

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**

**Departamento de Engenharia Electrotécnica e de Computadores**

# **ESTUDO MORFOLÓGICO-DINÂMICO DO TRACTO VOCAL HUMANO**

SANDRA MOREIRA RUA

Licenciada em Radiologia

pela Escola Superior de Tecnologia da Saúde do Porto – IPP

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

Dissertação realizada sob a supervisão de

Prof. Doutor Diamantino Freitas

do Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

Porto, Setembro de 2006

## RESUMO

A produção da fala humana, desde sempre, tem suscitado interesse a nível morfológico e acústico, pois deste conhecimento, advém informação útil à compreensão de todos os mecanismos envolvidos e à construção de modelos articulatórios. Apesar dos aspectos anatómicos e fisiológicos serem comuns a todos os indivíduos, o mecanismo da produção da fala é complexo e individual. Exige, por isso, que a criação de modelos tridimensionais para caracterização individual seja a mais exacta possível. A imagem por ressonância magnética (IRM) é uma técnica valiosíssima, que tem sido utilizada em várias investigações no âmbito da produção da fala, permitindo o estudo sem carácter ionizante, de todo o tracto vocal (avaliação morfológica), a aquisição de imagens multiplanares, com elevada resolução de contraste tecidual, e possibilitando o cálculo de funções de área úteis à compreensão do mecanismo da sua produção. Os estudos dinâmicos estiveram muito condicionados às limitações dos equipamentos, no entanto actualmente com o desenvolvimento de equipamentos com campos magnéticos mais elevados, tem vindo a ser possível alguma análise dinâmica. Neste sentido, o principal objectivo deste estudo foi a combinação da caracterização morfológica e dinâmica da maioria dos sons da fala do Português, culminando na criação de modelos tridimensionais e modelos dinâmicos bidimensionais. Este estudo consistiu, numa primeira fase, num estudo estático (conjunto de imagens obtidas durante a sustentação artificial de vinte e cinco sons do Português), e posteriormente, num estudo dinâmico (através da repetição sucessiva de três sílabas em contexto consoante-vogal (/tu/, /ma/ e /pa/) durante a aquisição). Da análise das stacks de imagem foi possível a extracção de contornos do tracto vocal e conseqüentemente a reconstrução 3D, não-convencional, através da combinação de stacks ortogonais, permitindo a visualização e medição parcial da forma do tracto vocal.

### **PALAVRAS-CHAVE:**

IRM do Tracto Vocal, Imagem, por Ressonância Magnética, Produção da fala, Estudo Morfológico, Modelo tridimensional do Tracto Vocal.

## **ABSTRACT**

The human speech production process have always been a subject of interest, concerning morphological knowledge and speech acoustics, aiming to reach a useful understanding and modelling of all mechanisms involved. The main anatomic aspects and physiology of the vocal tract are common to all people; however speech production is a complex and individual mechanism. This recommends that modelling be done with enough accuracy for individual characterization. A significant number of speech researchers have been using magnetic resonance imaging (MRI), that is a powerful tool for the study with enough safety, of the whole vocal tract (morphological measurement), to get multiplanar, high quality imaging of soft tissue, allowing the calculation of area functions for the understanding of the speech production mechanisms. More rarely dynamical studies have been undertaken to date, because of device limitations. Presently with high magnetic field value units, some dynamic analysis becomes feasible. Therefore, the aim of this study was a combination of the morphological and dynamical characterizations of speech, resulting in the production of three-dimensional models and dynamic 2D models of some relevant Portuguese sounds and syllables. This study consisted firstly of a static phase (a set of images collected during artificially sustained articulations of twenty five Portuguese sounds), and secondly of a dynamic one (several repetitions of sequences of three consonant-vowel syllables (/tu/, /ma/ and /pa/) during the acquisition). With the analysis of image stacks, the extraction of the vocal tract contours and a subsequent non-conventional 3D reconstruction was possible, by means of combination of orthogonal stacks, allowing the visualization and partial measurement of the vocal tract shape.

### **KEYWORDS:**

MRI of Vocal Tract, Magnetic Resonance Imaging, Speech Production, Morphological Study, Three-dimensional Vocal Tract Model.

## RÉSUMÉ

La production de la parole humaine, a toujours été un sujet d'intérêt, à niveau morphologique et acoustique, car de cette connaissance, parvient l'information utile à la compréhension et la modélisation de tous les phénomènes impliqués. Les aspects anatomiques et la physiologie du conduit vocal sont communs à toutes les personnes; cependant la production de la parole est un mécanisme complexe et individuel. Il exige, donc, que la création des modèles tridimensionnels pour que caractérisation individuelle soit la plus exacte possible. L'imagerie par résonance magnétique (IRM) est une technique inestimable, énormément employé en études de la parole, permettant l'étude avec assez de sécurité, de la totalité du conduit vocal (évaluation morphologique), en quelque plan, avec élevée résolution de contraste tissulaire, et en permettant le calcul des fonctions d'aire utiles à la compréhension du mécanisme de production de la parole. Les études dynamiques ont été conditionnées aux limitations des scanners, cependant, actuellement avec les champs magnétiques plus élevés, certain analyse dynamique est possible. Dans ce sens, le principal objectif de cette étude a été la combinaison des caractérisations morphologiques et dynamiques de la parole, pour le développement des modèles tridimensionnels et des modèles dynamic bidimensionnels de plusieurs articulations du Portugais. Cette étude a consisté, au début, dans une étude statique (ensemble d'images acquise pendant la sustentation artificielle de vingt-cinq articulations du Portugais), et deuxièmement, dans une étude dynamique (plusieurs répétitions de trois syllabes de consonne-voyelle (/tu/, /ma/ et /pa/) pendant l'acquisition). L'analyse des stacks d'images a permis l'extraction des contours du conduit vocal et subséquemment la reconstruction 3D, non-conventionnelle, par la combinaison de stacks perpendiculaires, possibilitant la visualisation e mesure partiel de la configuration du conduit vocal.

### MOTS-CLÉS :

IRM du Conduit Vocal, Imagerie par Résonance Magnétique, Production de la Parole, Étude Morphologique, Modèle Tridimensionnel du Conduit Vocal.

## PREFÁCIO

A fala é uma ferramenta singular de comunicação, que envolve um conjunto de órgãos que trabalham sincronamente na produção de sons. Os mecanismos envolvidos são complexos, sendo várias as técnicas de estudo da fala, quer do ponto de vista acústico, quer do ponto de vista morfológico. A IRM tem revelado potencialidades únicas no estudo do tracto vocal, das quais se destacam a sua inocuidade, a elevada resolução, o contraste e a capacidade multiplanar. Actualmente, com a informação tridimensional proveniente da IRM é possível a construção de modelos articulatórios para síntese da fala. Apesar dos vários estudos encontrados, existem ainda algumas limitações da IRM por superar, e portanto, a informação morfológica e dinâmica é ainda escassa.

Neste sentido, o tema deste trabalho versa o estudo morfológico-dinâmico do tracto vocal, para o Português Europeu (PE) usando a IRM. O presente trabalho está organizado em duas partes e vários capítulos, descritos mais adiante, cujo principal objectivo foi o exercício da técnica de IRM (do específico interesse profissional da autora) para a construção de modelos morfológico-dinâmicos destinados ao estudo e ao processamento da fala do PE.

Na realização do estudo experimental encontraram-se algumas dificuldades, relacionadas por um lado, pela inexistência de protocolos específicos para estudo do tracto vocal por IRM, e por outro, com limitações do próprio equipamento de ressonância magnética (quer em termos de funcionamento quer na programação das sequências). De facto, a bibliografia encontrada sobre o tema é pouco consensual quanto ao protocolo a aplicar, não só por limitação dos equipamentos existentes, mas também porque está condicionado ao tipo de estudo (estático ou dinâmico), ao corpus, e à finalidade do estudo. Isto é particularmente crítico, nos estudos dinâmicos por IRM, dado que apesar do desenvolvimento das sequências ultra-rápidas, nem todas são aplicáveis ao tracto vocal, pois são muito mais susceptíveis à produção de artefactos, que nesta região são abundantes devido ao movimento dos articuladores e ao fluxo do ar. Para além disso, não foi possível efectuar o registo acústico do sinal da fala, devido ao elevado ruído presente na sala, resultante do funcionamento do equipamento.

Acrescem ainda outras condicionantes relacionadas com os sujeitos, nomeadamente o esforço exigido durante a sustentação da fonação, a necessidade de imobilização da cabeça, e a postura corporal imposta pelo equipamento e a variabilidade dos próprios sujeitos na produção dos sons que influenciaram os resultados deste estudo – modelos tridimensionais. Por superar, está ainda o moroso e difícil processo de segmentação das imagens, assim como a finalização completa

dos modelos tridimensionais. A despeito destas dificuldades fica a convicção que o caminho agora trilhado neste trabalho tem um enorme potencial já na actualidade e no futuro crescerá a sua aplicação.

## **AGRADECIMENTOS**

Em primeiro lugar, agradeço ao meu orientador pela forma como orientou este trabalho, simpatia, e pelo apoio e incentivo que me prestou em todas as fases.

Agradeço à Professora Doutora Isabel Ramos do Serviço de Radiologia do Hospital S. João pela disponibilização de meios, bem como pela sua colaboração. Agradeço igualmente a toda a equipa técnica da unidade de RM deste hospital, pela colaboração e ajuda incansáveis. Ao Vítor, Sofia, Susana, Cláudia, Sandrine e Júlio, obrigada a todos.

Ao Luís Jacinto, colega do MEB, pela sua colaboração e ajuda, e pela iniciativa na construção de um molde do tracto vocal.

Ao João, à Cátia, à Maura, e ao André por toda a paciência, disponibilidade e colaboração prestadas, bem como pelas sugestões na área de Terapia da Fala.

À minha amiga Andreia, pelo apoio, ajuda, carinho e incentivo sempre constantes.

Dos meus colegas da ESTSP, agradeço ao Professor Doutor Rubim Santos pelos sábios conselhos, sugestões, e apoio constantes em todas as fases, à Dra. Maria João Cunha pela leitura e sugestões da versão preliminar da tese, especialmente na área de Terapia da Fala, assim como pelo fornecimento de material bibliográfico da área, e à Dra. Luísa Nogueira pelo apoio e incentivo.

Em especial, ao João Paulo pelo amor, ajuda e incentivo incondicional, obrigada por fazeres parte da minha vida.

E ao meu pai, mãe e irmão pelo permanente apoio e carinho.

## ÍNDICE

ÍNDICE DE FIGURAS	10
ÍNDICE DE QUADROS	12
LISTA DE ABREVIATURAS E SÍMBOLOS	13
GLOSSÁRIO	14
<b>INTRODUÇÃO GERAL</b>	<b>17</b>
<b>PARTE 1. ESTUDO DE REVISÃO BIBLIOGRÁFICA</b>	<b>20</b>
<hr/>	
<b>CAPÍTULO 1. REVISÃO DA LITERATURA</b>	<b>21</b>
1.1 UTILIZAÇÃO DA IRM NO ESTUDO DA PRODUÇÃO DA FALA	21
1.2 EFEITO DA SUSTENTAÇÃO DOS SONS E DA POSTURA CORPORAL EM IRM	24
1.3 ESTUDO DA NASALIDADE	25
1.4 MODELOS ARTICULATÓRIOS TRIDIMENSIONAIS	25
1.5 BREVE CONCLUSÃO	27
<b>CAPÍTULO 2. A FALA HUMANA</b>	<b>29</b>
2.1 ÓRGÃOS DO TRACTO VOCAL HUMANO	30
2.1.1 CAVIDADE ORAL	31
2.1.2 CAVIDADE NASAL	32
2.1.3 CAVIDADE FARÍNGEA	32
2.1.4 CAVIDADE LARÍNGEA	33
2.2 MECANISMO DA PRODUÇÃO DA FALA	36
2.2.1 MODELAÇÃO DO AR EXPIRADO PELAS CORDAS VOCAIS	37
2.3 ACÚSTICA DA FALA	38
2.4 PRODUÇÃO E ACÚSTICA DA FALA – RELAÇÃO	39
2.4.1 RESPIRAÇÃO	39
2.4.2 FONAÇÃO	40
2.4.3 ARTICULAÇÃO	40
2.5 A AUDIÇÃO DA FALA	43
<b>CAPÍTULO 3. MEDIÇÃO E ANÁLISE DA PRODUÇÃO DA FALA</b>	<b>45</b>
3.1 TÉCNICAS DE ANÁLISE DO SINAL DA FALA	46
3.2 TÉCNICAS DE ANÁLISE MORFOLÓGICAS /FUNCIONAIS DO TRACTO VOCAL HUMANO	47

3.2.1	TÉCNICAS INDIRECTAS	48
3.2.2	TÉCNICAS DIRECTAS	48
3.2.3	OBSERVAÇÃO DA LÍNGUA	48
<b>CAPÍTULO 4. PROCESSAMENTO DA FALA</b>		<b>52</b>
4.1	BREVE RESENHA HISTÓRICA	52
4.2	SÍNTESE DA FALA	53
<b>CAPÍTULO 5. LÍNGUA PORTUGUESA</b>		<b>56</b>
5.1	FONÉTICA ARTICULATÓRIA	56
5.1.1	CONSTRIÇÕES NA FALA	57
5.2	FONÉTICA ACÚSTICA	59
5.2.1	TERMINOLOGIA PORTUGUESA PARA OS TRAÇOS	60
5.3	FONÉTICA PERCEPTIVA	61
5.4	BREVES NOÇÕES DE FONOLOGIA	62
5.5	NOMENCLATURA PARA AS ARTICULAÇÕES VOCAIS E SONS USADOS NA FALA	63
<b>CAPÍTULO 6. IMAGEM POR RESSONÂNCIA MAGNÉTICA</b>		<b>65</b>
6.1	PRINCÍPIOS FÍSICOS	65
6.1.1	SEQUÊNCIAS EM IRM	67
6.2	INSTRUMENTOS EM RESSONÂNCIA MAGNÉTICA	68
6.2.1	FORMAÇÃO DA IMAGEM	69
6.3	IMPORTÂNCIA DA IRM EM SÍNTESE ARTICULATÓRIA E EM MEDICINA	70
<b>PARTE 2. ESTUDO EXPERIMENTAL</b>		<b>71</b>
<hr/>		
<b>CAPÍTULO 1. INTRODUÇÃO</b>		<b>72</b>
1.1	OBJECTIVOS	72
1.2	MATERIAL E MÉTODOS	73
1.2.1	CARACTERIZAÇÃO DA AMOSTRA	73
1.2.2	CORPUS DA FALA	75
1.2.3	EQUIPAMENTO UTILIZADO E PROCEDIMENTOS GERAIS	76
1.2.4	TÉCNICAS DE ANÁLISE DE IMAGEM UTILIZADAS	77

<b>CAPÍTULO 2. FASE DE TREINO E EXPLORAÇÃO DA IRM</b>	<b>78</b>
2.1 ESTUDO ESTÁTICO	79
2.2 ESTUDO DINÂMICO	84
2.3 INTERPRETAÇÃO DAS IMAGENS OBTIDAS	87
<b>CAPÍTULO 3. FASE DE PESQUISA E EXPLORAÇÃO DAS TÉCNICAS DE PROCESSAMENTO DE IMAGEM</b>	<b>91</b>
3.1 PROGRAMAS DE PROCESSAMENTO DE IMAGEM	93
3.1.1 IMAGE J	95
3.1.1.1 PLUGIN 3D EDITING TOOL	96
3.2 BLENDER	98
<b>CAPÍTULO 4. BASE DE DADOS DE NATUREZA MORFOLÓGICA PARA CARACTERIZAÇÃO DO PORTUGUÊS EUROPEU</b>	<b>100</b>
4.1 MODELOS TRIDIMENSIONAIS	100
4.2 DISCUSSÃO DOS RESULTADOS	108
4.3 SIMULAÇÃO ACÚSTICA REAL DO TRACTO VOCAL BASEADA NA RECOLHA DE DADOS MORFOLÓGICOS DE IRM	110
<b>CONCLUSÃO E PERSPECTIVAS PARA O FUTURO</b>	<b>112</b>
<b>REFERÊNCIAS</b>	<b>116</b>
<b>ANEXOS</b>	<b>122</b>
ANEXO A. INTERNATIONAL PHONETIC ALPHABET	123
ANEXO B. SPEECH ASSESSMENT AETHODS PHONETIC ALPHABET	125
ANEXO C. PROTOCOLO DE AVALIAÇÃO OROFACIAL	127
ANEXO D. FORMULÁRIO DE CONSENTIMENTO INFORMADO	144
ANEXO E. FOLHA DE REGISTO	147
ANEXO F. PROTOCOLO DO ESTUDO ESTÁTICO	149
ANEXO G. PROTOCOLO DO ESTUDO DINÂMICO	151
ANEXO H. CD COM ESTUDO ESTÁTICO E DINÂMICO	153
ANEXO I. FLUXOGRAMA CORRESPONDENTES ÀS ETAPAS DE COMBINAÇÃO E REPRESENTAÇÃO TRIDIMENSIONAL	155
ANEXO J. RELATÓRIO DE TRABALHOS PRÁTICOS DO INVESTIGADOR LUÍS JACINTO DO MEB	157
ANEXO K. BIBLIOGRAFIA	160

