):357–371.
- GONZALEZ e WOODS (2002). *Digital image processing*. Prentice Hall, 2 edição.
- GREENSPAN, H., PELED, S., OZ, G., e KIRYATI, N. (2002). “MRI inter-slice reconstruction using super-resolution”. *Magnetic Resonance Imaging*, 20(5):437–446.
- GREENWOOD, A. R., GOODYEAR, C. C., e MARTIN, P. A. (1992). “Measurements of vocal tract shapes using magnetic resonance imaging”. In “IEE Communications, Speech & Vision”, volume 139, páginas 553–560.
- GREGIO, F. N. (2006). *Configuração do trato vocal supraglótico na produção das vogais do Português Brasileiro: dados de imagens de ressonância magnética*. Tese de Mestrado, Pontifícia Universidade Católica de São Paulo.
- HAN, X., XU, C., e PRINCE, J. L. (2003). “A topology preserving level set method for geometric deformable models”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 25(6):755–768.
- HARALICK, R. M. e SHAPIRO, L. G. (1992). *Computer and Robot Vision*, volume I. Addison-Wesley, Reading (MA).
- HASHEMI, R. H., BRADLEY, W. G., e LISANTI, C. J. (1997). *MRI: the basics*. Lippincott Williams & Wilkins, Philadelphia.
- HEINZ, J. M. e STEVENS, K. N. (1964). “On the derivation of area functions and acoustic spectra from cineradiographic films of speech”. *Journal of the Acoustical Society of America (JASA)*, 23:1037–1038.
- HELBING, M. e ORGLMEISTER, R. (1993). “Anisotropic filtering for detecting left ventricular borders in echocardiographic images”. In “Computers in Cardiology”, páginas 197–200. London, England.
- HENTEA, T. I. (1993). “Algorithm for automatic threshold determination for image segmentation”. volume 1, páginas 535–538. Vancouver.
- HOFFMAN, E. A., GNANAPRAKASAM, D., GUPTA, K. B., HOFORD, J. D., KUGELMASS, S. D., e KULAWIEC, R. S. (1992). “VIDA: an environment for multidimensional image display and analysis”. In G. D. Acharya RS, Cogswell CJ, editor, “Biomedical Image Processing and Three-dimensional Microscopy”, volume 1660, páginas 694–711. The International Society for Optical Engineering (SPIE), Newport Beach, Calif.

- HOOLE, P., WISMÜLLER, A., LEINSINGER, G., KROOS, C., GEUMANN, A., e INOUE, M. (2000). “Analysis of tongue configuration in multi-speaker, multi-volume MRI data”. In “5th Seminar on Speech Production”, páginas 157–160. Kloster Seeon, Germany.
- ICHIGAYA, A., KUROZUMI, M., HARA, N., NISHIDA, Y., e NAKASU, E. (2006). “A method of estimating coding PSNR using quantized DCT coefficients”. *IEEE transactions on Circuits and Systems for Video Technology*, 16(2):251–259.
- JOHNSON, C. R. e WEINSTEIN, D. M. (2006). “Biomedical computing and visualization”. In “Twenty-Ninth Australasian Computer Science Conference (ACSC)”, volume 48 de *Conferences in Research and Practice in Information Technology (CRPIT)*, páginas 3–10. Hobart, Australia.
- KASS, M., WITKIN, A., e TERZOPOULOS, D. (1987). “Snakes: active contour models”. *International Journal of Computer Vision*, páginas 321–331.
- 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.
- KORNPROBST, P., PEETERS, R., VIEVILLE, T., MALANDAIN, G., MIERISOVA, S., SUNAERT, S., FAUGERAS, O., e HECKE, P. V. (2002). *Superresolution in MRI and its influence in statistical analysis*. Relatório Técnico 4513.
- KRÖGER, B. J., WINKLER, R., MOOSHAMMER, C., e POMPINO-MARSCHALL, B. (2000). “Estimation of vocal tract area function from Magnetic Resonance Imaging: preliminary results”. In “5th Seminar on Speech Production”, páginas 333–336. Kloster Seeon, Germany.
- LEE, S.-H. e SEO, J. K. (2006). “Level set-based bimodal segmentation with stationary global minimum”. *IEEE transactions on Image Processing*, 15(9):2843–2852.
- LEFOHN, A. E., CATES, J., e WHITAKER, R. (2003a). “Interactive, GPU-based level sets for 3D segmentation”. In “Medical Image Computing and Computer Assisted Intervention (MICCAI)”, páginas 564–572. Montréal, Canada.
- LEFOHN, A. E., KNISS, J. M., HANSEN, C. D., e WHITAKER, R. T. (2003b). “A streaming narrowband algorithm: interactive computation and visualization of level sets”. *IEEE transactions on Visualization and Computer Graphics*, 10(4):422–433.
- LI, H., ELMOATAZ, A., FADILI, J. M., e RUAN, S. (2003). “An improved image segmentation approach based on level set and mathematical morphology”. In “Third International Symposium on Multispectral Image Processing and Pattern Recognition”, volume 5286, páginas 851–854. The International Society for Optical Engineering (SPIE), Beijing, China.
- LI, H. e YEZZI, A. (2007). “Local or global minima: flexible dual-front active contours”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 29(1):1–14.
- LIANG, J., MCINERNEY, T., e TERZOPOULOS, D. (2006). “United snakes”. *Medical Image Analysis*, 10:215–233.

- LIANG, Z.-P. e LAUTERBUR, P. C. (1999). *Principles of magnetic resonance imaging: a signal processing perspective*. Wiley-IEEE Press, New York.
- LIEBERMAN, P. (1968). “Primate vocalizations and Human Linguistic Ability”. *The Journal of the Acoustical Society of America*, 44(6):1574–1584.
- LIM, J. S. (1990). “Two-dimensional signal and image processing”. Prentice Hall signal processing series, páginas 469–476. Prentice Hall, Englewood Cliffs.
- LIN, P., ZHENG, C.-X., e YANG, Y. (2004a). “Model-based medical image segmentation: a level set approach”. In “Fifth World Congress on Intelligent Control and Automation (WCICA)”, volume 6, páginas 5541– 5544. Hangzhou, China.
- LIN, P., ZHENG, C.-X., YANG, Y., e GU, J.-W. (2004b). “Statistical model based on level set method for image segmentation”. In “Fourth International Conference on Computer and Information Technology (CIT)”, páginas 143–148. Wuhan, China.
- MACIEL, J. e COSTEIRA, J. P. (2003). “A global solution to sparse correspondence problems”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 25(2):187–199.
- MAEDA, S., BERGER, M.-O., ENGWALL, O., LAPRIE, Y., MARAGOS, P., POTARD, B., e SCHOENTGEN, J. (2006). *Acoustic-to-articulatory inversion: methods and acquisition of articulatory data*. Relatório Técnico ASPI/2006/D1/v1.0.
- MALLADI, R. e SETHIAN, J. A. (1995). “Image processing via level set curvature flow”. In “National Academy of Sciences”, volume 92, páginas 7046–7050. Washington, D.C.
- MALLADI, R., SETHIAN, J. A., e VEMURI, B. C. (1995). “Shape modeling with front propagation: a level set approach”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175.
- MARQUES, C. (2003). “Ressonância magnética”. In F. d. M. Gaivão, editor, “Imagiologia clínica: princípios e técnicas”, página 369. Serviço de Imagiologia - Hospitais da Universidade de Coimbra, 1 edição.
- MARTINS, P., CARBONE, I., PINTO, A., SILVA, A., e TEIXEIRA, A. (2007a). “European Portuguese MRI based speech production studies”. *submitted to Speech Communication*.
- MARTINS, P., CARBONE, I., SILVA, A., e TEIXEIRA, A. (2007b). “Coarticulation effects on European Portuguese: a first MRI study”. In “Workshop AFCEP”, páginas 67–70. Montpellier.