## ÍNDICE DE FIGURAS

FIGURA 1. ÓRGÃOS DO TRACTO VOCAL.	29
FIGURA 2. CORTE SAGITAL DA CAVIDADE ORAL.	31
FIGURA 3. PORÇÕES DA FARINGE E SUAS RELAÇÕES.	33
FIGURA 4. FUNÇÃO DA LARINGE NA DEGLUTIÇÃO E RESPIRAÇÃO.	33
FIGURA 5. CARTILAGENS DA LARINGE E RESPECTIVOS MÚSCULOS.	34
FIGURA 6. LARINGE E CORDAS VOCAIS.	35
FIGURA 7. REPRESENTAÇÃO ESQUEMÁTICA DA PRODUÇÃO DA FALA.	37
FIGURA 8. PRESSÃO SUBGLOTAL.	38
FIGURA 9. REPRESENTAÇÃO ESQUEMÁTICA DE UMA ARTICULAÇÃO NASAL E ORAL.	41
FIGURA 10. TEORIA FONTE-FILTRO – PRODUÇÃO DA FALA.	42
FIGURA 11. CONSTITUINTES DO OUVIDO.	43
FIGURA 12. TÉCNICAS DE ESTUDO DA PRODUÇÃO DA FALA.	45
FIGURA 13. ELECTROPALATOGRAFIA.	49
FIGURA 14. DELIMITAÇÃO DO TRACTO VOCAL OBTIDA POR CINE-RADIOGRAFIA.	50
FIGURA 15. RECONSTRUÇÃO A TRÊS DIMENSÕES DA FORMA DO TRACTO VOCAL, ATRAVÉS DA EBCT, DA VOGAL /A/	51
FIGURA 16. MÁQUINA FALANTE CONSTRUÍDA POR JOSEPH FABER.	52
FIGURA 17. VODER DE HOMER DUDLEY.	53
FIGURA 18. MECANISMO DE PRODUÇÃO DA FALA E SUA CORRESPONDÊNCIA COM UM MODELO COMPUTACIONAL.	54
FIGURA 19. PARÂMETROS CHAVE DO MODELO DO TRACTO VOCAL HUMANO PARA SÍNTESE ARTICULATÓRIA.	55
FIGURA 20. VERSÃO DO PROGRAMA DE SÍNTESE ARTICULATÓRIA – CASY.	55
FIGURA 21. VOGAIS DO PORTUGUÊS.	58
FIGURA 22. REPRESENTAÇÃO ACÚSTICA DO TRACTO VOCAL.	59
FIGURA 23. CLASSIFICAÇÃO DAS VOGAIS ORAIS DO PORTUGUÊS.	60
FIGURA 24. CLASSIFICAÇÃO DOS SONS DO PORTUGUÊS EUROPEU.	64
FIGURA 25. ESQUEMA REPRESENTATIVO DA OBTENÇÃO DO SINAL DE IRM.	66
FIGURA 26. LOCALIZAÇÃO ESPACIAL DO SINAL DE RM, E SUA CORRELAÇÃO COM OS PLANOS ANATÓMICOS.	69
FIGURA 27. ESTUDO SAGITAL COMPLETO DA VOGAL /U/ SUSTENTADA POR UM SUJEITO (SEXO FEMININO).	80
FIGURA 28. IMAGEM MÉDIOSAGITAL DE REFERÊNCIA, PARA IDENTIFICAÇÃO DA IMAGEM DOS DENTES (CONTORNO A BRANCO).	80
FIGURA 29. IMAGENS MÉDIOSAGITAIS DAS VOGAIS ORAIS DO PORTUGUÊS, SUSTENTADAS POR UM SUJEITO DO SEXO MASCULINO (ESQUERDA) E UM DO SEXO FEMININO (DIREITA).	82
FIGURA 30. STACK DE IMAGENS CORONAIS DA CONSOANTE /CH/.	83
FIGURA 31. IMAGENS MÉDIOSAGITAIS DAS VOGAIS NAIAS DO PORTUGUÊS, SUSTENTADAS POR UM SUJEITO DO SEXO MASCULINO (ACIMA) E UM DO SEXO FEMININO (ABAIXO).	84
FIGURA 32. GRÁFICO DE MONITORIZAÇÃO POR ECG E RESPECTIVO REGISTO DO CONJUNTO DE IMAGENS A ADQUIRIR MEDIANTE CADA INTERVALO R-R (CAIXAS A VERDE POR BAIXO DO TRAÇADO).	85
FIGURA 33. REPRESENTAÇÃO ESQUEMÁTICA DA TÉCNICA DE SINCRONIZAÇÃO POR ECG EM IRM PARA SUB-AMOSTRAGEM DE SEQUÊNCIAS.	86

FIGURA 34. EXTRACÇÃO DE CONTORNOS DAS IMAGENS MÉDIOSAGITAIS OBTIDAS NO ESTUDO DINÂMICO, PELA REPETIÇÃO DA SÍLABA /tu/.	87
FIGURA 35. REPRESENTAÇÃO ESQUEMÁTICA COM CAIXAS DA POSIÇÃO DAS STACKS COM BASE APENAS NOS VALORES DO ATRIBUTO <i>IMAGE POSITION</i> , NA VISTA LATERAL (A), VISTA DE FRENTE (B) E VISTA DE CIMA (C). A CAIXA VERDE PRETENDE REPRESENTAR A STACK CORONAL, E A CAIXA AZUL REPRESENTA A STACK SAGITAL.	89
FIGURA 36. REPRESENTAÇÃO ESQUEMÁTICA DA LOCALIZAÇÃO ESPACIAL DE CADA PLANO.	90
FIGURA 37. FUGA DE CONTORNOS USANDO A TÉCNICA SEMI-AUTOMÁTICA DE <i>THRESHOLD</i> .	92
FIGURA 38. JANELA DE IMAGEM E MENU DO SOFTWARE CTMRREDIT.	93
FIGURA 39. AMBIENTE DE TRABALHO E MENU DO SOFTWARE MRICRO.	95
FIGURA 40. AMBIENTE DE TRABALHO E MENU DO PROGRAMA IMAGEJ.	96
FIGURA 41. <i>OUTLINES</i> DA STACK SAGITAL (A) E CORONAL (B) DA VOGAL ABERTA /a/.	97
FIGURA 42. JANELA DA FERRAMENTA SEGMENTINGASSISTANT.	98
FIGURA 43. MODELO TRIDIMENSIONAL DA VOGAL /u/ OBTIDO PELA UNIÃO DAS <i>OUTLINES</i> SAGITAIS E CORONAIS.	99
FIGURA 44. MENU <i>TRANSFORM PROPERTIES</i> DO BLENDER USADO NA COMBINAÇÃO E REPRESENTAÇÃO TRIDIMENSIONAL DAS STACKS.	99
FIGURA 45. MODELO 3D CONVENCIONAL (A) E MODELO 3D NÃO CONVENCIONAL (B) EM PERSPECTIVAS DIFERENTES, DA VOGAL /u/ SUSTENTADA PELO MESMO SUJEITO (SEXO FEMININO).	100
FIGURA 46. IMAGEM EM TAMANHO REAL, DO MODELO TRIDIMENSIONAL DA VOGAL (A), DE UM SUJEITO DO SEXO FEMININO, COM INDICAÇÃO DAS SUAS DIMENSÕES E CORRELAÇÃO ANATÓMICA.	102
FIGURA 47. MODELOS TRIDIMENSIONAIS DAS CINCO VOGAIS ORAIS DO PORTUGUÊS PRODUZIDAS POR UM SUJEITO DO SEXO MASCULINO (A) E DO SEXO FEMININO (B).	103
FIGURA 48. MODELOS TRIDIMENSIONAIS DAS CINCO VOGAIS NASAIS DO PORTUGUÊS PRODUZIDAS POR UM SUJEITO DO SEXO MASCULINO (A) E DO SEXO FEMININO (B).	104
FIGURA 49. MODELOS TRIDIMENSIONAIS DAS QUATRO VOGAIS ORAIS DO PORTUGUÊS PRODUZIDAS POR UM SUJEITO DO SEXO MASCULINO (A) E DO SEXO FEMININO (B).	105
FIGURA 50. MODELOS TRIDIMENSIONAIS DAS CONSOANTES LATERAIS E OCLUSIVAS NASAIS DO PORTUGUÊS PRODUZIDAS POR UM SUJEITO DO SEXO MASCULINO (A) E DO SEXO FEMININO (B).	106
FIGURA 51. MODELOS TRIDIMENSIONAIS DAS CONSOANTES FRICATIVAS DO PORTUGUÊS PRODUZIDAS POR UM SUJEITO DO SEXO MASCULINO (A) E DO SEXO FEMININO (B).	107
FIGURA 52. VISTA SAGITAL DO ALINHAMENTO DAS PLACAS, PARA POSTERIOR CONSTRUÇÃO DO MOLDE FÍSICO REAL DO TRACTO VOCAL PARA A VOGAL /i/.	110
FIGURA 53. VISTAS FRONTAIS (A), VISTAS DA BASE (B), VISTA FRONTAL E DA BASE (C) E VISTA INTERIOR (D) DO MOLDE DO TRACTO VOCAL PARA A VOGAL /i/.	111

## ÍNDICE DE QUADROS

QUADRO 1. ÓRGÃOS INTERVENIENTES NA PRODUÇÃO DA FALA.	29
QUADRO 2. MÚSCULOS DA LARINGE E SUA FUNÇÃO.	36
QUADRO 3. CLASSIFICAÇÃO DAS VOGAIS E SEMIVOGAIS.	58
QUADRO 4. MODO E PONTO DE ARTICULAÇÃO DAS CONSOANTES.	59
QUADRO 5. HIPÓTESES POSSÍVEIS DE PROCESSAMENTO DE INFORMAÇÃO.	61
QUADRO 6. SEQUÊNCIAS USADAS EM IRM.	67
QUADRO 7. AVALIAÇÃO OROFACIAL APLICADA AOS SUJEITOS DA AMOSTRA.	74
QUADRO 8. ATRIBUTOS USADOS PARA A REPRESENTAÇÃO TRIDIMENSIONAL NO ESPAÇO DE CADA STACK.	89
QUADRO 9. LISTA DE PROGRAMAS DE RECONSTRUÇÃO TRIDIMENSIONAL.	94
QUADRO 10. NOMENCLATURA UTILIZADA NOS MODELOS 3D PARA CARACTERIZAÇÃO DOS SONS DO PE.	101

## LISTA DE ABREVIATURAS E SÍMBOLOS

<b>2D</b>	Bidimensional
<b>3D</b>	Tridimensional
<b>ASY</b>	Articulatory synthesizer
<b>Bmp</b>	Bitmap
<b>CASY</b>	Configurable articulatory synthesizer
<b>dB</b>	Decibel
<b>DICOM</b>	Digital Imaging and Communications in Medicine
<b>EBCT</b>	Tomografia computadorizada por feixe de electrões
<b>ECG</b>	Electrocardiograma
<b>EGG</b>	Electroglotografia
<b>EMMA</b>	Articulografia electromagnética
<b>EPG</b>	Electropalatografia
<b>ETL</b>	Factor turbo
<b>EUA</b>	Estados Unidos da América
<b>F<sub>0</sub></b>	Frequência fundamental
<b>F<sub>1</sub></b>	Primeira formante
<b>F<sub>2</sub></b>	Segunda formante
<b>FA</b>	Flip angles
<b>FDA</b>	Food and Drug Administration
<b>FID</b>	Free induction decay
<b>FOV</b>	Field of view
<b>IPA</b>	International Phonetic Alphabet
<b>IRM</b>	Imagem por ressonância magnética
<b>OPG</b>	Optopalatografia
<b>p.</b>	página
<b>PE</b>	Português Europeu
<b>PAOF</b>	Protocolo de avaliação orofacial
<b>ROI</b>	Região de interesse (do Inglês, <i>Region Of Interest</i> )
<b>RF</b>	Radiofrequência
<b>SAMPA</b>	Speech Assessment Methods Phonetic Alphabet
<b>SAR</b>	Limite de absorção de RF pelos tecidos
<b>T</b>	Tesla
<b>T/m</b>	Tesla por metro
<b>T/m/s</b>	Tesla por metro por segundo
<b>TAD</b>	Tempo de aquisição efectivamente disponível
<b>TD</b>	Trigger delay
<b>TE</b>	Tempo de eco
<b>TR</b>	Tempo de repetição
<b>TW</b>	Trigger window
<b>VME</b>	Vector de magnetização efectivo

## GLOSSÁRIO

**ANTERIOR** – EM ANATOMIA, ANTERIOR É UM DOS ADJECTIVOS FREQUENTEMENTE UTILIZADOS NA DESCRIÇÃO DO CORPO DOS SERES VIVOS. SIGNIFICA QUE DETERMINADA ZONA DO CORPO ESTÁ DIRIGIDA PARA A FRENTE, CONSIDERANDO-SE "FRENTE" COMO A DIRECÇÃO NORMAL DE MOVIMENTO DO SER VIVO.

**ARTEFACTO** – EM IMAGEM MÉDICA, SIGNIFICA TODO E QUALQUER OBJECTO PRESENTE NA IMAGEM MAS QUE NÃO PERTENCE AO OBJECTO EM ESTUDO, PREJUDICANDO A SUA OBSERVAÇÃO.

**ARTICULAÇÃO** – EM FONÉTICA, DENOMINAÇÃO DADA À FORMAÇÃO DE UMA CONSTRIÇÃO NO TRACTO VOCAL.

**ATRIBUTOS** – NO PROTOCOLO DICOM, ITENS QUE DESCREVEM QUE DESCREVEM ALGUMA COISA. SÃO USADOS PARA DESCREVER INFORMAÇÃO RELATIVA A UM OBJECTO.

**CAVIDADE DE RESSONÂNCIA** – EM FONÉTICA, O MESMO QUE CAVIDADE RESSONANTE. TERMO USADO PARA DESIGNAR O FENÓMENO DE RESSONÂNCIA (TRANSFERÊNCIA DE ENERGIA) NUMA DETERMINADA CAVIDADE, QUE PELA SUA FORMA E VOLUME PRODUZEM EFEITOS NOS SONS.

**COARTICULAÇÃO** – EM FONÉTICA, TERMO QUE DESIGNA A INFLUÊNCIA DE UM SOM SOBRE O OUTRO.

**CORPUS DA FALA** – CORPO DE LINGUAGEM, A FALA NESTE CASO EM PARTICULAR, PARA SER RECOLHIDO E ANALISADO PARA POSTERIOR UTILIZAÇÃO EM INVESTIGAÇÃO.

**CORTE** – EM IMAGEM MÉDICA, SIGNIFICA SECÇÃO, FATIA.

**CRESCIMENTO POR REGIÕES** – EM PROCESSAMENTO DE IMAGEM, É UM MÉTODO DE SEGMENTAÇÃO QUE CONSISTE EM AGRUPAR PIXELS COM CARACTERÍSTICAS SIMILARES FORMANDO REGIÕES HOMOGÉNEAS.

**DECÚBITO** – EM ANATOMIA, POSIÇÃO DEITADA DO CORPO.

**DICOM** – EM IMAGEM MÉDICA, É UM PROTOCOLO DE COMUNICAÇÃO PARA ENVIO E RECEPÇÃO DE IMAGENS E DADOS RELACIONADOS COM O PACIENTE.

**DICOM CONFORMANCE STATEMENT** – EM IMAGEM MÉDICA, DOCUMENTO QUE DESCREVE O CUMPRIMENTO DICOM PARA UM EQUIPAMENTO ESPECÍFICO.

**DORSAL** – DE "DORSO", EM ANATOMIA SIGNIFICA DE COSTAS.

**ELEMENTOS DE DADOS** – NO PROTOCOLO DICOM, SÃO ATRIBUTOS DESCRITIVOS QUE FORNECEM INFORMAÇÃO SOBRE AS CARACTERÍSTICAS DE DETERMINADA ENTIDADE.

**ESPECTROGRAMA** – SEQUÊNCIA DE ESPECTROS RESULTANTE DA ANÁLISE DA FALA NAS SUAS COMPONENTES DE FREQUÊNCIAS.

**ESTROBOSCOPIA** – EFEITO LUMINOSO INTERMITENTE GERALMENTE COM SUB-AMOSTRAGEM.

**ESTUDO DINÂMICO** – EM INVESTIGAÇÃO DA FALA, DESIGNA UM ESTUDO REALIZADO PARA A CRIAÇÃO DE IMAGENS A PARTIR DA ARTICULAÇÃO DOS SONS ACTIVA POR PARTE DOS SUJEITOS, NECESSITANDO PARA TAL DE SINCRONIZAÇÃO.

**ESTUDO EM TEMPO-REAL** – EM INVESTIGAÇÃO DA FALA, DESIGNA UM ESTUDO NO QUAL SE REALIZA A CAPTURA OU AQUISIÇÃO DIRECTA DE IMAGENS DO TRACTO VOCAL EM MOVIMENTO.

**ESTUDO ESTÁTICO** – EM INVESTIGAÇÃO DA FALA, DESIGNA UM ESTUDO NO QUAL AS IMAGENS SÃO ADQUIRIDAS DURANTE A SUSTENTAÇÃO DE UM SOM.

**FANTOMA** – EM IMAGEM MÉDICA, ESTE TERMO DESIGNA UM OBJECTO ANTROPOMÓRFICO (ISTO É, COM CARACTERÍSTICAS SIMILARES AOS TECIDOS DO CORPO HUMANO) USADO NA OBTENÇÃO DE IMAGENS PARA TESTAR OU CALIBRAR EQUIPAMENTOS, E TAMBÉM, NO CASO DA IRM, PARA MEDIÇÃO DO LIMITE DE ABSORÇÃO POR RF.

**FONEMA** – EM LINGUÍSTICA, UNIDADE DISTINTIVA MÍNIMA DE UMA LÍNGUA (MENOR UNIDADE SONORA).

**FORMANTES** – BANDAS DE RESSONÂNCIA (FREQUÊNCIAS) VISÍVEIS NO ESPECTROGRAMA PARA AS QUAIS A ENERGIA LOCAL É MÁXIMA.

**INTERPOLAÇÃO** – ESTIMATIVA DE VALORES DE ENTRE VALORES CONHECIDOS, PARA A REPRESENTAÇÃO DE UMA IMAGEM.

**MATRIZ** – CONJUNTO DE PIXELS DISTRIBUÍDOS EM COLUNAS E LINHAS. AS SUAS DIMENSÕES SÃO USADAS PARA DESCREVER A IMAGEM.

**METADADOS** – INFORMAÇÃO ACERCA DE DADOS. EM LINGUAGEM INFORMÁTICA, DEFINE UM CONJUNTO DE DADOS ESTRUTURADOS E CODIFICADOS.

**MODELOS ARTICULATÓRIOS** – MODELOS FÍSICO QUE PERMITE A PRODUÇÃO DA FALA.

**MODELAÇÃO** – EM FONÉTICA, SIGNIFICA MODIFICAÇÃO DA PASSAGEM DO FLUXO DE AR PELAS CORDAS VOCAIS.

**MODULAÇÃO** – EM FONÉTICA, SIGNIFICA MODIFICAÇÃO DAS PROPRIEDADES DE UM SOM, COMO A VARIAÇÃO DE AMPLITUDE, INTENSIDADE OU FREQUÊNCIA.

**MONEMA** – EM LINGUÍSTICA, SEGMENTO MÍNIMO SIGNIFICATIVO DA LÍNGUA, PODE SER LÉXICO (LEXEMA) OU GRAMATICAL (MORFEMA).

**OBJECTOS** – NO PROTOCOLO DICOM, DIZ RESPEITO A IMAGENS, RELATÓRIOS, OU QUALQUER OUTRA INFORMAÇÃO RELATIVA AO PACIENTE.

**OUTLINES** – CONJUNTO DE CONTORNOS BIDIMENSIONAIS OU GRUPO DE CURVAS QUE DEFINEM UM OBJECTO.

**“PIXEL”** – ELEMENTO DE IMAGEM, UNIDADE COM MENOR INFORMAÇÃO DE UMA IMAGEM (PICTURE ELEMENT).

**PLANO AXIAL** – EM ANATOMIA, PLANO SECCIONAL QUE DIVIDE O CORPO HUMANO EM PARTE SUPERIOR E PARTE INFERIOR. O MESMO QUE TRANSVERSAL, TRANSAXIAL.

**PLANO CORONAL** – EM ANATOMIA, PLANO SECCIONAL QUE DIVIDE O CORPO HUMANO EM PARTE ANTERIOR E PARTE POSTERIOR.

**PLANO MÉDIOSAGITAL** – EM ANATOMIA, PLANO SAGITAL LOCALIZADO NA LINHA MÉDIA DO CORPO HUMANO.

**PLANO SAGITAL** – EM ANATOMIA, PLANO SECCIONAL QUE DIVIDE O CORPO HUMANO EM PARTE DIREITA E PARTE ESQUERDA.

**POSTERIOR** – EM ANATOMIA, O CONTRÁRIO DE ANTERIOR.

**PROTOCOLO** – EM SENTIDO LATO, SIGNIFICA ALGO QUE SE PREDEFINE E QUE ESTÁ PRONTO A SER UTILIZADO, ATRAVÉS DE RECURSOS, OU AINDA, É A PADRONIZAÇÃO DE LEIS E PROCEDIMENTOS QUE SÃO DISPOSTOS A

EXECUÇÃO DE UMA DETERMINADA TAREFA. EM IMAGEM MÉDICA, SIGNIFICA UM CONJUNTO DE REGRAS QUE PERMITEM A COMUNICAÇÃO ENTRE DOIS COMPUTADORES OU EQUIPAMENTOS DE IMAGEM.