- (2007c). “MRI study of coarticulation in European Portuguese”. In “154th Meeting of the Acoustical Society of America”, volume 122, página 3031. New Orleans, Louisiana.
- (2007d). “An MRI study of European Portuguese nasals”. In “Interspeech”, páginas 58–61. Antwerp, BE.

- MARTINS, P., CARBONE, I., TEIXEIRA, A., e SILVA, A. (2006). *Ressonância Magnética estática e dinâmica aplicada ao estudo da Fonética Articulatória do Português Europeu (PE)*. Projecto de investigação.
- MARTINS, P. M. V. (2007). *Ressonância Magnética no Estudo da Produção do Português Europeu*. Tese de Mestrado, Universidade de Aveiro.
- MATHIAK, K., HERTRICH, I., KINCSES, W. E., KLOSE, U., ACKERMANN, H., e 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. e TERZOPOULOS, D. (1995). “A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis”. *Computerized Medical Imaging and Graphics*, 19(1):69–83.
- (1996). “Deformable models in medical image analysis: a survey”. *Medical Image Analysis*, 1(2):91–108.
- (2000). “T-snakes: topology adaptive snakes”. *Medical Image Analysis*, 4:73–91.
- MÁDY, K., SADER, R., ZIMMERMANN, A., HOOLE, P., BEER, A. J., e ZEILHOFER, H.-F. (2001). “Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction”. In R. K. H. P. P. v. L. B. Maassen, W. Hulstijn, editor, “4th International Speech Motor Conference”, Speech motor control in normal and disorderd speech, páginas 142–145. Nijmegen, the Netherlands.
- MÁDY, K., SADER, R., ZIMMERMANN, A., HOOLE, P., BEER, A. J., ZEILHOFER, H.-F., e HANNIG, C. (2002). “Assessment of consonant articulation in glossectomee speech by dynamic MRI”. In “7th International Conference on Spoken Language Processing (ICSLP)”, páginas 961–964. Denver, Colorado, USA.
- MENDONÇA, A. M., SILVA, J. A. D., e CAMPILHO, A. (2004). “Automatic delimitation of lung fields on chest radiographs”. In “IEEE International Symposium on Biomedical Imaging”, páginas 1287–1290. Arlington, VA, USA.
- MENDONÇA, T., MARCAL, A. R. S., VIEIRA, A., NASCIMENTO, J., SILVEIRA, M., MARQUES, J., e ROZEIRA, J. (2007). “Comparison of segmentation methods for automatic diagnosis of dermoscopy images”. In “29th Annual International Conference of the IEEE (EMBC)”, páginas 6572–6575. Lyon, France.
- MERRIMAN, B., BENCE, J., e OSHER, S. (1992). “Diffusion generated motion by mean curvature”. In T. J. E., editor, “Computational Crystal Growers Workshop”, páginas 73–83. Amer. Soc., Rhode Island.
- MISITI, M., MISITI, Y., OPPENHEIM, G., e POGGI, J.-M. (2007). “Wavelet toolbox user’s guide”.

- MITRA, A. (2005). “Restoration of noisy document images with an efficient bi-level adaptive thresholding”. *International Journal of Computational Intelligence*, 2(2):118–123.
- MOHAMMAD, M., MOORE, E., CARTER, J. N., SHADLE, C. H., e GUNN, S. R. (1997). “Using MRI to image the moving vocal tract during speech”. In “5th European Conference on Speech Communication and Technology (Eurospeech)”, volume 4, páginas 2027–2030. Rhodes, Grécia.
- MOHAMMAD, M. A.-J. (1999). *Dynamic measurements of speech articulators using MRI*. Tese de Doutorado, University of Southampton.
- MORTENSEN, E. N. e BARRETT, W. A. (1995). “Intelligent scissors for image composition”. In “Computer Graphics (SIGGRAPH)”, páginas 191–198. Los Angeles.
- (1998). “Interactive segmentation with intelligent scissors”. *Graphical Models Image Processing*, 60(5):349–384.
- MORTENSEN, E. N., MORSE, B. S., e BARRETT, W. A. (1992). “Adaptive boundary detection using “live-wire” two-dimensional dynamic programming”. In “Computers in Cardiology”, páginas 635–638. Durham, NC, USA.
- MOUTINHO, L. D. C. (2000). *Uma introdução ao estudo da Fonética e Fonologia do Português*. Plátano - Edições Técnicas.
- MUNIM, H. E. A. E. e FARAG, A. A. (2007). “Curve/surface representation and evolution using vector level sets with application to the shape-based segmentation problem”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 29(6):945–958.
- MURASE, K., KAWAKAMI, K., MOCHIZUKI, T., KIKUCHI, K., YAMAZAKI, Y., SHINOHARA, M., MIKI, H., e IKEZOE, J. (2001). “An anisotropic diffusion method for denoising dynamic susceptibility contrast-enhanced magnetic resonance images”. *Physics in Medicine & Biology*, 46(10):2713–2723.
- NARAYANAN, S., ALWAN, A., e HAKER, K. (1995). “An articulatory study of fricative consonants using MRI”. *Journal of the Acoustical Society of America (JASA)*, 98(3):1325–1347.
- (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., e ALWAN, A. (1996). “Liquids in Tamil”. In “4th International Conference on Spoken Language Processing (ICSLP)”, páginas 797–800. Philadelphia.
- NARAYANAN, S., NAYAK, K., LEE, S., e 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.

- NASCIMENTO, J. e MARQUES, J. S. (1999). “Estimação de superfícies tridimensionais com modelos activos”. In “Jornadas de Engenharia de Telecomunicações e Computadores”, volume II, páginas 21–26. Lisbon, Portugal.
- (2002). “Robust multi-model filter for shape tracking in the presence of outliers”. *Pattern Recognition*, 35:2711–2718.
- NASCIMENTO, J. C. e MARQUES, J. S. (2005). “Adaptive snakes using the EM algorithm”. *IEEE transactions on Image Processing*, 14(11):1678–1686.
- NASCIMENTO, J. C. M. P. D. (1998). *Modelos Deformáveis em estimação e seguimento de objectos bidimensionais*. Tese de Doutoramento, Universidade técnica de Lisboa.
- NASON, G. P. e SILVERMAN, B. (1995). “The stationary wavelet transform and some statistical applications”. *Lecture Notes in Statistics*, 103:281–299.
- OLABARRIAGA, S. D. (1999). *Human-Computer interaction for the segmentation of medical images*. Tese de Doutoramento, University of Amsterdam.
- OLIVEIRA, L. D. M. (2007). *Segmentação fuzzy de imagens e Vídeos*. Tese de Doutoramento, Universidade Federal do Rio Grande do Norte.
- OSHER, S. e SETHIAN, J. A. (1988). “Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulation”. *Journal of computational physics*, 79:12–49.
- PARAGIOS, N. e DERICHE, R. (2000). “Geodesic active contours and level sets for the detection and tracking of moving objects”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, páginas 266–280.
- PENEDO, M., LADO, M., TAHOCES, P., SOUTO, M., e VIDAL, J. (2006). “Effects of JPEG2000 data compression on an automated system for detecting clustered microcalcifications in digital mammograms”. *IEEE Transactions on Information Technology in Biomedicine*, 10(2):354–61.
- PERKELL, J. (1996). “Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling”. *Journal of Phonetics*, 24:3–22.
- PERONA, P. e MALIK, J. (1990). “Scale-space and edge detection using anisotropic diffusion”. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–638.
- PESQUET, J. C., KRIM, H., e CARFATAN, H. (1996). “Time-invariant orthonormal wavelet representations”. *IEEE Transactions on Signal Processing*, 44(8):1964–1970.
- PETERSON, H. A. (1990). *Image segmentation using human visual system properties with applications in image compression*. Tese de Doutoramento, Univ., West Lafayette, IN.