**PROTOCOLO DE EXAME** – CONJUNTO DE PARÂMETROS DE AQUISIÇÃO E PROCEDIMENTOS A USAR NA REALIZAÇÃO DE UM EXAME IMAGIOLÓGICO.

**REGISTO** – EM IMAGEM MÉDICA, SIGNIFICA ALINHAMENTO ESPACIAL DE IMAGENS DE ACORDO COM O MESMO EIXO DE REFERÊNCIA.

**SEGMENTAÇÃO** – EM PROCESSAMENTO DE IMAGEM, DESIGNA A SEPARAÇÃO DA IMAGEM NOS SEUS COMPONENTES QUE A CARACTERIZAM.

**SÍNTESE DA FALA** – GERAÇÃO DE UM SINAL ACÚSTICO ATRAVÉS DE UM COMPUTADOR.

**“SKIN”** – OBJECTO TRIDIMENSIONAL OBTIDO PELA UNIÃO DE OUTLINES.

**“STACK”** – CONJUNTO DE CORTES QUE REPRESENTAM UM VOLUME.

**“TAG”** – NO PROTOCOLO DICOM, É A DENOMINAÇÃO NUMÉRICA DE UM ATRIBUTO OU ELEMENTOS DE DADOS.

**“THRESHOLD”** – EM PROCESSAMENTO DE IMAGEM, É UM MÉTODO DE SEGMENTAÇÃO BASEADO NOS NÍVEIS DE CINZENTO (HISTOGRAMA) PARA A EXTRACÇÃO DE CONTORNOS.

**“TRIGGER”** – O MESMO QUE DISPARO, QUE PROVOCA O DESENCADear DE OUTRA ACÇÃO.

**VENTRAL** – DE "VENTRE", EM ANATOMIA SIGNIFICA ABDÓMEN.

## **INTRODUÇÃO GERAL**

A comunicação oral é o modo de comunicação mais familiar e frequentemente usado pelo Homem, e que resulta do trabalho organizado e sincronizado de um conjunto de órgãos que constituem o tracto vocal. A caracterização da produção da fala é possível fazer-se em três etapas distintas, a respiração (fonte de energia), a fonação (fonte de som) e a articulação (formação do som). A articulação resulta da actividade de um conjunto de órgãos articuladores que se movimentam, modificando a posição e a forma durante a passagem do ar, originando diferentes sons e conseqüentemente, diferentes representações acústicas. O seu entendimento é, por isso, amplamente estudado, com recurso a diferentes técnicas, mas ainda não totalmente conhecido, principalmente, no que diz respeito ao Português Europeu. A IRM, pelo elevado detalhe anatómico que fornece, destaca-se de outras técnicas de análise da fala, possibilitando o estudo morfológico e dinâmico, bidimensional e tridimensional do tracto vocal, com utilidade quer para a Medicina quer para a síntese da fala.

A multidisciplinaridade desta temática é marcante, envolvendo a Medicina (pelo estudo anatómico e funcional do tracto vocal), a Engenharia (através da acústica e processamento da fala), a Fonética (pelo estudo dos sons da fala, da sua produção e percepção), da Terapia da Fala (na avaliação dos aspectos anátomo-fisiológicos relacionados com perturbações da comunicação, linguagem e de fala), e a Imagiologia (pelo desenvolvimento da técnica de IRM aplicada ao estudo da fala).

Desta forma, o presente trabalho, organizado em duas partes, pretende, numa primeira parte, descrever de forma clara e sucinta, como estas disciplinas científicas se relacionam com a fala humana. No capítulo um, retrata-se o estado da arte no estudo da fala, seguindo-se a descrição dos aspectos anatómicos e fisiológicos do tracto vocal, no segundo capítulo, possibilitando ao leitor a familiarização com a terminologia anatómica e o entendimento dos mecanismos envolvidos. Dada a individualidade da fala humana são descritas, no capítulo três, as várias técnicas existentes para a medição e análise da sua produção, e que assentam fundamentalmente no estudo da actividade muscular, do movimento e forma dos articuladores, e na análise acústica do sinal da fala. No quinto capítulo é feita uma abordagem à Língua Portuguesa, não só através da descrição e classificação dos sons do PE (e nomenclatura usada), como da fonética acústica e perceptiva. A técnica de IRM é explorada no sexto capítulo, através da breve explicação dos seus princípios físicos e formação da imagem, e também é dada ênfase à importância que desempenha neste âmbito.

A segunda parte é destinada ao estudo experimental desenvolvido, explorando as técnicas e procedimentos executados para o efeito, bem como as dificuldades e limitações encontradas. Este estudo consistiu em duas fases distintas, inicialmente um estudo estático da maioria dos sons da

fala do PE para a caracterização morfológica, e posteriormente, um estudo dinâmico. São apresentados os modelos tridimensionais dos 25 sons estudados e modelos bidimensionais dos eventos dinâmicos na produção de três sílabas escolhidas para o efeito.

A IRM fornece informação morfológica útil e precisa acerca da posição e forma dos diferentes articuladores envolvidos na produção da fala, bem como da sua dinâmica, e com aplicações em diferentes áreas como a fonética articulatória e a síntese da fala.

**PARTE 1**

**ESTUDO DE REVISÃO**

**BIBLIOGRÁFICA**

## **CAPÍTULO 1. REVISÃO DA LITERATURA**

O tracto vocal permite a produção e a modulação dos sons – a fala, dotando o Homem de uma ferramenta singular de comunicação e que é comum a todos os seres humanos. No entanto, a produção dos sons é um processo complexo, sendo escassos os conhecimentos acerca das características anatómicas e funcionais do tracto vocal (Teixeira et al., 2001), e enorme a variabilidade na produção dos sons (Faria et al., 1996); ou seja, o mesmo indivíduo é praticamente incapaz de produzir um som várias vezes da mesma forma. Acresce ainda que, para além das diferenças anatómicas entre indivíduos, cada um adopta diferentes estratégias no controlo do tracto vocal, fazendo com que exista uma variabilidade do sinal acústico entre oradores.

Desta forma, desde muito cedo o estudo do tracto vocal despertou o interesse da comunidade científica, com carácter multidisciplinar, encontrando-se na literatura vários artigos (descritos adiante neste capítulo) que ilustram não só a diversidade de disciplinas envolvidas à volta desta temática, como também a especificidade de cada uma das abordagens.

Esta revisão da literatura visa, por um lado, a apresentação temporal dos diferentes estudos realizados para o estudo da produção da fala, com base na IRM, apontando os principais resultados e dificuldades e por outro, a descrição do estado da arte actual.

Também, neste capítulo, são descritos estudos relativos ao efeito da sustentação dos sons e da postural corporal em IRM (imposta pelo equipamento utilizado), dada a sua importante influência no estudo da produção da fala. A nasalidade é uma característica importante e complexa que tem suscitado o interesse de vários investigadores, e que será aqui também destacada. Por último, são apresentados alguns estudos relacionados com a construção de modelos tridimensionais, obtidos com base na IRM.

### **1.1 UTILIZAÇÃO DA IRM NO ESTUDO DA PRODUÇÃO DA FALA**

A compreensão da produção da fala seria impossível sem o conhecimento morfológico do tracto vocal bem como das suas relações com outros órgãos. Dada a escassez de informação morfológica têm sido realizadas inúmeras investigações usando técnicas directas e/ou indirectas no estudo dos mecanismos da produção da fala, revelando-se a Imagem por Ressonância Magnética (IRM) uma das mais promissoras. Para além da sua inocuidade, a IRM permite o estudo completo de todo o tracto vocal e suas relações, em cortes seccionais e também a três dimensões.

A primeira demonstração a três dimensões da forma do tracto vocal foi apresentada por Baer et al. (1991). Deste estudo impulsionador outros se repetiram; contudo, a informação morfológica ainda estava condicionada pelas limitações iniciais da técnica de IRM (uso de baixos campos magnéticos dado o elevado custo dos equipamentos, longos tempos de exame, imagem com muitos artefactos e com baixa resolução).

Os primeiros estudos descritos no estudo do tracto vocal por IRM baseavam-se na representação estática dos sons obrigando aos sujeitos à sustentação prolongada dos mesmos. Para além do grande esforço a que os indivíduos estavam submetidos, a informação morfológica ainda não era a desejada, pois a baixa resolução temporal da IRM que se verificava em detrimento da relação sinal-ruído e a presença de muitos artefactos de imagem, condicionava a maioria das investigações ao estudo das vogais orais e nasais (Dang et al., 1996; Demolin et al., 1996).

Rapidamente, e com o desenvolvimento de aquisições mais rápidas por IRM, foi possível a realização de estudos dinâmicos durante a produção da fala, através da repetição da mesma sequência de sons sincronizados com o equipamento, durante um certo período de tempo, e posterior reconstrução dos movimentos articulatorios.

Um estudo realizado por Di Girolamo et al. (1996), com o objectivo de avaliar por IRM os movimentos laríngeos e orofaríngeos durante a fala, concluíram que os cortes médiosagittais seriam a melhor modalidade no estudo dos diferentes padrões durante a fala, e que poderiam até mesmo substituir outras técnicas de imagem usadas até então (como a radiografia de perfil e cineradiografia) no estudo das vogais.

Com o objectivo de descrever um método para a obtenção da IRM do tracto vocal em movimento, Mohammad et al., (1997), realizaram um estudo que consistia numa primeira fase em aquisições estáticas para caracterização das vogais, consoantes fricativas, nasais, plosivas e líquidas e, numa segunda fase, em aquisições dinâmicas. Estas aquisições dinâmicas consistiam na obtenção de imagens seccionais que posteriormente eram sincronizadas com o sinal áudio e reconstruídas sem a utilização do processador do equipamento de IRM. Apesar dos longos tempos de exame constituírem a grande limitação da IRM, as suas vantagens são inequívocas, destacando-se a boa qualidade de imagem no estudo de tecidos moles e a sua capacidade multiplanar.

Vários estudos se seguiram até então, nomeadamente, estudos estáticos e/ou dinâmicos na caracterização morfológica das vogais em cortes médiosagittais, que vieram reforçar o futuro promissor da IRM neste âmbito, e nos quais já era possível a medição da forma do tracto vocal e a criação de modelos a três dimensões (Demolin et al., 1997; Engwall, 1999; Apostol et al., 1999; Soquet et al., 2002).

Badin et al. (1998) realizaram um estudo no qual um sujeito sustentava (durante 43s) uma vogal no início ou final de uma sequência vogal-consoante-vogal, com o objectivo de criar um modelo articulatório a três dimensões do tracto vocal com base na informação proveniente da IRM (em cortes sagitais, coronais e oblíquos). Este estudo demonstra a importância da informação tridimensional para a construção dos modelos articulatórios, pois estes constituem, segundo os autores, uma forma privilegiada no estudo dos fenómenos da produção da fala e em particular no seu controlo.

Mais tarde, num estudo dinâmico realizado por Shadle et al. (1999) é feita a descrição do princípio de funcionamento de aquisições dinâmicas por IRM, sendo estas apresentadas como nova ferramenta na investigação da produção da fala. Este estudo vem reforçar as vantagens desta técnica no estudo do tracto vocal, que segundo os autores, supera as suas limitações (longos tempos para aquisição das imagens e dificuldades no registo do sinal da fala a partir do intercomunicador devido ao elevado ruído produzido durante as aquisições).

Outros estudos se seguiram, tais como alguns estudos estáticos e dinâmicos para a caracterização de consoantes em diferentes contextos vogal-consoante ou vogal-consoante-vogal (Badin et al., 2000; Engwall, 2000a; Engwall, 2000b, Engwall et al., 2000; Mády et al., 2002) e também os primeiros estudos com aquisição directa do movimento dos articuladores envolvidos durante a fala, em tempo real (Engwall, 2001a; Engwall, 2001b; Stone et al., 2001; Demolin et al., 2002; Takemoto et al., 2003).

Demolin et al. (2000) compararam as distâncias médiosagitais obtidas num estudo estático (durante 12s) e depois em tempo real (durante 20s) de um sujeito, para as mesmas vogais. Desta análise, constataram que a IRM em tempo real fornecia informação fiável e exacta da posição dos articuladores envolvidos na produção da fala, apesar das dificuldades da recolha do sinal acústico pelo elevado ruído durante as sequências (não fornecendo qualidade suficiente para a sua segmentação). Segundo os autores, comparativamente a outras técnicas, a IRM em tempo real permite o estudo da deformação dinâmica do tracto vocal, em qualquer plano, permitindo obter novas informações e criando novas perspectivas para o estudo de processos de coarticulação.

Também segundo Mathiak et al. (2000), a IRM fornece imagens de todo o tracto vocal com boa resolução temporal e espacial, permitindo visualizar e quantificar aspectos dinâmicos dos diferentes articuladores durante a fala. Este estudo realça a importância do conhecimento morfológico e dinâmico do tracto vocal, no entendimento dos mecanismos subjacentes às patologias da fala e no desenvolvimento de algoritmos para a síntese da fala.

Mády et al. (2001) realçaram o uso da IRM em tempo real na avaliação da articulação das consoantes antes e após glossectomia parcial e reconstrução oral. Segundo os autores, esta técnica de imagem conjuntamente com a história clínica pode ser útil na identificação de factores com influência na diminuição da fala pós-cirurgia. O estudo articulatório e dos mecanismos compensatórios numa situação de alteração oral (patologia) não tem só relevância teórica, como também pode ser útil no fornecimento de linhas de orientação clínica para técnicas de reconstrução oral.

O surgimento de novas abordagens e o desenvolvimento de sequências ultra-rápidas por IRM permitiu a realização de mais estudos em tempo real (pela aquisição de imagens a cada 110ms), passando o futuro destes pela melhoria da qualidade de imagem e pela avaliação quantitativa da variação contínua da forma do tracto vocal (Narayanan et al., 2004).

Inicialmente a frequência de aquisição da IRM em tempo real dependia do compromisso entre o tempo e a resolução espacial. Actualmente, depende das alterações subjacentes ao desenvolvimento dos equipamentos de IRM e das sequências utilizadas, que contribuem para o aumento contínuo da resolução temporal (Engwall, 2004).

## **1.2 EFEITO DA SUSTENTAÇÃO DOS SONS E DA POSTURA CORPORAL EM IRM**

Um estudo realizado por Engwall (2003), demonstrou o efeito da sustentação dos sons durante longos períodos de tempo (como a realizada nos estudos estáticos por IRM), e qual a influência da posição de decúbito dorsal (deitado com apoio das costas) ou ventral (com apoio do ventre) na avaliação da produção da fala devido à força da gravidade. Segundo este, a sustentação acima dos 30s dos sons originava fenómenos de hiperarticulação (exagero na articulação), como já tinha sido demonstrado anteriormente pelo mesmo autor. Nas conclusões deste estudo, o autor recomenda que nos estudos estáticos por IRM se usem tempos de aquisição mais curtos (permitindo a sustentação dos sons) mas se possível, que seja utilizada a IRM em tempo real. Relativamente à posição de decúbito quer dorsal quer ventral, concluiu que ambas afectam a forma e posição da língua, especialmente se os sons forem sustentados.

A forma do tracto vocal durante a fala difere de acordo com a postura do sujeito (posição de decúbito ou em pé, por exemplo) devido à diferença de força da gravidade nas duas posições. Os estudos do tracto vocal por IRM estão condicionados pelos equipamentos, obrigando à realização dos mesmos na posição de decúbito. Os equipamentos de campo aberto, menos potentes, podem

minimizar este efeito da força da gravidade, restringindo, no entanto, a investigação ao estudo estático da produção da fala (sustentação dos sons). Num estudo realizado por Kitamura et al. (2005), foram comparadas imagens médiosagitais durante a articulação de vogais, em duas posições (sentado e decúbito), usando um equipamento de ressonância magnética de campo aberto. Os autores constataram que existe influência nos tecidos moles e estruturas ósseas, que se deformam de acordo com a força de gravidade. Nas estruturas ósseas, esta modificação deve-se também à mudança de postura corporal. A retracção da língua, na posição de decúbito, é mais marcante para a articulação das vogais posteriores que das anteriores.

### **1.3 ESTUDO DA NASALIDADE**

A nasalidade é uma característica complexa de alguns sons e que está relacionada com o abaixamento do véu do palato. Esta mudança de configuração introduzida pelos movimentos deste órgão, traduz-se em mudanças significativas no sinal acústico da fala, descritos em vários estudos quer do ponto de vista da sua produção (Dang et al., 1996; Demolin et al., 1998; Etena et al., 2002; Kane et al., 2002) quer da sua percepção.

Nesse sentido, Serrurier et al. (2005) descreveram modelos articulatórios bidimensionais e tridimensionais do véu do palato, na tentativa de explicar o possível mecanismo do movimento articulatório envolvido. Para tal, estudaram 46 sons do Francês (vogais orais e nasais e consoantes em três contextos diferentes) sustentados por um sujeito, tendo cada uma das articulações sido mantida durante 45 segundos. Os contornos do véu do palato foram extraídos manualmente, sendo estes posteriormente adaptados para a construção de modelos tridimensionais. Os autores constataram que a forma tridimensional descrita para 28 das articulações sustentadas, nas quais ocorria contacto entre a língua e o véu, poderia ser criada através de um modelo linear.

### **1.4 MODELOS ARTICULATÓRIOS TRIDIMENSIONAIS**

A construção de modelos articulatórios tridimensionais exige a utilização de técnicas de segmentação (extracção dos contornos do tracto vocal) e análise das imagens do tracto vocal provenientes da IRM, e sua posterior representação num sistema de coordenadas.

A IRM fornece informação acerca de toda a geometria do tracto vocal, possibilitando a delimitação das cavidades de passagem do ar dos restantes tecidos que as circundam, e a sua

localização. As tarefas de segmentação e localização têm causado problemas de diferentes ordens (por exemplo, dificuldade na identificação dos dentes), abordados desde cedo na literatura (Baer et al., 1991; Demolin et al., 1996).

Neste sentido, e pela variedade de técnicas de segmentação utilizadas, Soquet et al. (1998) elaboraram um estudo comparativo entre três métodos de segmentação usados em IRM: o método *threshold* (semi-automático) descrito por Baer et al. (1991) e Badin et al. (1998), o método manual descrito por Demolin et al. (1996), e o método elástico (colocação de uma curva na região de interesse e posterior coincidência). Segundo os autores, estes três métodos têm resultados comparáveis, o método *threshold* no entanto apresenta uma menor dispersão, e os parâmetros estabelecidos por cada método (como o contraste, limiar, tipo de curva) são factores que condicionam a área resultante no processo de segmentação.

Na literatura revista, são encontradas várias descrições de métodos de reconstrução tridimensional (Kroger et al., 2000), assim como várias técnicas de segmentação das imagens de ressonância magnética, cada uma com vantagens e desvantagens, pois esta tarefa, para além de árdua, exige ainda um grande esforço por parte dos investigadores.

Nesse sentido, Beherends et al. (2001) propuseram um método de segmentação e análise das imagens de ressonância magnética de forma a minimizar o tempo e esforço dos investigadores. Os autores descrevem, a utilização da técnica de crescimento por regiões (também descrita por Kenneth et al. (2001) para segmentação de imagens obtidas por tomografia computadorizada por feixe de electrões) para extracção de contornos após fecho da abertura da boca, e a inclusão da imagem dos dentes (obtidas por tomografia computadorizada) por sobreposição nas obtidas por IRM. Propõem também a utilização do algoritmo de Kohonen para o cálculo da linha média do tracto vocal, dado que tem como vantagem a obtenção das coordenadas nas três dimensões sem a necessidade de estabelecer pontos de referência manualmente.