- PHAM, D. L., XU, C., e PRINCE, J. L. (2000). “A survey of current methods in medical image segmentation”. *Annual Review of Biomedical Engineering*, 2:315–337.

- PLUEMPITIWIRIYAWAJ, C., MOURA, J. M. F., WU, Y.-J. L., e HO, C. (2005). “STACS: new active contour scheme for cardiac MR image segmentation”. *IEEE transactions on medical imaging*, 24(5):593–603.
- POON, K., HAMARNEH, G., e ABUGHARBIEH, R. (2007). “Segmentation of complex objects with non-spherical topologies from volumetric medical images using 3D livewire”. In J. M. R. Josien P. W. Pluim, editor, “Medical Imaging: Image Processing”, volume 6512, páginas 6512–6531. The International Society for Optical Engineering (SPIE), San Diego-USA.
- PRATT, W. K. (1978). *Digital Image Processing*. Wiley-Interscience, New York, 3 edição.
- PRITCHARD, D. (2002). *Vocal tract visualisation*. EECE593 winter 2002 project report.
- RENDON, J. M., CAPISTRAN, M., e LARA, B. (2006). “A discrete reaction-diffusion operator for moving curves and edge detection”. In “Electronics, Robotics and Automotive Mechanics Conference (CERMA)”, páginas 24–29. Cuernavaca, Morelos, Mexico.
- RIBEIRO, C. M. (2003). “Produção de fala”.
- RUA, S. e FREITAS, D. (2006). “Morphological dynamic study of human vocal tract”. In “Computational Modelling of Objects Represented in Images (CompIMAGE)”, páginas 381–386. Coimbra, Portugal.
- RUA, S. M. (2007). *Estudo Morfológico-Dinâmico de Tracto Vocal Humano*. Tese de Mestrado, Universidade do Porto.
- SANTOS, B., FERREIRA, C., SILVA, J., SILVA, A., e TEIXEIRA, L. (2004). “Quantitative evaluation of a pulmonary contour segmentation algorithm in X-ray computed tomography images”. *Academic Radiology*, 11(8):868–878.
- SCHEFFLER, K. e LEHNHARDT, S. (2003). “Principles and applications of balanced SSFP techniques”. *European Radiology*, 13(11):2409–2418.
- SCHNEIDER, M., FIEGUTH, P., KARL, W., e WILLSKY, A. (2000). “Multiscale methods for the segmentation and reconstruction of signals and images”. *IEEE transactions on Image Processing*, 9(3):456–468.
- SERRURIER, A. e BADIN, P. (2005a). “A three-dimensional linear articulatory model of velum based on MRI data”. In “Interspeech - Eurospeech”, páginas 2161–2164. Lisbon, Portugal.
- (2005b). “Towards a 3D articulatory model of the velum based on MRI and CT images”. *ZAS Papers in Linguistics (ZASPIL)*, 40:195–211.
- SETHIAN, J. A. (1985). “Curvature and the evolution of fronts”. *Communications in Mathematical Physics*, 101(4):487–499.
- (1996). “Fast marching level set methods for three-dimensional photolithography development”. In G. E. Fuller, editor, “Optical Microlithography IX”, volume 2726, páginas

- 262–272. The International Society for Optical Engineering (SPIE), Santa Clara, California.
- (1997). “Level set methods: an act of violence - evolving interfaces in geometry, fluid mechanics, computer vision and materials sciences”. *American Scientist*, 85(3):254–263.
- SEZGIN, M. e SANKUR, B. (2004). “Survey over image thresholding techniques and quantitative performance evaluation”. *Journal of Electronic Imaging*, 13(1):146–165.
- SHADLE, C. H., MOHAMMAD, M. A.-J., CARTER, J. N., e JACKSON, P. J. B. (1999). “Multi-planar dynamic magnetic resonance imaging: new tools for speech research”. In “XIVth International Congress of Phonetic Sciences (ICPhS)”, páginas 623–626. San Francisco, CA.