A extracção de contornos nos estudos dinâmicos e em tempo real de IRM é mais complexa pois as imagens apresentam muito ruído (provocado pelo movimento dos articuladores), e artefactos devido ao uso de sequências rápidas (pela maior susceptibilidade). Na tentativa de solucionar este problema Avila-García et al. (2004) desenvolveram um método automático para extracção da forma e dimensões da língua por IRM, durante a fala. A combinação de modelos activos de forma e a transformada dinâmica de Hough, permite a localização e extracção da forma da língua, com vantagens na sua aplicação e apresentando resultados promissores.

A utilização da IRM foi expandida da imagem bidimensional para a imagem tridimensional, permitindo a medição da forma do tracto vocal e o cálculo das funções de área. A identificação dos

dentos no processo de segmentação, como já havia sido referido, constitui um problema sério da IRM, devido à quase inexistência de prótons de hidrogénio nos dentes.

Esta limitação torna indistinguível os dentes da cavidade oral, sendo o tamanho desta cavidade facilmente sobrestimado, dependendo também da habilidade do investigador no processo de segmentação. Vários métodos foram descritos na tentativa de exclusão dos dentes da cavidade oral, através do uso de moldes ou placas dentárias, ou na tentativa da sua inclusão por combinação com imagens dos dentes obtidas por tomografia computadorizada, mas que no entanto, acrescentam ainda mais tempo e esforço no processo de segmentação.

Neste sentido, Takemoto et al. (2004) desenvolveram outro método para sobrepor imagens tridimensionais dos dentes à informação obtida por IRM, de forma a permitir a medição do tracto vocal com exactidão. Obtiveram imagens dos dentes, realizando, através da IRM, uma aquisição na qual o sujeito tinha que manter um produto de contraste na cavidade oral, em decúbito ventral. Após sobreposição destas na informação volumétrica alvo (com o sujeito em decúbito dorsal), usando marcas de referência, concluíram que este método constitui um procedimento automático e exacto para solucionar a limitação específica da técnica de IRM na identificação dos dentes. O principal inconveniente desta técnica é a dificuldade na sobreposição das imagens obtidas, não só pelas diferentes posições adoptadas pelo sujeito (alteração devido à força da gravidade, como demonstrado anteriormente), mas também pela diferença resultante da articulação dos sons.

Inicialmente, a maioria dos estudos realizados no estudo da produção da fala estavam limitados ao plano médiosagital, obrigando à inferência para cálculo das funções de área e à impossibilidade de caracterizar determinadas articulações (por exemplo, as consoantes laterais caracterizam-se por uma oclusão ápico-alveolar no plano sagital, com abertura lateral para passagem do ar). Tais motivos vieram justificar a necessidade de construir modelos tridimensionais da forma do tracto vocal, informação útil não só no domínio da síntese da fala, como também na reabilitação (Badin et al., 2006).

## **1.5 BREVE CONCLUSÃO**

Desta revisão da literatura, verifica-se que dadas as condicionantes da investigação da fala (pela variabilidade na produção individual da fala e variabilidade inter-oradores), a maioria dos estudos revistos relacionados com a Engenharia e Electrónica tem amostras reduzidas (um a cinco sujeitos). Os estudos encontrados no âmbito da Medicina são ainda mais escassos, apresentando

contudo amostras maiores. No que diz respeito aos sons estudados, as vogais têm sido mais amplamente estudadas, em diferentes línguas, como Francês, Sueco, Japonês, Inglês ou Alemão, pois para além de ser mais fácil a sua sustentação, a segmentação das imagens é menos complexa pois são produzidas pela livre passagem do ar pelo tracto vocal. As consoantes têm sido estudadas por associação às vogais em diferentes contextos, mas não de forma isolada pois são mais difíceis de sustentar.

É notório, também, o interesse que a produção da fala desperta, bem como o estudo da nasalidade, da coarticulação, e de outros aspectos relacionados, por um lado com o tipo de estudo efectuado, e por outro, com condicionantes do equipamento, cuja influência é descrita em vários estudos. De realçar também, é a evolução e aperfeiçoamento da técnica de IRM, e do desenvolvimento de diferentes abordagens na procura de um melhor entendimento da fala.

Como é demonstrado em alguns estudos, apesar das limitações ainda verificadas, a IRM é sem dúvida uma técnica promissora no estudo da fala, inócua, e que fornece informação fiável e necessária à construção de modelos articulatórios tridimensionais.

De notar, por fim, os elevados tempos de sustentação usados nalguns estudos estáticos, e o ainda elevado tempo de aquisição usado para os estudos dinâmicos. O processo de segmentação das imagens é, sem dúvida, uma tarefa que acima de tudo requer rigor e exactidão, mas que apresenta ainda algumas limitações.

Na literatura revista não se encontrou nenhum estudo sobre a produção da fala, com base na IRM, do Português Europeu.

## CAPÍTULO 2. A FALA HUMANA

A comunicação oral entre duas pessoas constitui uma tarefa humana rotineira aparentemente simples, pois trata-se de um processo dinâmico de interacção através do qual as pessoas expressam os seus sentimentos, pensamentos, necessidades, e emoções, associando expressões faciais e/ou gestuais. No entanto, o processo de comunicação oral é bem mais complexo, pois requer a acção de actividades complexas como o pensamento, a audição, a fala e a coordenação motora e neurológica (Martins, 1988).

Todo o ser humano tem a capacidade de comunicar; no entanto, qualquer disfunção no processo de comunicação pode comprometer significativamente as suas relações interpessoais.

Por definição, a fala é o acto motor que pela articulação dos sons permite a expressão da linguagem. A linguagem é a capacidade que o ser humano tem para usar um código (língua) partilhado socialmente, que permite a transformação de pensamentos, sentimentos, num conjunto de símbolos arbitrários com significado próprio através da fala, escrita, língua gestual, ou ainda outras formas de comunicação.

Na análise dos sons da fala inserem-se duas grandes áreas de especial interesse (Faria et al., 1996):

- A fonética que se ocupa das propriedades dos sons da fala, do ponto de vista físico e do ponto de vista da sua produção e percepção;
- A fonologia estuda a organização dos sistemas dos sons das línguas.

O tracto vocal é a “ferramenta” que serve para a articulação e produção dos sons para a fala, sendo constituído por um conjunto de órgãos com funções distintas entre si (Quadro 1).

Fundamentalmente, o tracto vocal mede cerca de 17 cm nos homens e 14 cm nas mulheres, podendo ser considerado um tubo que se estende desde a laringe até às cavidades oral e nasal, e é composto:

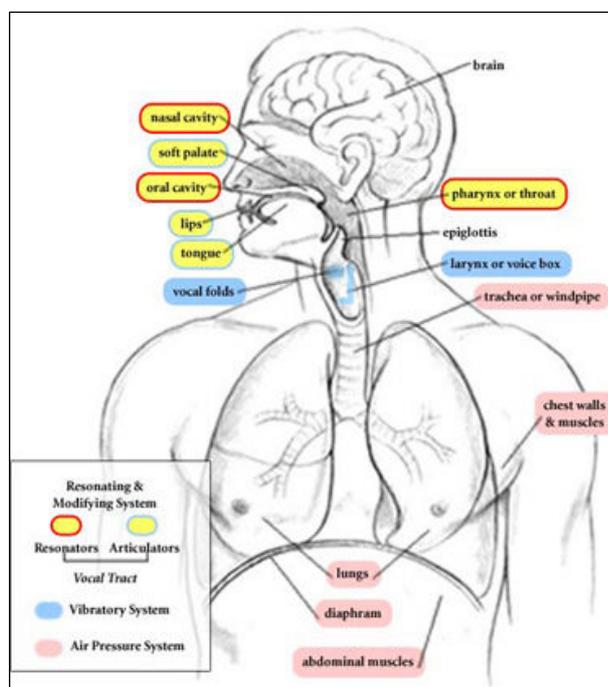
- Pelos pulmões (que vão ser a fonte de energia);
- Pela laringe (cordas vocais), que com a sua vibração numa cavidade ressonante vão produzir diferentes sons;
- Pelas cavidades supraglóticas (faringe, cavidade nasal e oral), que funcionam como cavidades de ressonância e de modulação dos sons produzidos.

**Quadro 1.** Órgãos intervenientes na produção da fala (Faria et al., 1996).

Órgão	Função trivial	Função na produção da fala
<b>Pulmões</b>	Troca de oxigénio e dióxido de carbono	Fonte de som (fornecimento do fluxo de ar)
<b>Cordas vocais (laringe)</b>	Servem como válvulas prevenindo a entrada de líquidos e alimentos para os pulmões	Produzem vibrações numa Caverna ressonante
<b>Língua</b>	Gustação, mastigação e deglutição	Articulação dos sons
<b>Dentes</b>	Mastigação	Articulador passivo e Separação acústica
<b>Véu do palato</b>	Funciona como válvula entre as cavidades oral e nasal.	Articulador activo
<b>Lábios</b>	Fecho da cavidade oral	Articulação dos sons

## 2.1 ÓRGÃOS DO TRACTO VOCAL HUMANO

A produção da fala depende dos órgãos do sistema respiratório, dos quais os pulmões são o motor de funcionamento de um conjunto de cavidades (nasal, oral, faríngea e laringea), produzindo o fluxo de ar necessário (Denes et al., 1993). Os brônquios e a traqueia vão permitir a ligação entre os pulmões e a cavidade laríngea (Figura 1).

**Figura 1.** Órgãos do tracto vocal.

Fonte: <http://www.voiceproblem.org/anatomy/learning.asp>.

### 2.1.1 Cavidade oral

Esta cavidade é constituída por uma porção externa, menor, o vestíbulo da boca, e por uma porção interna, maior, que é a cavidade oral propriamente dita (Gardner et al., 1988).

O vestíbulo corresponde ao espaço situado entre os lábios e as bochechas externamente, e pelos dentes e gengivas internamente.

A cavidade oral tem como limites anterior e laterais os arcos alveolares, dentes e gengivas. O palato constitui o tecto desta cavidade, sendo o chão da boca ocupado pela língua, que é mantida por uma série de músculos (intrínsecos e extrínsecos) e outros tecidos moles (Figura 2). A cavidade oral comunica-se posteriormente com a orofaringe.

A língua é um órgão muscular, volumoso, encontrando-se presa ao osso hióide, à mandíbula, à apófise estilóide e à faringe por diversos músculos. Para além do seu importante papel na mastigação, gustação e deglutição, tem um papel deveras importante na fala (pois é o articulador com maior mobilidade e mais flexível). Estruturalmente apresenta um ápex, um dorso, uma superfície inferior e uma raiz.

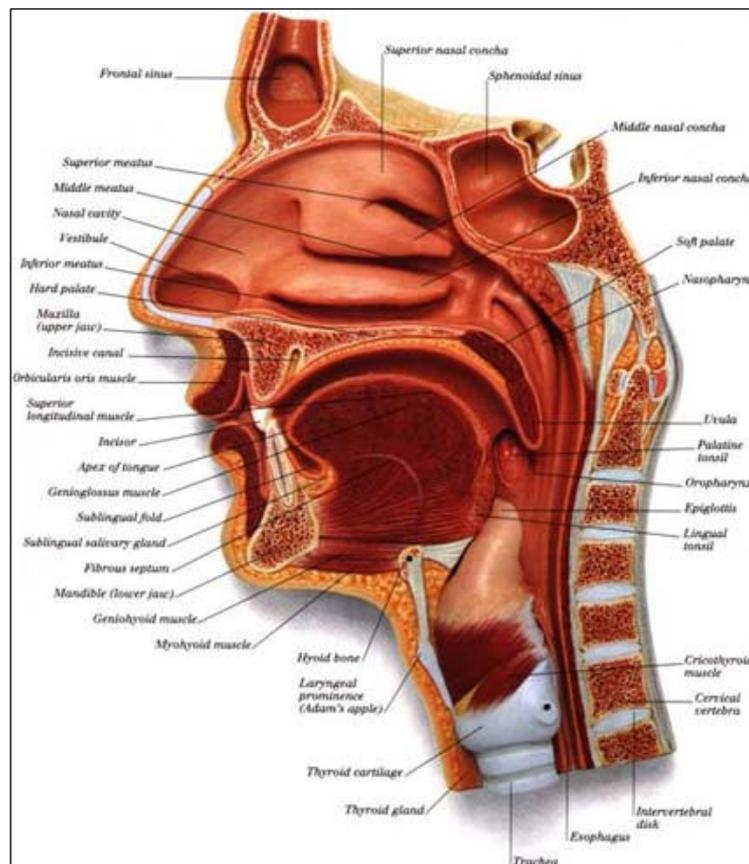


Figura 2. Corte sagital da cavidade oral.

Fonte: [http://www.ling.upenn.edu/courses/Summer\\_2004/ling001/lecture2.html](http://www.ling.upenn.edu/courses/Summer_2004/ling001/lecture2.html).

A dentição adulta é composta por trinta e dois dentes, sendo a zona interna de implantação denominada de zona alveolar.

O palato, ou tecto da boca, constitui o chão da cavidade nasal, estendendo-se em direcção posterior, permitindo a separação completa entre as duas cavidades. Divide-se em duas partes, correspondendo os dois terços anteriores ao palato duro e o terço posterior ao palato mole. O palato duro ou palato ósseo é constituído pelas apófises palatinas do maxilar (anteriormente), e pelas lâminas horizontais do osso palatino (posteriormente).

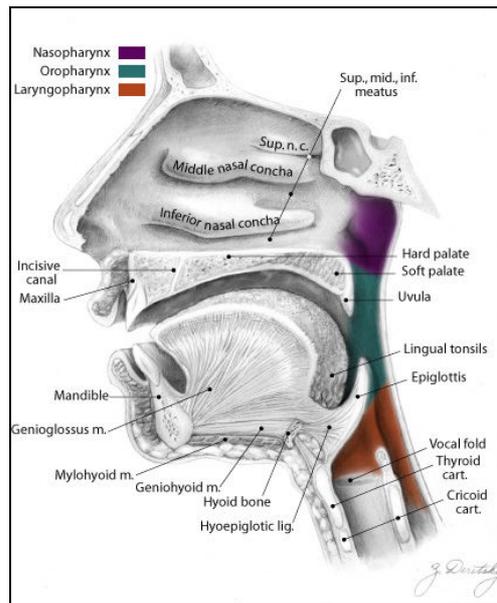
O palato mole, ou véu do palato, é uma prega fibromuscular móvel, e que está suspensa ao bordo posterior do palato duro. A sua função é fechar o istmo faríngeo durante a deglutição e durante a fala para a produção de sons orais. O bordo livre do palato mole, apresenta medialmente uma projecção de comprimento variável, a úvula, continuando-se lateralmente com duas pregas, denominadas de arcos palatoglosso e palatofaríngeo.

### 2.1.2 Cavidade nasal

A cavidade nasal é constituída externamente pelo nariz, e internamente por duas entradas de ar (separadas pelo septo nasal) – as fossas nasais, estendendo-se em direcção posterior até às coanas para se comunicar com a nasofaringe. A sua função é o controlo da pureza, humidade e temperatura do ar.

### 2.1.3 Cavidade faríngea

A faringe situa-se posteriormente à cavidade nasal e oral, e atrás da laringe, podendo por esse motivo ser dividida em três porções a nasofaringe (parte nasal), a orofaringe (parte oral) e a laringofaringe (parte laríngea). Estende-se desde a base do crânio, inferiormente até ao bordo inferior da cartilagem cricóide, continuando-se com o esófago. É um canal comum tanto para a deglutição como para a respiração. A nasofaringe comunica-se com a orofaringe através do istmo faríngeo (ou hiato nasofaríngeo). A orofaringe estende-se a partir do palato mole ao bordo superior da epiglote, comunicando-se anteriormente com a cavidade oral. A laringofaringe estende-se a partir do bordo superior da epiglote ao bordo inferior da cartilagem cricóide, continuando-se com o esófago (Figura 3).



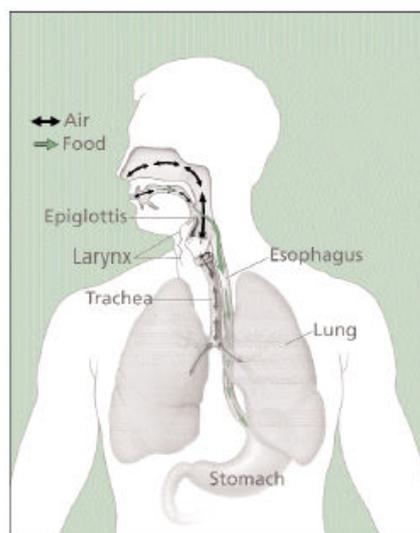
**Figura 3.** Porções da faringe e suas relações.

Fonte: <http://www.cortexity.com:8080/nicksblog/images/pharynx.jpg>.

#### 2.1.4 Cavidade laringea

A laringe localiza-se na porção anterior do pescoço, comunica-se superiormente com a faringe, e inferiormente com a traqueia, tendo como principais funções (Figura 4):

- Servir de válvula de forma a proteger as vias aéreas, especialmente durante a deglutição;
- Manter a via aérea permeável durante a respiração;
- Auxiliar na fala.

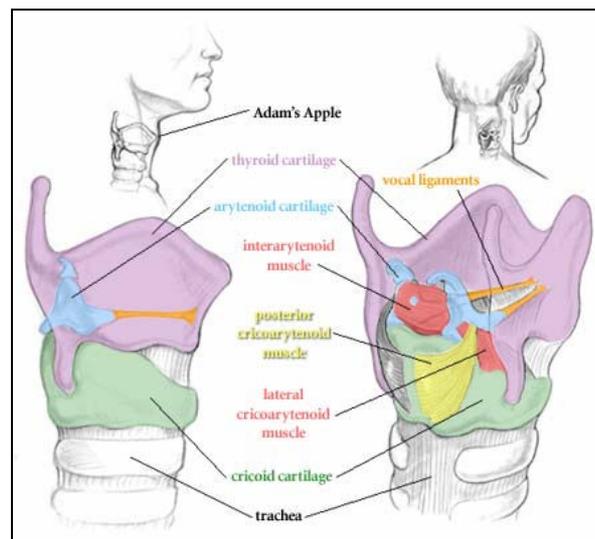


**Figura 4.** Função da laringe na deglutição e respiração.

Fonte: [http://www.cancer.gov/pdf/WYNTK/WYNTK\\_larynx.pdf](http://www.cancer.gov/pdf/WYNTK/WYNTK_larynx.pdf).

A laringe é constituída por três cartilagens ímpares (a cartilagem tiróide, a cartilagem cricóide e a epiglote) e três cartilagens pares (cartilagens aritenóides, cartilagens corniculadas, cartilagens cuneiformes). As cartilagens ímpares da laringe são constituídas por cartilagem hialina, podendo com o seu desenvolvimento calcificar, tornando-se assim visíveis radiograficamente. As cartilagens pares são cartilagens elásticas.

A cartilagem tiróide é a maior das cartilagens da laringe, e é constituída por duas placas (lâminas) que se unem formando uma elevação mediana denominada de proeminência laríngea ou maçã-de-adão (Figura 5). O ângulo formado pelas duas lâminas faz aproximadamente um ângulo recto no homem, sendo algumas vezes um pouco maior na mulher. No homem para além deste ângulo ser mais agudo do que na mulher, a proeminência laríngea é mais evidente, as cordas vocais são mais longas e a voz é um pouco mais profunda no tom.



**Figura 5.** Cartilagens da laringe e respectivos músculos.

Fonte: <http://www.voiceproblem.org/anatomy/learning.asp>.