- SHAH, S. (2007). “Segmenting biological particles in multispectral microscopy images”. In I. c. society, editor, “IEEE workshop on applications of computer vision (WACV)”, página 44. Austin, Texas, USA.
- SHECHTER, G., DECLERCK, J., ÖZTÜRK, C., e MCVEIGH, E. R. (1999). “Fast template based segmentation of cine cardiac MR”. In “7th Scientific Meeting & Exhibition of the International Society for Magnetic Resonance in Medicine”, página 480. Pennsylvania, USA.
- SHI, Y. e KARL, W. C. (2005a). *A fast implementation of the level set method without solving partial differential equations*. Technical Report ECE-2005-02, Department of Electrical and Computer Engineering, Boston University.
- (2005b). “A fast level set method without solving PDEs”. In “IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)”, volume II, páginas 97–100. Philadelphia, PA, USA.
- SIEMENS (2003). *Spins, magnets and resonances: an introduction to the basics of magnetic resonance*. Erlanger.
- SILVA, A. (2003). “Formação de imagem em ressonância magnética”.
- SILVA, J. A. D., CAMPILHO, A. J. C., e SANTOS, J. C. M. D. (1996). “3-D data acquisition and scene segmentation system”. In “13th International Conference on Pattern Recognition”, volume 3, páginas 563–567. Vienna.
- SILVA, J. S. S. (2005). *Segmentação Pulmonar em Estudos de Tomografia Axial Computorizada*. Tese de Doutorado, Universidade de Aveiro.
- SILVA, R. J. (2006). *JPEG and JPEG2000 compression standards*. Relatório técnico.
- SONG, B. (2003). *Topics in Variational PDE Image Segmentation, Inpainting and Denoising*. Tese de Doutorado, University of California.
- SONKA, M., HLAVAC, V., e BOYLE, R. (1999). *Image processing, analysis and machine vision*. Pacific Grove (CA), 2 edição.

- SOQUET, A., LECUIT, V., METENS, T., e 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.
- SOQUET, A., LECUIT, V., METENS, T., NAZARIAN, B., e DEMOLIN, D. (1998). “Segmentation of the airway from the surrounding tissues on magnetic resonance images: a comparative study”. In “5th International Conference on Spoken Language Processing (ICSLP)”, páginas 3083–3086. Sydney, Australia.
- STONE, M., DAVIS, E., DOUGLAS, A., AIVER, M., GULLAPALLI, R., LEVINE, W., e LUNDBERG, A. (2001). “Modeling tongue surface contours from cine-MRI images”. *Journal of speech, language, and hearing research (JSLHR)*, 44(5):1026–40.
- STONE, M., DICK, D., DOUGLAS, A. S., DAVIS, E. P., e OZTURK, C. (2000). “Modelling the internal tongue using principal strains”. In “5th Seminar on Speech Production”, páginas 133–136. Kloster Seeon, Seeon, Germany.
- STONE, M. e 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., TITZE, I. R., e HOFFMAN, E. A. (1996). “Vocal tract area functions from magnetic resonance imaging”. *Journal of the Acoustical Society of America (JASA)*, 100(1):537–554.
- SURI, J. S., LIU, L., SINGH, S., LAXMINARAYAN, S., ZENG, X., e REDEN, L. (2001). “Shape recovery algorithms using levels sets in 2D/3D medical imagery: a state-of-the-art review”. *IEEE transactions on Information Technology in Biomedicine*, 6(1):8–28.
- SZÉKELY, G., HUG, J., BAJKA, M., MANESTAR, M., GROSCURTH, P., e HALLER, U. (1998). “Anatomical model generation for laparoscopic surgery simulation”. In “The Second Visible Human Project Conference”, páginas 45–46. Bethesda, Maryland USA.
- TAI, X.-C., CHRISTIANSEN, O., LIN, P., e SKJÆLAAEN, I. (2007). “Image segmentation using some piecewise constant level set methods with MBO type of projection”. *International Journal of Computer Vision*, 73(1):61–76.