A cartilagem cricóide tem a forma de anel de sineta, e localiza-se imediatamente abaixo da cartilagem tiróide. As cartilagens aritenóides articulam-se com o bordo superior da cartilagem cricóide, apresentando uma forma piramidal. As cartilagens corniculadas constituem-se num par de nódulos localizados nos ápices das cartilagens aritenóides e nas pregas ariepiglóticas da membrana mucosa. Nestas pregas e anteriormente às cartilagens corniculadas localizam-se dois pares de bastonetes – as cartilagens cuneiformes. A epiglote tem a forma de folha e encontra-se praticamente toda coberta por membrana mucosa. Localiza-se atrás da raiz da língua, do corpo do osso hióide e anteriormente ao ádito da laringe.

O ádito da laringe comunica com a cavidade laríngea, cavidade esta que está dividida em três porções: o vestíbulo laríngeo, ventrículos e a cavidade infraglotica, por dois pares de pregas horizontais – as pregas vestibulares e as pregas vocais (Figura 6). O vestíbulo laríngeo é limitado anteriormente pela epiglote e posteriormente pelas cartilagens aritenóides que estão unidas pelas pregas ariepiglóticas no sentido ânteroposterior. Lateralmente a essas pregas situam-se os seios piriformes, que se comunicam, anteriormente à epiglote, com as valéculas.

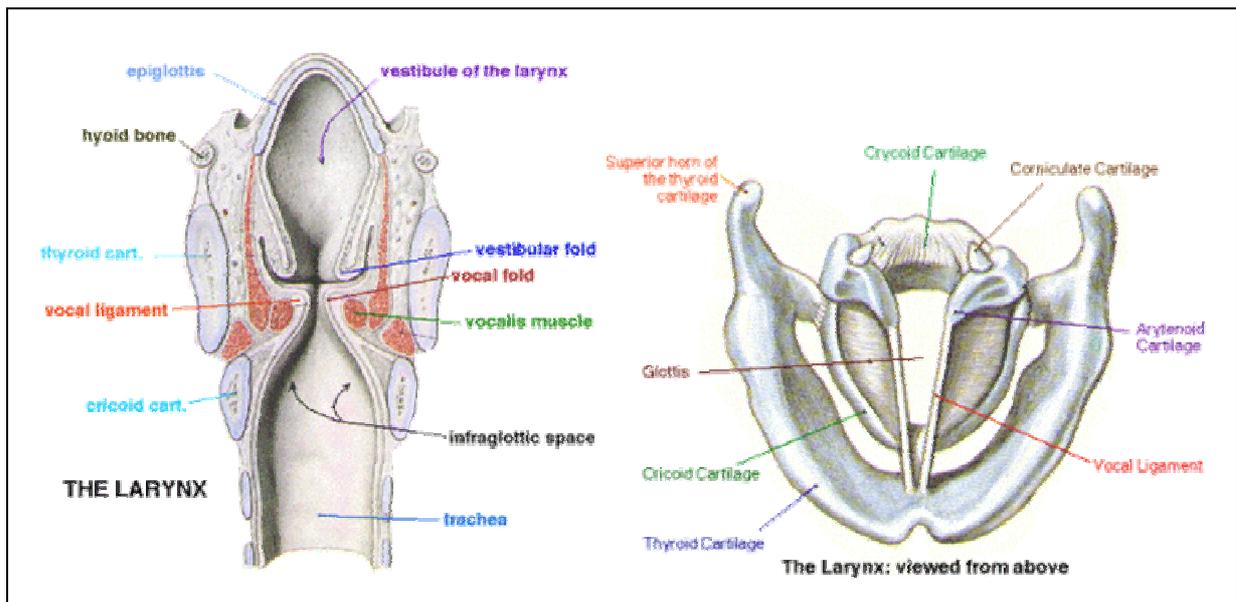


Figura 6. Laringe e cordas vocais.

Fonte: <http://www.hopkinsmedicine.org/voice/anatomy.html>.

As duas pregas vestibulares (ou “falsas cordas vocais”) estendem-se desde a cartilagem tiróide, anteriormente, à região das cartilagens cuneiformes, posteriormente. A sua função é apenas de protecção, não afectando normalmente a produção da fala.

As pregas vocais (ou cordas vocais verdadeiras) são anteparos músculo-membranáceos, móveis, de cor branco-pérola, e que se localizam medialmente e abaixo das pregas vestibulares. Cada uma destas pregas contém o ligamento vocal. O espaço entre as cordas vocais denomina-se de glote, que é controlado pelas cartilagens aritenóides. A rima da glote é a parte mais estreita da cavidade laríngea.

O fecho da laringe pode ser efectuado por músculos esfíntéricos num dos três terços ou níveis da laringe:

- Ádito – fecha-se durante a deglutição, protegendo as vias respiratórias da passagem dos alimentos;

- Pregas vestibulares – o seu fecho retém o ar inferiormente, permitindo um aumento da pressão intra-torácica (como na tosse), e um aumento da pressão intra-abdominal;
- Pregas vocais – as suas aproximações ocorrem durante a fala.

A laringe está suspensa por membranas, ligamentos e músculos (Quadro 2) ligados ao osso hióide (em forma de ferradura e localizado por baixo da raiz da língua). O volume das cavidades supraglotais pode ser alterado e conseqüentemente a pressão do ar, devido aos movimentos da laringe em altura e largura.

**Quadro 2.** Músculos da laringe e sua função.

Músculos da laringe	Função
<b>Extrínsecos</b>	Movem a laringe como um todo, sendo por isso classificados como levantadores (exs. músculo tiro-hióideu, músculo palatofaríngeo) e depressores (exs. músculo esterno-hióideu, músculo esternotiróideu).
<b>Intrínsecos</b>	Fecham ou abrem a laringe, sendo por isso classificados em esfíntéricos (ou adutores) e em dilatadores (ou abdutores) respectivamente. Exs.: músculo cricotiróideu, músculo cricoaritenóideu (posterior e lateral), músculo vocal, músculo aritenóideu (transverso e oblíquo) e músculo tiroaritenóideu.

A laringe pode ser observada *in vivo* através da laringoscopia, através de um espelho (laringoscopia indirecta) ou através de um tubo de endoscopia (laringoscopia directa).

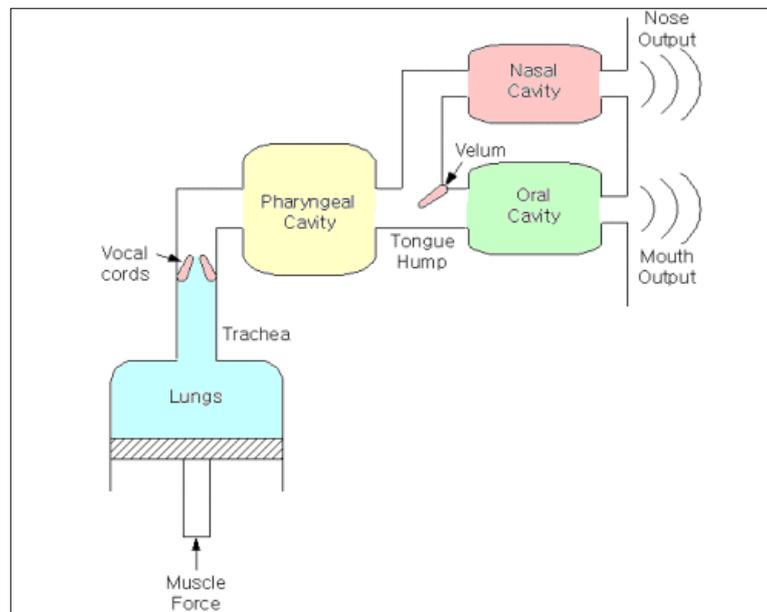
## **2.2 MECANISMO DA PRODUÇÃO DA FALA**

O tórax aloja os pulmões e é constituído por uma caixa torácica revestida por uma série de músculos que auxiliam no processo da respiração. A separação entre o tórax e a cavidade abdominal é feita pelo diafragma. O movimento da caixa torácica e do diafragma provocam uma diferença de volume dos pulmões durante a respiração.

Quando o volume dos pulmões aumenta, a pressão de ar diminui dando-se a entrada de ar nos pulmões. Este processo é conhecido por inspiração. No final deste processo, a pressão iguala a pressão atmosférica dando-se o processo inverso – a expiração.

É nesta fase de saída de ar para o exterior através dos diferentes órgãos – cavidades ressonantes (pulmões – brônquios – traqueia – laringe – faringe – cavidade oral e/ou nasal), que o

ar é modelado pela laringe e cavidades superiores nasais e orais, dando-se a produção da fala (Figura 7). Dadas as potencialidades dos diferentes órgãos do sistema respiratório, há na fase de expiração um aumento de pressão, permitindo a produção de ondas sonoras audíveis.



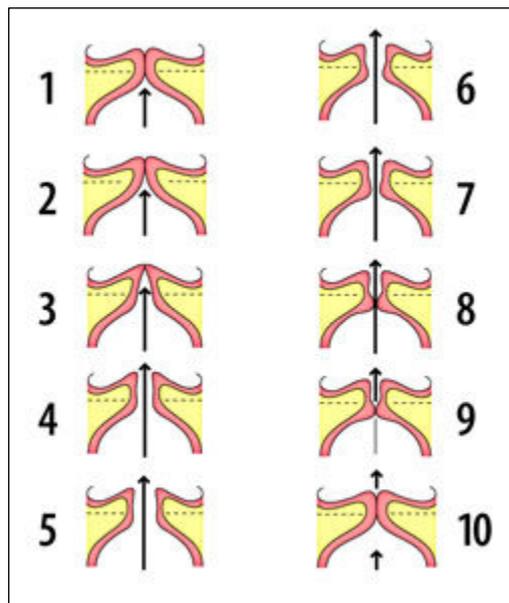
**Figura 7.** Representação esquemática da produção da fala.

**Fonte:** [www.ispl.korea.ac.kr/~wikim/research/speech.html](http://www.ispl.korea.ac.kr/~wikim/research/speech.html).

### 2.2.1 Modelação do ar expirado pelas cordas vocais

Durante a respiração, as cordas vocais estão abertas para a passagem do ar. Durante a fala, as cordas vocais actuam como um gerador de som, abrindo e fechando rapidamente (vibrando), modificando a passagem de fluxo de ar que provém dos pulmões.

A pressão subglotal criada pela junção das cordas vocais vai aumentando até ser suficiente para afastar as cordas vocais (Martins, 1988; Faria et al., 1996). Após o seu afastamento dá-se a passagem do ar, e consequentemente, a pressão diminui levando as cordas vocais a aproximar-se novamente; este ciclo repete-se durante a fala (Figura 8).



**Figura 8.** Pressão subglotal.

Fonte: <http://www.voiceproblem.org/pdfs/anatomy.pdf>.

Em suma, a fala compreende (Denes et al., 1993):

- A expiração do ar dos pulmões pelo diafragma e músculos abdominais e intercostais;
- A vibração das cordas vocais sendo a tensão e posição das mesmas controladas por acção muscular;
- Ressonância e articulação nas cavidades nasal, oral e faríngea, auxiliadas pelos músculos labial, lingual e palatino.

### **2.3 ACÚSTICA DA FALA**

O som pode ser entendido como a deslocação do ar que atinge o ouvido, ou seja, a vibração acústica susceptível ou não de provocar uma sensação auditiva. Este é produzido por uma fonte vibratória, podendo as variações ser transmitidas por meio gasoso, sólido ou líquido.

Uma onda sonora (vibração com determinadas características que a permitem ser audível), gerada pela fonte vibratória, provoca nas partículas do ar compressão / rarefacção, transmitindo-se a partir dessa fonte sob a forma de círculos (Martins, 1988).

As suas principais características são:

- Ciclo: oscilação completa da onda;
- Período: tempo correspondente a uma oscilação completa (ciclo), ou vibração dupla;

- Frequência: número de ciclos por segundo próprio de cada corpo; uma frequência de vibração alta origina um tom mais alto ou agudo, enquanto que uma frequência menor origina um tom mais baixo ou grave – altura do som; as frequências utilizadas na linguagem humana não ultrapassam os 10.000 ciclos/s;
- Amplitude: distância entre o ponto de repouso até ao ponto máximo de vibração.

A onda sinusoidal é uma onda simples, sendo denominada toda e qualquer onda não sinusoidal de onda complexa. A frequência das componentes sinusóides denomina-se de frequência fundamental (também conhecida por tom ou *pitch*), enquanto que nas ondas complexas são designadas por frequências harmónicas.

Um som composto por várias frequências aleatórias combinadas, forma e amplitudes aleatórias, denomina-se de ruído. Esta onda tem forma não repetitiva e não harmónica.

A resposta de ressonância é, por definição, a reacção de um corpo aos sons, ou seja, a absorção, armazenamento e radiação de energia. A frequência, para a qual o máximo de energia, é absorvida e irradiada denomina-se de frequência de ressonância. O corpo que responde ao som incidente é denominado de cavidade ressonante ou ressoador.

A intensidade do som é definida pela amplitude, pois consiste num reforço do som resultante do aumento da pressão do ar interno, que se obtém na fala humana, pela contracção dos músculos abaixadores do tórax. É considerada um factor relevante para a definição do acento, em particular para a língua portuguesa.

## **2.4 PRODUÇÃO E ACÚSTICA DA FALA – RELAÇÃO**

A produção da fala dá-se por três etapas independentes, a respiração (fonte de energia), a fonação (fonte de som) e a articulação (formação do som), permitindo a sua caracterização (Martins, 1988).

### **2.4.1 Respiração**

A respiração, assim como um sistema acústico, é equivalente à fonte de energia acústica para a fonação. A energia produzida pelos pulmões – ar, sofre variações na sua pressão, durante a respiração, quer por variação de volume e como também de velocidade.

O tracto vocal humano actua como um filtro acústico, pois vai mudando as características do som que vai passando através do mesmo, originando a fala.

### 2.4.2 Fonação

Na fonação existem três tipos de fontes sonoras, a fonte laríngea, a fonte de ruído e a fonte de transição. Na fonte laríngea, as cordas vocais funcionam como um gerador de ondas periódicas. O funcionamento das cordas vocais exerce-se pela pressão subglotal e pelo efeito de Bernoulli (que assenta no princípio de conservação de energia). Segundo este efeito, quando um fluxo de ar chega às cordas vocais quase ou totalmente fechadas, a área de passagem do mesmo diminui, e aumenta a velocidade do ar. A velocidade do ar atinge o seu máximo na glote, conseqüentemente a pressão baixa à medida que o fluxo de ar sai atravessando a glote (Martins, 1988; Mateus et al., 1990).

As modificações exercidas pelas cordas vocais permitem as variações do tom de voz, ou seja, definem a frequência fundamental de cada voz (Martins, 1988). Também outros factores como o tamanho da laringe são responsáveis pela definição do tom, definindo a voz:

- Nos homens: frequência fundamental entre 100 a 150Hz;
- Nas mulheres: frequência fundamental entre 200 a 250Hz;
- Nas crianças: frequência fundamental entre 300 a 500Hz.

Do mesmo modo, define o tom natural da voz cantada (baixo, soprano e contra-alto).

O ruído (fonte de ruído) pode ser produzido na cavidade oral sob a condição da passagem do ar poder criar uma turbulência. Denominam-se de fricativas sonoras, os tons produzidos com a passagem do ar mas sem fonte laríngea, ocorrendo na cavidade oral uma constrição suficiente que provoca uma fonte de ruído (fricatização).

A fonte de ruído transiente resulta de uma abertura repentina, brusca, de uma oclusão no tracto vocal, dando-se posteriormente uma rápida igualização da pressão de ar produzindo uma “explosão” sonora. Nas consoantes ditas oclusivas há um local do tracto vocal onde ocorre uma obstrução à passagem do ar.

### 2.4.3 Articulação

Articulação é a denominação dada à formação de uma constrição no tracto vocal, como resultado da actividade organizada de um articulador ou de um conjunto de articuladores.

As cavidades ressonantes durante a passagem de ar vão originar, por variação de configuração das mesmas, diferentes tipos de articulação, e consequentemente diferentes representações acústicas do som produzido.

Existem cinco tipos de configuração (Martins, 1988):

- Configuração vocálica – formação de cavidades de ressonância supraglotais; estas cavidades assumem formas e tamanhos diferentes para cada articulação, cujos ressoadores têm características específicas para certas frequências;
- Configuração lateral – passagem do ar livre pelos lados da boca com oclusão na parte média da língua, que se eleva em direcção ao palato duro;
- Configuração fricativa – obstrução parcial da passagem do ar no tracto vocal, causando uma fonte de ruído;
- Configuração oclusiva – obstrução total da passagem do ar nas cavidades supraglotais;
- Configuração nasal – passagem do ar pelas fossas nasais (Figura 9) por abaixamento do véu do palato, sendo a passagem nasal acoplada à passagem oral (vogais nasais). No entanto, a passagem nasal pode estar livre e a passagem oral fechada (consoantes nasais).

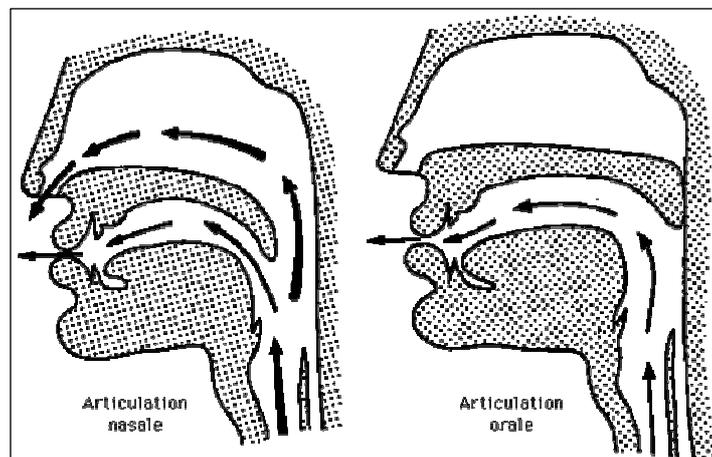


Figura 9. Representação esquemática de uma articulação nasal e oral.

Fonte: <http://www.unil.ch/ling/page24434.html>.

O modo de articulação depende:

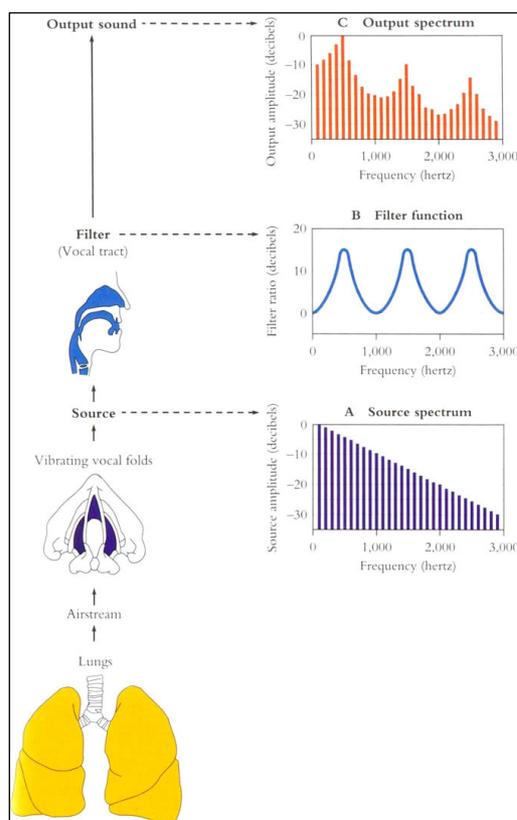
- Forma de vibração das cordas vocais;
- Obstrução do fluxo de ar acima da epiglote;
- Passagem do fluxo de ar através da cavidade nasal acoplada à cavidade oral;

- Passagem do fluxo de ar através da parte média da cavidade oral ou pelos seus lados.

O ponto de articulação refere-se aos locais ou partes do tracto vocal que contribuem para a definição do modo, sendo que existem partes estáveis (como o maxilar superior, os alvéolos, o palato, a parede faríngea) e outras móveis, ou seja, que se deslocam alterando a configuração articulatória (como a língua, lábios, véu do palato).

A variabilidade é uma característica própria da fala, isto porque tem um carácter dinâmico (Faria et al,1996). Para além dos diferentes articuladores estarem em constante movimento, ao fenómeno de produção de um som são muitas vezes acrescentadas características próprias de um som vizinho, devido à falta de sincronização dos movimentos necessários à produção de sons sucessivos. A esta influência de um som sobre o outro dá-se o nome de coarticulação (Barbosa, 1994).

O tracto vocal, segundo a teoria fonte-filtro (Fant, 1960), é considerado um filtro acústico de sons produzidos na laringe, onde as cordas vocais constituem a fonte sonora, e a forma do tracto vocal (cavidades ressonantes) constitui o filtro pois determina a forma como o som é modificado (Figura 10).



**Figura 10.** Teoria fonte-filtro – produção da fala.

**Fonte:** [http://www.spectrum.uni-bielefeld.de/~thies/HTHS\\_WiSe2005-06/session\\_05.html](http://www.spectrum.uni-bielefeld.de/~thies/HTHS_WiSe2005-06/session_05.html).

## 2.5 A AUDIÇÃO DA FALA

Ouvir é o mecanismo pelo qual o ouvido responde à vibração mecânica das ondas sonoras do ar por intermédio de propagação de ondas, e o processo que permite a transmissão da informação auditiva ao cérebro permitindo a sua compreensão – percepção (Martins, 1988).

O ouvido divide-se em três partes mediante a função específica que desempenham na audição (Figura 11):

- **Ouvido externo:** é composto pelo pavilhão externo e canal auditivo externo; acusticamente funciona como um filtro que capta e selecciona as vibrações sonoras que lhe chegam;
- **Ouvido médio:** é composto pela membrana timpânica e cadeia de ossículos (martelo, bigorna e estribo); a cadeia de ossículos actua pelo efeito de alavanca e pela diminuição da área onde se aplica a energia transmitida, contribuindo a própria membrana timpânica (pela sua curvatura) para a obtenção de uma boa transformação de impedâncias, actuando como um transformador mecânico das ondas sonoras que chegam pelo ouvido externo;
- **Ouvido interno:** é composto pela cóclea (inserida no labirinto) e pela membrana basilar; responsável pela transmissão do som, actua como transformador de energia mecânica em padrões de excitação nervosa.

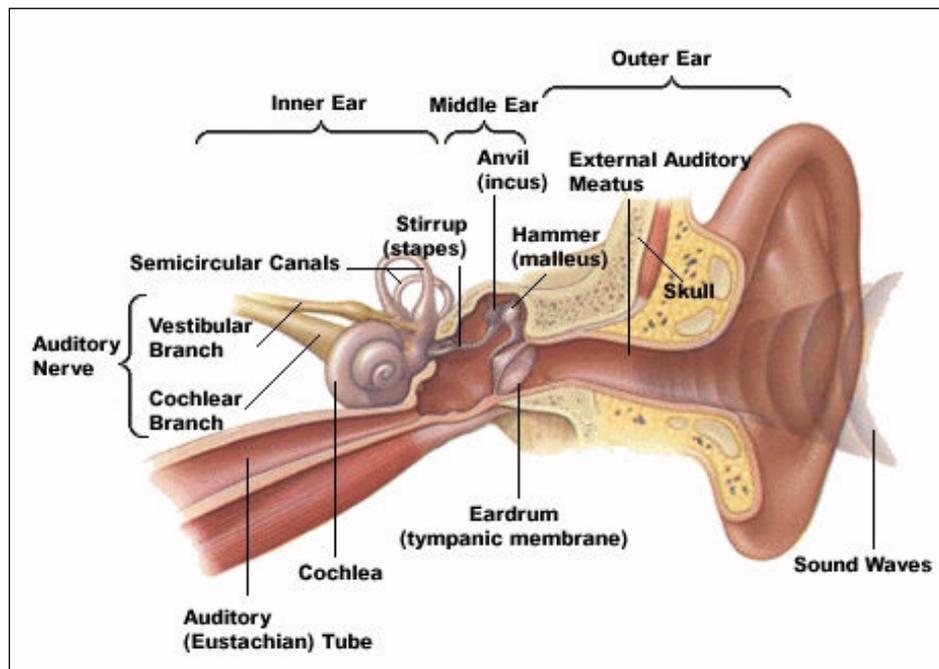


Figura 11. Constituintes do ouvido.

Fonte: <http://www.well-net.com/prevent/ear1.jpg>.

Uma onda sonora ao atingir a membrana timpânica promove a sua vibração; essas vibrações são transmitidas pelos ossículos até à janela oval. A pressão criada pelos movimentos do estribo promove a compressão dos fluidos da cóclea, que por sua vez provoca o movimento da membrana basilar. O órgão de Corti é o receptor que gera os impulsos nervosos (estímulo) em resposta às vibrações da membrana basilar. Posteriormente, o nervo auditivo conduz esse estímulo ao cérebro (Barbosa, 1994).

O ouvido humano consegue detectar um grande número de tons diferentes existindo, no entanto, variabilidade na forma são percebidos. Para frequências abaixo de 1000Hz o ouvido consegue detectar uma diferença entre tons distintos por 2 ou 3Hz, enquanto que para frequências acima de 1000Hz, a diferença de tons tem que ser maior para ser percebida (Martins, 1988).

A audição depende não só da frequência, como também da intensidade percebida.

A percepção auditiva é da responsabilidade do córtex cerebral do lobo temporal – o córtex auditivo, cujo funcionamento é pouco conhecido. Os sons da fala, dentro da área auditiva, ocupam uma área mais restrita, cuja frequência se situa na faixa de 0 até 8000Hz, e com uma zona de intensidade até 60/70dB. Acima de 120dB situa-se o limiar da dor ou desconforto.

Grande parte dos neurónios do córtex auditivo não responde aos sons ouvidos, pois a associação dos estímulos auditivos é feita com outras áreas de informação como as áreas sensoriais, áreas da linguagem, de forma a permitir a codificação dos mesmos em significado.

### CAPÍTULO 3. MEDIÇÃO E ANÁLISE DA PRODUÇÃO DA FALA

Dada a complexidade anatómica e fisiológica do tracto vocal humano, várias técnicas de análise têm sido desenvolvidas, na tentativa de perceber todos os mecanismos da fala.

Apesar dos aspectos fundamentais anatómicos e fisiológicos do tracto vocal serem comuns a todos os humanos, a fala é única e individual, dotando-nos da capacidade de comunicar, interrelacionar e viver em comunidade. Mas, acima de tudo, torna-se particularmente complexa pois o próprio indivíduo é incapaz de produzir um som várias vezes exactamente da mesma forma, dependendo de factores como a entoação, o timbre, e até mesmo por influência do local onde vive (dialectos da língua).

Várias investigações, no âmbito da fonética articulatória têm sido desenvolvidas na tentativa de estabelecer uma relação entre as unidades fonéticas e o conhecimento espacial e temporal dos vários articuladores da fala, como por exemplo com base na actividade muscular e, na regulação do suprimento de ar. De uma forma geral, os estudos assentam basicamente na actividade muscular, no movimento e forma dos articuladores, e na análise do sinal da fala (Figura 12).

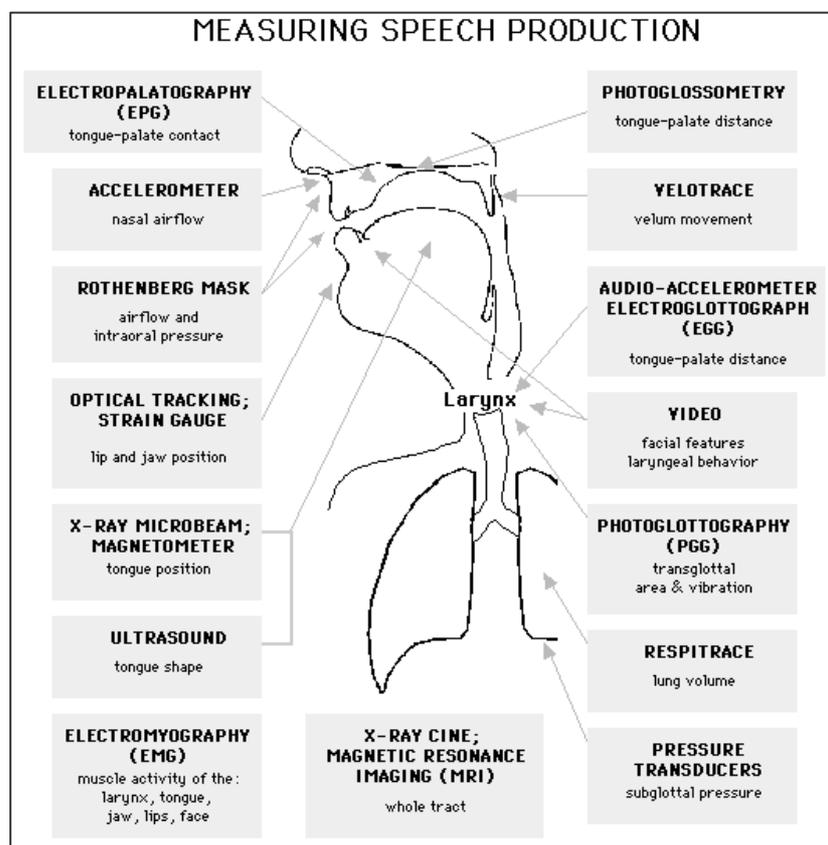


Figura 12. Técnicas de estudo da produção da fala.

Fonte: <http://www.haskins.yale.edu/haskins/HEADS/MMSP/tract.html>.

Várias técnicas e engenhos foram desenvolvidos para o estudo do tracto vocal, de forma directa ou indirecta, e durante a produção ou não da fala.

No entanto, as técnicas directas são escassas, pois apenas algumas estruturas são acessíveis à visão externa não invasiva. Como por exemplo, o movimento dos lábios e mandíbula podem ser observados optoelectronicamente, através de um sensor óptico, ou através da análise de vídeo (que permite também a análise de expressões faciais, bem como o posicionamento da mandíbula).

### **3.1 TÉCNICAS DE ANÁLISE DO SINAL DA FALA**

Os sons da fala são criados pela actividade vibratória do tracto vocal humano, sendo transmitidos através do ar (variação na pressão do ar) para o ouvido ou um microfone. Quando o receptor é um microfone, torna-se necessário a conversão destas variações de pressão, na mesma proporção, em corrente eléctrica.

Esta conversão é efectuada através de um conversor analógico-digital, e envolve duas fases:

- Amostragem – consiste em ler o sinal da fala num determinado período de tempo, tantas vezes quantas as necessárias para o poder reproduzir. Ou seja, transformar um sinal contínuo nas suas componentes discretas.
- Quantificação – cada valor de pressão amostrado é quantificado num determinado número de bits.

A fala consiste numa série de vibrações, que podem ser representadas em função das suas ocorrências ao longo do tempo, através de uma forma de onda. A análise desta forma de onda, nas suas componentes de frequências, permite a obtenção de um espectrograma.

A frequência de cada uma destas ocorrências determina a frequência fundamental ( $F_0$ ) da fonte laríngea, e contribui para a percepção do som produzido. A forma do espectro é determinada pela ocorrência de eventos de fecho e abertura das cordas vocais.

Acusticamente, o tracto vocal consiste numa série de tubos (cavidades de ressonância) com determinado comprimento, forma e volume, interligados entre si. A variação de cada um destes parâmetros numa determinada câmara, modulam a livre passagem do ar originando vibrações na gama da nossa audição.

No espectrograma são visíveis bandas de ressonância ou formantes, ou seja, frequências para as quais a energia local é máxima. Em conversação, o espectro da fala é muito limitado, sendo necessário atingir os  $7 \times 10^3$  Hz nas mais altas frequências, e os 50Hz nas mais baixas, para possibilitar uma boa inteligibilidade.

A análise da fala compreende vários níveis (Rowden, 1992):

- Nível de sinal – em que se pretende obter uma réplica eléctrica do sinal de pressão acústica ou de velocidade volumétrica associada.
- Nível segmental – consiste na segmentação do sinal em intervalos suficientemente grandes para discriminar uma palavra das restantes.
- Nível de frase – consiste numa mensagem definida em forma falada.
- Nível de aplicação – representa a interacção entre sistemas automáticos de fala e pessoas.

Várias ferramentas computacionais foram desenvolvidas para a análise da fala, sendo sem dúvida o programa PRAAT, o mais solicitado em investigação pelos foneticistas. Este programa permite a análise, síntese e manipulação do sinal da fala.

### **3.2 TÉCNICAS DE ANÁLISE MORFOLÓGICAS /FUNCIONAIS DO TRACTO VOCAL HUMANO**

Várias técnicas de análise do tracto vocal foram desenvolvidas através da criação de engenhos electrónicos ou dispositivos, permitindo a avaliação directa (técnicas invasivas e que requerem a cooperação do sujeito, principalmente na colocação de sensores e equipamentos de transdução) ou indirecta (técnicas não invasivas) dos aspectos anatómicos e fisiológicos relacionados com a produção da fala.

A electromiografia é útil na avaliação da actividade eléctrica dos diferentes músculos e articuladores da fala, podendo ser utilizada indirectamente, através do uso de eléctrodos de superfície, ou directamente através de eléctrodos agulha.

### 3.2.1 Técnicas indirectas

Das técnicas não invasivas destacam-se a máscara de Rothemberg (que permite a medição indirecta do fluxo de ar proveniente das cavidades nasal e oral) e os espirómetros ou pletismógrafos (que permitem a medição do volume pulmonar).

A electroglotografia (EGG) é uma técnica de análise indirecta, que consiste na medição da impedância eléctrica entre dois eléctrodos colocados externamente e a cada lado da laringe, com o objectivo de determinar a abertura e fecho (vibração) das cordas vocais.

A EGG consiste na representação do ciclo oscilatório das cordas vocais, permitindo também determinar a sua frequência de oscilação. Através do cálculo do quociente de fecho (adução) das cordas vocais, é possível a identificação de fenómenos de hipo e hiper adução.

As suas principais aplicações clínicas são a avaliação do grau de fecho das cordas vocais, auxiliar no tratamento de disfunções da voz, e auxiliar no diagnóstico e investigação do mecanismo laríngeo.

A nasometria, apesar de não constituir um instrumento de medição da articulação da fala, é um excelente método para a determinação da presença de nasalidade na produção da mesma. Através da medição da energia acústica (outputs acústicos nasais e orais), a nasometria permite avaliar a extensão de “fuga” nasal em situações de fenda do palato, e também fornece um feedback durante a terapia.

### 3.2.2 Técnicas directas

As técnicas mais invasivas são as mais comuns, e permitem uma observação directa através da colocação de sensores, ou outros dispositivos. Através de um endoscópio de fibra óptica é possível a observação em vídeo da laringe, bem como efectuar o registo fotográfico das estruturas da laringe e áreas transglotais através da fotoglografia. Também é possível, através da colocação de sensores de pressão miniaturizados, e efectuar a medição das pressões supra e subglotal.

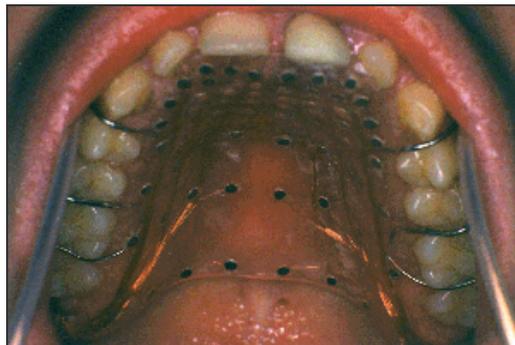
### 3.2.3 Observação da língua

O estudo dos articuladores é uma tarefa complicada, pois possuem um elevado grau de liberdade durante a produção da fala, originando um problema no seu controlo e observação.

A observação dos movimentos da língua é de extrema importância, pois é o órgão mais versátil e complexo, e um importante interveniente na articulação, podendo ser efectuada através do uso da optoelectrónica, da ultrassonografia, da imagem por radiação X e da indutância electromagnética.

A fotoglossometria (ou glossometria) é uma técnica optoelectrónica que mede por reflexão, a distância entre a superfície da língua e pontos do palato duro.

A electropalatografia (EPG) é uma técnica que permite a medição da porção de contacto entre a língua e determinados pontos do palato duro (Figura 13). Através da colocação de uma placa em acrílico, constituída por 62 eléctrodos divididos em três regiões (região alveolar, palatal e velar), que se adapta ao palato, é possível avaliar o tipo de articulação da língua relativamente ao palato. Também é possível avaliar a posição da língua, sendo o contacto adquirido sob a representação dinâmica. A EPG é utilizada nas investigações de fonética, no tratamento clínico de pacientes com dificuldades de articulação ou fenda do palato, e também no ensino de segundas línguas.



**Figura 13.** Electropalatografia.

**Fonte:** <http://www.cstr.ed.ac.uk/artic>.

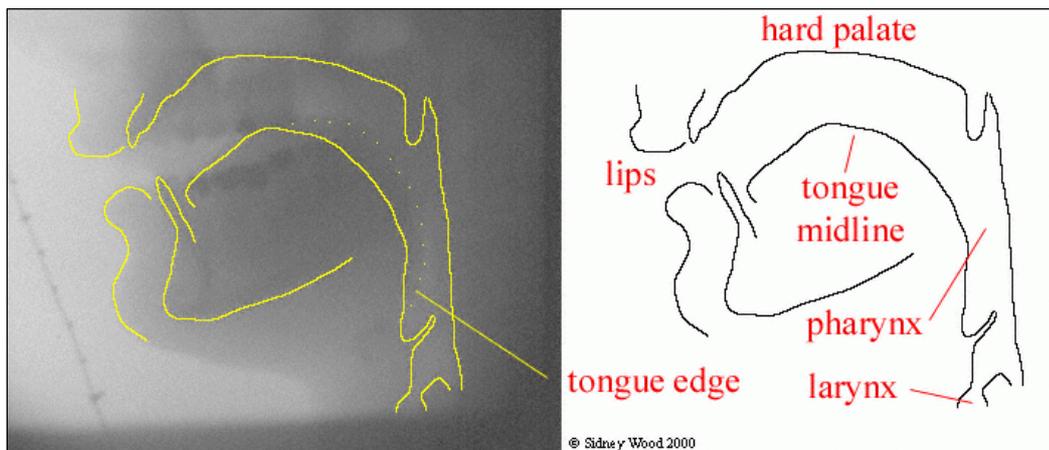
A optopalatografia (OPG) combina os benefícios da EPG e da glossometria, permitindo a medição do volume e forma da cavidade intra-oral e suas constricções. Através da utilização de um palato artificial com dezasseis sensores de fibra óptica, é possível a medição dinâmica do contacto, distância e pressão entre a língua e o palato (Wrench et al., 1996).

A ultrassonografia (ecografia) permite a aquisição de informação dinâmica da língua e de outras partes moles, sendo amplamente usada em investigação da fala. Para além da sua inocuidade, actualmente a ultrassonografia permite a obtenção de modelos tridimensionais da língua.