- TAKANO, S. e HONDA, K. (2007). “An MRI analysis of the extrinsic tongue muscles during vowel production”. *Speech Communication*, 49(1):49–58.
- TAKEMOTO, H., HONDA, K., MASAKI, S., SHIMADA, Y., e FUJIMOTO, I. (2003). “Measurement of temporal changes in vocal tract area function during a continuous vowel sequence using a 3D cine-MRI technique”. In “6th Seminar on Speech Production”, páginas 84–289. Sydney.
- TAKEMOTO, H., KITAMURA, T., NISHIMOTO, H., e HONDA, K. (2004). “A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions”. *Acoustical Science and Technology*, 25(6):468–474.

- TEIXEIRA, A., MOUTINHO, L., e COIMBRA, R. L. (2001). “Acerca das vogais nasais do Português Europeu”. *Revista da Universidade de Aveiro - Letras*, 18:241–274.
- TEIXEIRA, A. e VAZ, F. (1999). *SAP - Síntese Articulatoria do Português*. Relatório SAP 1/1999, IEETA, Universidade de Aveiro.
- (2001). “European Portuguese nasal vowels: an EMMA study”. In “7th European Conference on Speech Communication and Technology (EuroSpeech)”, volume 2, páginas 1483–1486. Scandinavia.
- TSAI, R., ANTHONY YEZZI, J., e WILLSKY, A. S. (2001). “Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification”. *IEEE transactions on Image Processing*, 10(8):1169–1186.
- TSAI, R. e OSHER, S. (2004). “Level set methods and their applications in image science”. *Communications in Mathematical Sciences*, 1(4):1–20.
- VESE, L. e CHAN, T. (2001). *A multiphase level set framework for image segmentation using the Mumford and Shah model*. CAM report 01-25, UCLA Department of Mathematics.
- VOCI, F., EIHO, S., SUGIMOTO, N., e SEKIBUCHI, H. (2004). “Estimating the gradient threshold in the Perona-Malik equation”. *IEEE Signal Processing Magazine*, 21(3):39–65.
- WANG, Z., BOVIK, A., SHEIKH, H., e SIMONCELLI, E. (2004). “Image quality assessment: from error visibility to structural similarity”. *IEEE transactions on Image Processing*, 13(4):600–612.
- WEISHAUP, D., KÖCHLI, V. D., e MARINCEK, B. (2003). *How does MRI work? An introduction to the physics and function of magnetic resonance imaging*. Berlin.
- WELLS, J. C. (1997). *SAMPA computer readable phonetic alphabet*, volume Part IV de *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, Berlin and New York.
- WHALEN, D. H., KANG, A. M., MAGEN, H. S., FULBRIGHT, R. K., e 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.
- WILKINSON, M. H. F., WIJBENGA, T., VRIES, G. D., e WESTENBERG, M. A. (2003). “Blood vessel segmentation using moving-window robust automatic threshold selection”. In “International Conference on Image Processing (ICIP)”, volume 2, páginas 1093–1096. Barcelona.
- WRIGHT, G. A. (1994). “Signal acquisition and processing for magnetic resonance imaging”. In 1994, editor, “IEEE International Conference on Image Processing”, volume 3, páginas 523–527. Austin, Texas.
- (1997). “Magnetic resonance imaging”. *IEEE Signal Processing Magazine*, 14(1):56–66.

- WU, Z., PAULSEN, K. D., e JOHN M. SULLIVAN, J. (2005). “Adaptive model initialization and deformation for automatic segmentation of T1-weighted brain MRI data”. *IEEE transactions on Biomedical Engineering*, 52(6):1128–1131.
- YANG, B. (1999). “Measurement and synthesis of the vocal tract of Korean monophthongs by MRI”. In “XIVth International Congress of Phonetic Sciences (ICPhS)”, páginas 2005–2008. San Francisco, CA.
- YE, Z., ZHENG, Z., YU, X., e NING, X. (2005). “Automatic threshold selection based on ant colony optimization algorithm”. volume 2, páginas 728–732. Beijing, China.