A cineradiografia (ou vídeofluoroscopia) sobrepõe-se à ecografia, devido à sua elevada resolução espacial e por permitir uma visão global de todo o tracto vocal; no entanto, com os

progressos da ultrassonografia e utilização da IRM, e dado o seu carácter ionizante tem um papel cada vez menor para fins de investigação.

Esta técnica consiste na obtenção de radiogramas sucessivos durante a produção de um segmento da fala, de forma a mostrar a evolução da posição dos órgãos articulatórios (Figura 14). Cada imagem representa uma posição articulatória num dado instante, e que é determinada pelas medidas radiológicas obtidas do traçado de eixos relativamente a pontos de referência fixos (Martins, 1988).

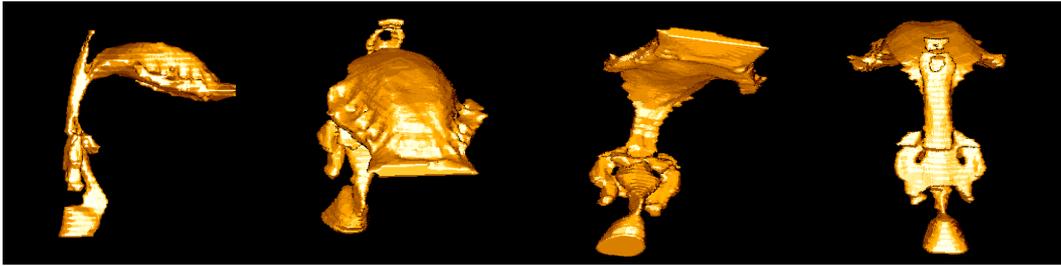


**Figura 14.** Delimitação do tracto vocal obtida por cineradiografia.

**Fonte:** <http://www.ling.lu.se/persons/Sidney/coartdem/tracing.html>.

A articulografia electromagnética (EMMA) permite, através de um sensor, a obtenção de informação relativa à posição e trajecto dos articuladores a duas dimensões, comparável à informação obtida por radiação X. A EMMA destina-se ao estudo da coarticulação, permitindo a monitorização, através de sensores, dos movimentos de articuladores como a língua, a mandíbula, os lábios ou o véu do palato (Hoole et al., 1999).

A tomografia computadorizada por feixe de electrões (EBCT) permite a obtenção de imagens tridimensionais do tracto vocal por reconstrução de imagens axiais. Com tempos de aquisição mais rápidos comparativamente à IRM e melhor resolução espacial, a EBCT consiste na emissão de um feixe de electrões usando a tecnologia de tomografia computadorizada, permitindo a obtenção de cortes axiais (Kenneth et al., 2001). As imagens são posteriormente reconstruídas a três dimensões usando interpolação dos dados. A sua principal limitação é a dose de radiação ionizante à qual o paciente é submetido (Figura 15).



**Figura 15.** Reconstrução a três dimensões da forma do tracto vocal, através da EBCT, da vogal /a/.

**Fonte:** [http://web1.dcpa.org/brad\\_html/ctgallery.html](http://web1.dcpa.org/brad_html/ctgallery.html).

A imagem por Ressonância Magnética permite uma visão global do tracto vocal, multiplanar e a três dimensões com elevado detalhe anatómico, e assim como a ecografia tem baixa resolução espacial comparativamente com a imagem por radiação X. Contudo, é uma técnica valiosíssima e que tem sido alvo de inúmeras investigações no âmbito da produção da fala, e cuja descrição é efectuada mais adiante.

## CAPÍTULO 4. PROCESSAMENTO DA FALA

### 4.1 BREVE RESENHA HISTÓRICA

O interesse pela reprodução da fala surge em 1773, quando um professor de Fisiologia em Copenhaga – Christian Kratzenstein, constrói uma máquina que produzia vogais, usando tubos ressoadores ligados a palhetas vibratórias activadas pela passagem do ar. Vivia-se então a era Mecânica (Barros, 2002).

Na mesma época, Wolfgang Von Kempelen tentara também construir uma máquina que falasse. A máquina construída por Von Kempelen, considerado o primeiro foneticista, em 1791, foi a primeira a produzir palavras inteiras e pequenas frases, e não apenas fonemas.

Em 1835, Joseph Faber inclui no seu mecanismo de produção da fala, um modelo de língua e uma cavidade laríngea, cuja forma podia ser controlada, permitindo à máquina falar em qualquer língua europeia (Figura 16).



**Figura 16.** Máquina falante construída por Joseph Faber.

**Fonte:** <http://mambo.ucsc.edu/psl/smus/smus.html>.

R. Riesz (EUA), em 1937, constrói uma máquina similar às anteriores, mas com um tracto vocal mais aproximado da realidade.

No início do século XX, o progresso da engenharia eléctrica tornou possível a síntese eléctrica contrastando com a era mecânica. Homer Dudley desenvolve o primeiro dispositivo deste tipo – *Voder*, em 1939, tratando-se basicamente de um sintetizador espectral (Figura 17).

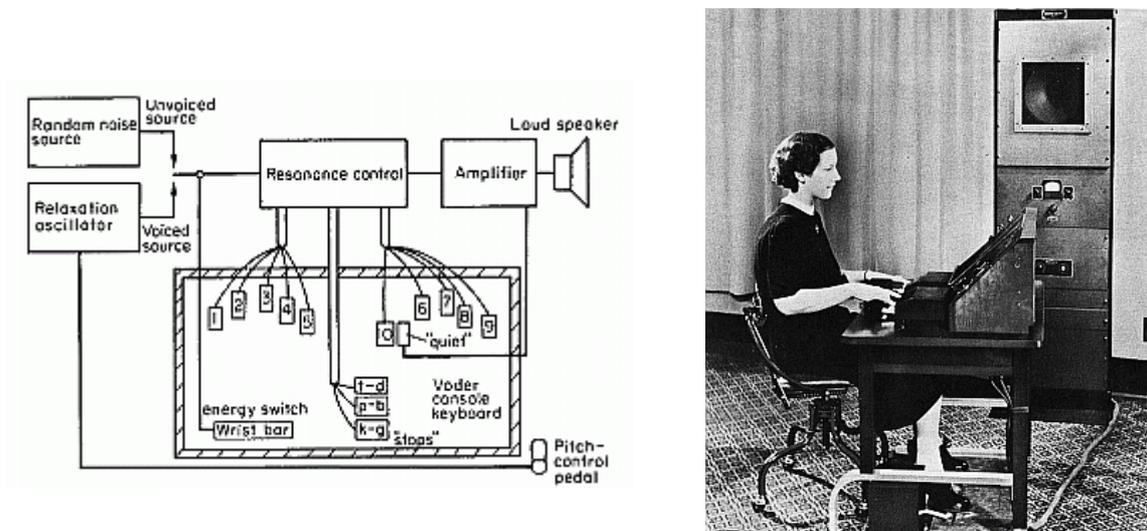


Figura 17. Voder de Homer Dudley.

Fonte: <http://mambo.ucsc.edu/psl/smus/smus.html>.

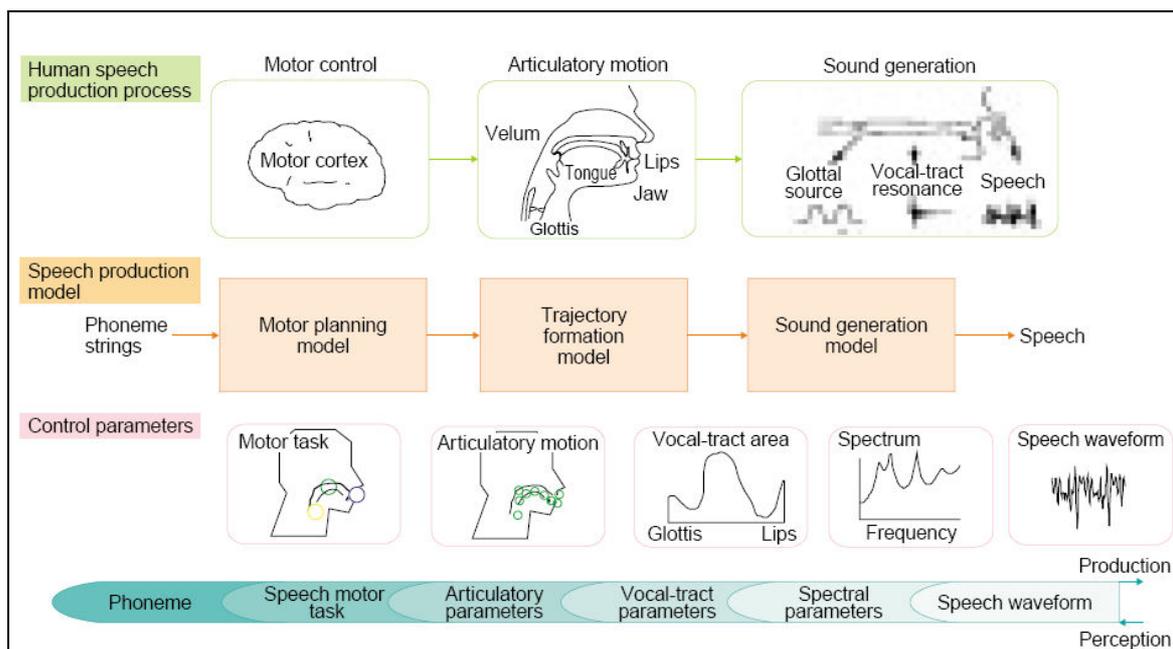
Já em 1950, Frank Cooper desenvolveu um sintetizador da fala que funcionava como um inversor do espectrograma de som, e cujo objectivo era a investigação da percepção da fala. O primeiro investigador de formantes em cascata foi desenvolvido em 1953, por Gunnar Fant.

A partir dos finais dos anos 60, a introdução da era digital, permite o desenvolvimento da síntese da fala. Mas só em meados dos anos 80 se começou a perceber que, de facto, a intensidade e a duração, desempenhavam um papel importante na fala, constituindo um grande avanço no desenvolvimento da área.

## 4.2 SÍNTESE DA FALA

A síntese articulatória constitui uma ferramenta fundamental na síntese da fala, permitindo o estudo dos aspectos linguísticos e perceptivos dos eventos articulatórios, simulando a produção da fala (Teixeira et al.; 2001).

Na construção de um modelo para produção da fala tem de se ter conta todos os processos envolvidos (processamento da linguagem, controlo motor, movimentos articulatórios e a produção do som) de forma a permitir uma simulação deste mecanismo (Figura 18).



**Figura 18.** Mecanismo de produção da fala e sua correspondência com um modelo computacional.

**Fonte:** <http://www.ntt.co.jp/tr/0305/files/ntr200305024.pdf>.

Vários estudos têm sido desenvolvidos na tentativa de obter informação relativa à função articulatória directamente a partir do sinal acústico da fala. Efectivamente, existem problemas na síntese da fala, pois os sistemas de reconhecimento baseiam-se normalmente em aproximações estatísticas. O conhecimento da produção da fala e percepção é pequeno. Surge, portanto, a necessidade de obter informação relativamente à configuração do tracto vocal durante a produção dos diferentes sons.

O sistema de síntese articulatória (*Articulatory Synthesizer – ASY*), desenvolvido pelos Laboratórios Haskins, permite o estudo dos eventos articulatórios. Este modelo do tracto vocal humano é constituído por um conjunto de parâmetros que permitem controlar grande parte dos articuladores (língua, véu do palatino, osso hióide, mandíbula e lábios) usados durante a fala.

Estes parâmetros chave fornecem uma descrição da forma do tracto vocal, com base na determinação de áreas (Figura 19). Outros parâmetros como a fonte sonora (excitação) e a informação relativa aos tempos de cada movimento, permitem controlar variações temporais e de posição dos articuladores.

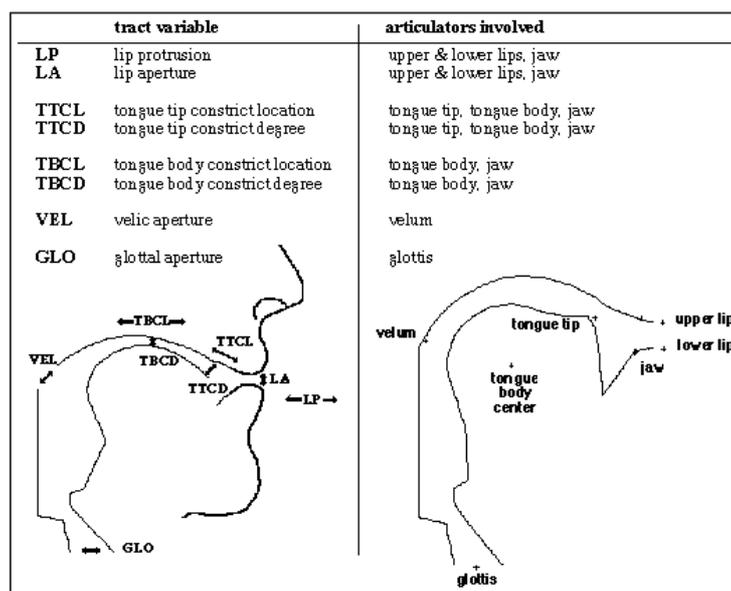


Figura 19. Parâmetros chave do modelo do tracto vocal humano para síntese articulatória.

Fonte: <http://www.haskins.yale.edu/haskins/MISC/RESEARCH/GesturalModel.html>.

A criação de modelos de simulação da produção da fala requer não só conhecimento dos aspectos anatómicos e fisiológicos como também da forma como esses aspectos variam ao longo do tempo.

De forma a conseguir um melhor feedback, foi criada uma versão do programa de síntese articulatória – o *Configurable Articulatory Synthesizer (CASy)*, que permite ao utilizador sobrepôr uma linha delimitativa do tracto vocal numa imagem sagital obtida por IRM tipicamente, ou raio X. Esta versão permite ajustar graficamente os parâmetros do modelo, de forma a determinar as suas dimensões (Figura 20). O CASy tem a capacidade de adaptação a qualquer imagem de IRM, independentemente do indivíduo, e permite alterar os parâmetros chave.

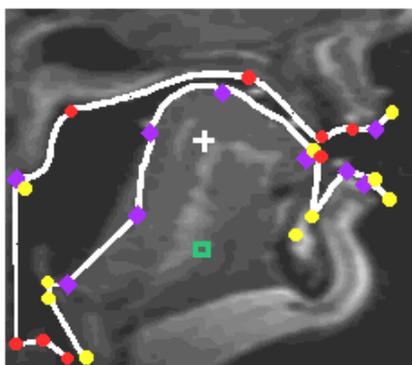


Figura 20. Versão do programa de síntese articulatória – CASy.

Fonte: <http://www.haskins.yale.edu/haskins/HEADS/ASY/INFO/CASY.html>.

## CAPÍTULO 5. LÍNGUA PORTUGUESA

A fonética é a disciplina científica que se ocupa dos sons da fala, e do modo como estes são produzidos e percebidos (Faria et al., 1996). Desta forma, mediante este conceito torna-se relativamente simples a identificação das suas três vertentes:

- **Fonética articulatória:** dedica-se ao estudo da anatomia e da fisiologia da produção da fala e descreve a forma como os sons da fala são articulados pelo aparelho fonador;
- **Fonética acústica:** dedica-se às propriedades acústicas dos sons da fala, análise da composição, produção, propagação do tipo de ondas sonoras, e dos filtros e ressoadores;
- **Fonética perceptiva:** estuda os processos de audição da fala humana e do processamento das suas características pelo cérebro humano.

### 5.1 FONÉTICA ARTICULATÓRIA

Todos os sons requerem energia, e na fala os sons são produzidos com auxílio de um fluxo de ar proveniente dos pulmões durante a respiração. Uma série de órgãos permitem o aumento ou diminuição da pressão no tracto vocal, criando áreas de constrição no mesmo.

Existem três fontes moduladoras que perturbam a livre passagem do ar:

- A fonte pulmonar (sistema sub-laríngeo) que é constituída pelos pulmões que forçam a passagem do ar pelo tracto vocal;
- A fonte glotal (sistema laríngeo) que envolve o fecho da glote e a elevação ou abaixamento da laringe, variando a pressão no tracto vocal;
- A fonte velar (sistema supraglotal) que envolve o fecho do véu do palatino impedindo ou modulando a passagem do ar pela via nasal.

Estas fontes podem ser classificadas em *egressivas*, se forçam a saída de ar, ou em *ingressivas*, se retêm a saída de ar. As consoantes produzidas por fontes egressivas são conhecidas por consoantes ejectives. Estas consoantes são inerentemente não vozeadas, enquanto que os sons produzidos por fontes ingressivas (consoantes implosivas) são geralmente vozeados.

Contudo, alguns sons da linguagem podem ser formados sem auxílio da respiração, designando-se de **cliques**. Os cliques apesar de frequentes noutras línguas exóticas, são inexistentes na Europa.

A passagem do ar pelas fontes modeladoras pode ser livre, comprimida ou detida momentaneamente por uma oclusão completa.

### 5.1.1 Constrições na fala

A resistência ao fluxo de ar na fala advém principalmente das constrições criadas na laringe e, em diferentes regiões do tracto vocal, para a geração de fontes sonoras.

Qualquer processo de produção da fala envolve sempre a formação de uma ou mais constrições supra-pulmonares, onde o fluxo de ar é modulado, com vista à geração de vozeamento (na laringe) e/ou ruído (na laringe ou supra-laringe).

Cada unidade fonética pode ser classificada, de acordo com o modo e ponto de articulação, em duas classes: as vogais e as consoantes. O modo de articulação define a qualidade do som de acordo com o tipo de formação, e o ponto de articulação diz respeito à parte anatómica que contribui para a definição do modo.

As **vogais** são sons pronunciados com livre passagem do ar pelo tracto vocal (Malmberg, 1954), e apesar de estarem situadas nas zonas dos pontos de articulação são classificadas em classes diferentes: abertas, fechadas, médias e centrais, de acordo com a posição da língua e projecção dos lábios, e grau de abertura da boca (Quadro 3).

**Quadro 3.** Classificação das vogais e semivogais.

Ângulo de abertura do maxilar inferior	Projecção dos lábios	Posição da língua em relação ao palato
Fechadas <mi <u>to</u> >	Arredondadas	Anteriores <vi <u>i</u> >
Semi-fechadas <me <u>do</u> >	<q <u>vo</u> >	Centrais <a <u>nde</u> >
Semi-abertas <re <u>gra</u> >	Não-arredondadas	Posteriores <su <u>mo</u> >
Abertas <p <u>á</u> >	<si <u>pó</u> >	

As **semivogais** têm as características articatórias das vogais com uma duração menor, e ocorrem sempre junto de uma vogal, formando um ditongo. Na língua portuguesa apenas as vogais [i] e [u] funcionam como semivogais, sendo representadas por [j] e [w] respectivamente (Barbosa, 1994).

As **formantes** são zonas de frequências intensificadas pelas cavidades de ressonância, e são específicas para cada vogal, definindo a sua qualidade. A qualidade vocálica é essencialmente determinada pelos dois primeiros formantes ( $F_1$  e  $F_2$ ) que permitem só por si o reconhecimento da vogal. Assim, as vogais e semivogais são definidas pela fonte laríngea, pela livre passagem do ar e pelas formantes (Martins, 1988).

O sistema tónico do PE é formado por nove vogais (Figura 21), que podem ser definidas pela relação dos seus dois primeiros formantes  $F_1$  e  $F_2$ , constituindo o chamado triângulo acústico das vogais.

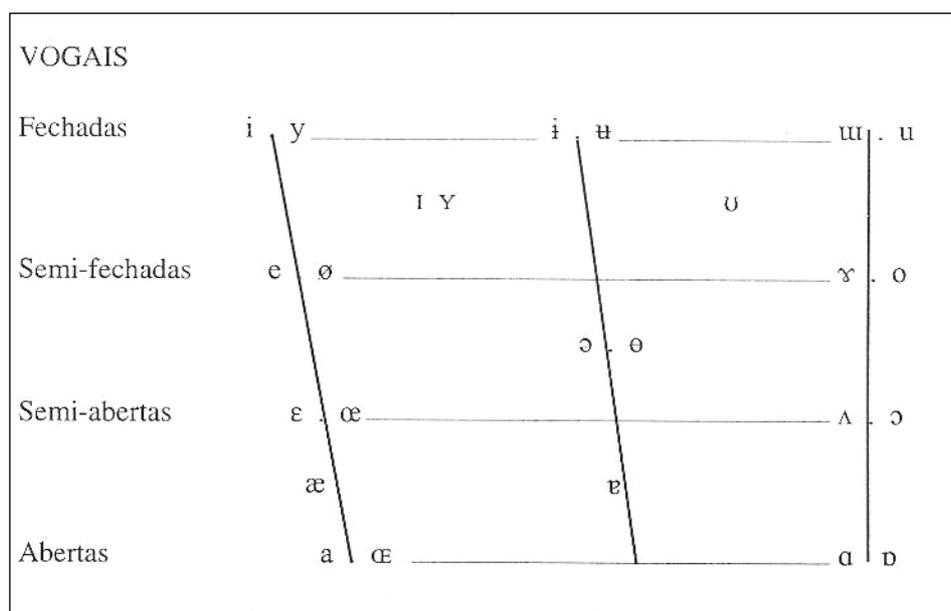


Figura 21. Vogais do Português (Faria et al., 1996).

As **consoantes** são sons que se caracterizam pela compressão ou fecho completo (momentâneo) da passagem do ar (Malmberg, 1954). Na produção das consoantes, a passagem de ar pode ser totalmente obstruída em determinado ponto do tracto vocal, durante um certo período de tempo, ou parcialmente obstruída, dificultando em determinado ponto a sua passagem.

Cada som é referido quanto ao seu ponto de articulação pelo nome do órgão que se desloca (articulador activo), seguido do nome do órgão para o qual este se desloca (articulador passivo) (Quadro 4). Quanto ao modo, as consoantes são classificadas em cinco classes, dentro das quais se

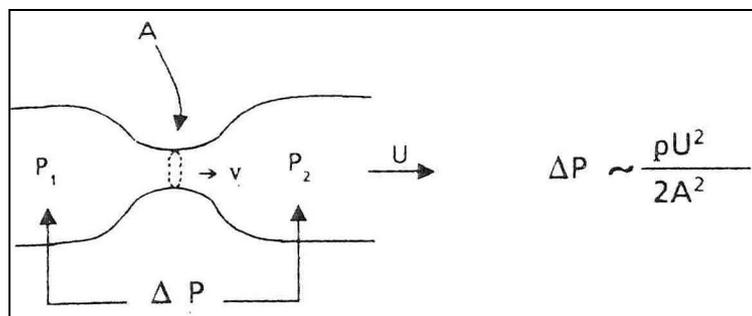
podem dividir em *surdas* ou *sonoras* (ausência ou presença de vibração das cordas vocais respectivamente) – característica de vozeamento.

**Quadro 4.** Modo e ponto de articulação das consoantes.

Modo de articulação		Ponto de articulação	
Oclusivas <capa>	Vibrantes <raro>	Bilaterais <pomo>	Lábio-dentais <fava>
Fricativas <safar>	Africadas <dia>	Dentais <tudo>	Alveolares <samo>
Laterais <luja>		Palatais <chá>	Velares <gago>
		Uvulares <rei>	

## 5.2 FONÉTICA ACÚSTICA

A síntese da fala (teoria acústica) tem em conta um modelo simples de representação acústica do tracto vocal, que consiste em considerar o tracto vocal como um tubo no qual é imposta uma ou mais constrições de área e posição variáveis (Figura 22).



**Figura 22.** Representação acústica do tracto vocal (Mateus et al., 1990).

Da criação de uma constrição na passagem de ar resulta uma baixa de pressão que depende (Mateus et al., 1990):

- da velocidade do volume do fluxo ( $U$ );
- e da área de secção recta da constrição ( $A$ ).

Segundo uma estimativa de Stevens (1975) existe um valor crítico para a área de secção recta da constrição ( $A$ ) para a qual acima de  $0,1 \times 10^{-2} \text{m}^2$  tem um papel determinante, sendo os valores abaixo desta fronteira negligenciáveis. Quando  $A$  assume valores na ordem dos  $0,1 \times 10^{-2} \text{m}^2$  ou menos, como acontece nas fricativas, a relação entre a pressão e o fluxo é mais complexa. Neste caso, a geometria da constrição tem um papel importante sendo o valor de  $A$  negligenciável.

Quando  $A$  é maior que  $0,1 \times 10^{-2} \text{m}^2$ , as diferenças de pressão à frente e atrás variam inversamente, enquanto que variam directamente com a quantidade de ar por unidade de tempo e a densidade do ar ( $\rho$ ).

Os sons da fala são constituídos por ondas complexas. As vogais e as semivogais são sons periódicos ou musicais (ou vozeados), pois são produzidos por uma fonte glotal com regularidade nas variações de pressão.

As consoantes sonoras são sons mistos (pois podem ter duas fontes sonoras, a glotal e a bucal) produzindo frequências periódicas e ruídos. As consoantes surdas são sons aperiódicos, ou seja, têm frequências aleatórias.

As cavidades oral e nasal funcionam como filtros, enfraquecendo certas frequências, enquanto que a laringe, o véu do palatino, a cavidade nasal e lábios funcionam como ressoadores, incrementando outras frequências.

### 5.2.1 Terminologia portuguesa para os traços

Em 1952, pela primeira vez, Jakobson, Fant e Halle utilizaram a análise espectral para a análise fonológica e linguística, constituindo uma etapa importante na descrição e classificação das línguas. A obra "Preliminaries to Speech Analysis" destes autores, propõe traços para a língua inglesa, desconhecendo-se investigação sobre o assunto para a língua portuguesa. Em 1956, os mesmos autores apresentam os traços acústicos e suas correspondentes realizações articulatórias.

Para a língua portuguesa foi proposto um exemplo de classificação das vogais orais do Português a partir do estudo acústico realizado por Jakobson e seus colaboradores (Figura 23).

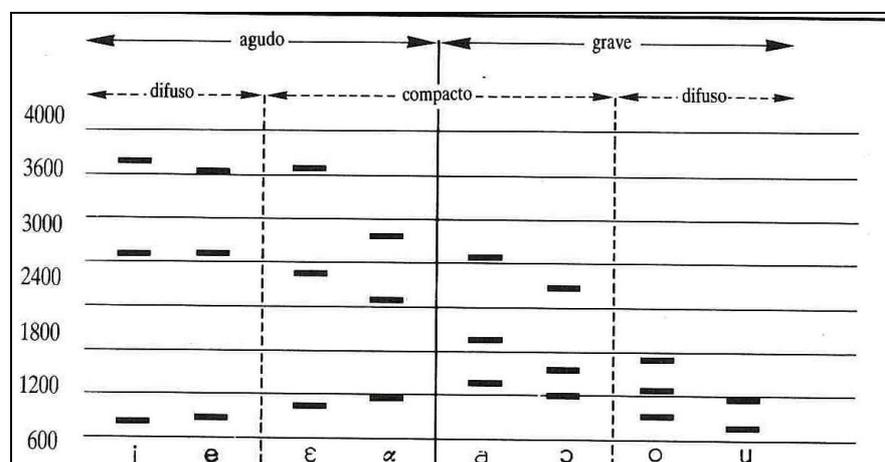


Figura 23. Classificação das vogais orais do Português (Martins, 1988).

As vogais compactas são aquelas que apresentam as duas formantes ( $F_1$  e  $F_2$ ) em zonas centrais do espectro, tal como [  $\alpha$  ], [  $\varepsilon$  ], [ a ] e [  $\text{ɔ}$  ].

As vogais difusas apresentam as formantes fora desta zona central, tal como [ e ] e [ i ] têm  $F_1$  abaixo desta zona e  $F_2$  acima da mesma, e [ o ] e [ u ] têm os dois primeiros formantes abaixo da zona central.

O traço grave/ agudo distingue as vogais cujo segundo formante se situa numa zona de frequências baixa (por exemplo, [ a ], [  $\text{ɔ}$  ], [ o ] e [ u ]), e das vogais cujo segundo formante se situa numa região de frequências alta (por exemplo, [  $\alpha$  ], [  $\varepsilon$  ], [ e ] e [ i ]).

### 5.3 FONÉTICA PERCEPTIVA

A fala é um contínuo acústico que para a sua compreensão exige não só a audição, como também a sua percepção (Martins, 1988). A percepção integra todos os processos que possibilitam a compreensão do sinal da fala em significado linguístico a nível do sistema nervoso central, existindo várias hipóteses possíveis de processamento de informação (Quadro 5).

**Quadro 5.** Hipóteses possíveis de processamento de informação (Martins, 1988).

Hipótese	Descrição
<b>“Bottom-up”</b> (Linear de baixo para cima)	Processo linear do aspecto fonético para o aspecto semântico e que segue a análise em níveis progressivamente mais complexos.
<b>“Top-down”</b> (Linear de cima para baixo)	O processamento da fala pode processar-se no sentido inverso, ou seja, ouve-se por exemplo uma frase e compreende-se globalmente, fazendo depois a análise dos seus componentes.
<b>Interactivo</b>	O processamento pode fazer-se de forma não linear, ou seja, interactivamente ou em círculo aberto ou fechado. Ou seja, pode-se compreender uma palavra, não compreender o sentido da frase, recorrer a outra palavra ou a seguir à análise sintáctica ou até compreender o sentido geral da frase ouvida.

O modelo de «acesso lexical», de Klatt, propõe a realidade psicológica do segmento fonético como uma unidade de reconhecimento no processo “bottom-up” de formação lexical.

A teoria motriz surge a partir dos anos 60, e foi desenvolvida pelos Laboratórios Haskins, que segundo esta, existe mediação da sensação dos movimentos articulatorios entre o sinal acústico e a memória auditiva (nível neuromotor).

Posteriormente, Stevens e House desenvolveram a teoria da percepção de análise por síntese (semelhante à anterior) situando a associação auditiva-articulatória ao nível neuroacústico.

A partir desta teoria, Klatt, através de uma complementação com o modelo de acesso lexical, propõe um modelo computacional capaz de simulá-lo.

Em 1971, foi desenvolvido por Linbdlom e Sundberg, outro modelo de percepção – o modelo de programação por simulação. A ideia base deste modelo é que o sistema consiga seleccionar a resposta motora apropriada com maior flexibilidade e com maior elasticidade, permitindo a aprendizagem a partir dos seus próprios erros simulados, antes de decidir.

#### **5.4 BREVES NOÇÕES DE FONOLOGIA**

A fonologia descreve e explica o funcionamento das unidades significativas da fala, ou seja, estuda a segunda articulação de uma língua (Barbosa, 1994).

A segmentação dos enunciados (descrição global de uma língua) em unidades mínimas significativas (monemas), e posteriormente a segmentação dos significantes dos monemas em unidades distintivas mínimas e sucessivas denominam-se de fonemas.

Portanto, um fonema é uma unidade distintiva, mínima e sucessiva que serve para formar os significados dos monemas (Barbosa, 1994).

Por exemplo, o monema *gato* [ˈgatu] pode ser segmentado em quatro fonemas sucessivos [g], [a], [t] e [u], uma vez que todos eles podem ser comutados com outros. Ou seja, [g] é substituível por [b] (bato), ou por [p] (pato), assim como [a] é comutável com [o] (gato). Estas comutações dão lugar a outros significantes diferentes do primeiro.

Um par mínimo é um par de monemas cuja diferença reside apenas num segmento (fonema), como por exemplo, gato-pato ou pato-bato.

As características funcionais dos fonemas são as que permitem distinguir um fonema de outro, e determinam-se de acordo com uma pertinência linguística. Estas características funcionais são chamadas de traços pertinentes.

## **5.5 NOMENCLATURA PARA AS ARTICULAÇÕES VOCAIS E SONS USADOS NA FALA**

Os alfabetos fonéticos descrevem o modo como as palavras se representam graficamente. Para isso, a cada som corresponde sempre um símbolo e só esse símbolo. Embora a generalidade das palavras tenha um segmento, chamado de sílaba, esta pode ser pronunciada com maior intensidade do que os seus vizinhos, intensidade essa definida por acento ou acentuação.

Assim, nos alfabetos fonéticos a sílaba acentuada está marcada com um sinal de acento agudo, e seja qual for o timbre da vogal da sílaba acentuada, este acento é colocado antes do símbolo que representa o primeiro som da mesma.

Os parênteses rectos mostram que se trata de uma transcrição fonética. Os diacríticos [˘] e [˙] colocados sob o símbolo que representa uma vogal indica que esta é mais aberta ou menos aberta respectivamente.

O mais divulgado e completo dos alfabetos fonéticos é o International Phonetic Alphabet – IPA (Anexo A). Hoje, este alfabeto é essencialmente constituído por caracteres do alfabeto romano, acrescido de alguns diacríticos e outros símbolos, permitindo uma descrição completa de todas as línguas conhecidas.

Em 1987-89, um grupo internacional de foneticistas desenvolveram um alfabeto fonético denominado SAMPA (Speech Assessment Methods Phonetic Alphabet), destinado à aplicação nas línguas da Comunidade Europeia, mas só em 1993 foi aplicado à Língua Portuguesa (Anexo B). O SAMPA consiste basicamente no mapeamento de símbolos do IPA em códigos ASCII, e possui guias de orientação para as transcrições fonéticas.

Em Portugal, Aniceto Gonçalves Viana, foi o primeiro foneticista Português, autor de uma vastíssima bibliografia filológica, linguística e fonética. Armando de Lacerda foi fundador e director do Laboratório de Fonética Experimental da Faculdade de Letras da Universidade de Coimbra e da Revista desse laboratório, sendo autor ou co-autor de vários trabalhos de fonética.

Apesar destas obras serem muito minuciosas, a Revista Portuguesa de Filologia de Manuel Paiva Boléo é mais prática e a mais usada.

Na tentativa de abranger a variedade regional da Língua Portuguesa foi elaborado um alfabeto extenso com os símbolos e diacríticos necessários, e que constitui o Projecto do Atlas Linguístico do Centro de Linguística da Universidade de Lisboa.

Dada a diversidade de alfabetos, apresenta-se, por ser menos ambíguo, um subconjunto de símbolos do IPA elaborado Faria et al. (1996) para a classificação do PE (Figura 24).

(4)	IPA	ORTOGRAFIA	TRANSCRIÇÃO FONÉTICA
<b>VOGAIS</b>	i	fita	[ˈfitɐ]
	e	pera	[ˈperɐ]
	ɛ	seta	[ˈsetɐ]
	a	cara	[ˈkarɐ]
	ɐ	cama	[ˈkɐmɐ]
	ɨ	que; se	[ki], [si]
	ɔ	corda	[ˈkordɐ]
	o	mofo	[ˈmofu]
	u	mudo	[ˈmudu]
<b>SEMI-VOGAIS</b>	j	pai; piada	[pai], [ˈpjadɐ]
	w	pau; soalho	[paw], [ˈswaɫu]
<b>CONSOANTES</b>	p	papo	[ˈpapu]
	t	tia; fato	[ˈtiɐ], [ˈfatu]
	k	casa; baque	[ˈkazɐ], [ˈbaki]
	b	barba	[ˈbarbɐ]
	d	data; arde	[ˈdatɐ], [ˈardi]
	g	gato; mago	[ˈgatu], [ˈmagu]
	f	férias; bafo	[ˈferjeʃ], [ˈbafu]
	s	selo; caça; passa	[ˈselu], [ˈkase], [ˈpase]
	ʃ	chave; festas	[ˈʃavi], [ˈfeʃteʃ]
	v	vaca, cava	[ˈvakɐ], [ˈkavɐ]
	z	asa; azul; exacto	[ˈazɐ], [vˈzul], [iˈzatu]
	ʒ	jacto; agir; asma	[ˈzatu], [vˈʒir], [ˈazme]
	l	lado; sala	[ˈladu], [ˈsalɐ]
	ʎ	folha, alho	[ˈfoʎɐ], [ˈaʎu]
	r	caro; parva	[ˈkaru], [ˈparvɐ]
	R	raiva; palra; carro	[ˈrajvɐ], [ˈpalrɐ], [ˈkaru]
	m	marca; turma	[ˈmarkɐ], [ˈturmɐ]
	n	neta; cena	[ˈnetɐ], [ˈsenɐ]
	ɲ	unha; pinha	[ˈunɐ], [ˈpinɐ]

Figura 24. Classificação dos sons do Português Europeu (Faria et al., 1996).

O abaixamento do véu do palato faz com que o ar se escape também para a cavidade nasal, originando sons nasalizados; estes são assinalados por um til sobre o seu símbolo. No PE, para além das cinco vogais nasais, as consoantes oclusivas /m/, /n/ e /nh/ também são nasaladas, e ao contrário das consoantes orais são sempre vozeadas.

## CAPÍTULO 6. IMAGEM POR RESSONÂNCIA MAGNÉTICA

A IRM é um método de imagem que consiste em gerar imagens seccionais do corpo humano, tendo por base a interacção de um elevado campo magnético com pulsos de radiofrequência (RF) e com os núcleos dos átomos que constituem o corpo humano. É, por isso, uma técnica inócua e com elevada sensibilidade dado ter em conta os processos anatómo-fisiológicos dos átomos (Lufkin, 1999).

### 6.1 PRINCÍPIOS FÍSICOS

Os núcleos em IRM, para serem úteis, devem possuir um momento angular ou spin diferente de zero, ou seja, devem desenvolver um movimento de rotação sobre o seu eixo. Assim, na presença de um campo magnético, o núcleo em causa descreverá um movimento de acordo com este – precessão (similar ao movimento de um pião), condição necessária para que ocorra o fenómeno de ressonância. O núcleo de hidrogénio é a base da IRM, não só pela sua elevada abundância no corpo humano, mas também porque tem elevada sensibilidade magnética.

Após colocação do paciente num campo magnético externo, potente e uniforme, os prótons de hidrogénio tendem para se orientar:

- Paralelamente ao campo (estado de equilíbrio);
- Anti-paralelamente ao campo (estado de excitação);

Por questões de equilíbrio, a orientação paralela será preferencialmente superior, dando-se a formação de um vector (soma das orientações magnéticas) paralelo ao campo magnético, denominado de vector de magnetização efectivo (VME).

Posteriormente, é aplicado um pulso de RF de  $90^\circ$ , que introduz energia ao sistema, deslocando o VME para o plano transversal. Para que isto aconteça é necessário que este segundo campo esteja sincronizado com a frequência de precessão dos núcleos. A amplitude e duração do pulso de RF podem ser controladas para produzir diversos graus de angulações (*flip angle*) do VME, do plano longitudinal ao transversal.

Para além de introduzirem energia ao sistema, os pulsos de RF ( $90^\circ$  ou  $180^\circ$ ), fazem com que os prótons estejam em fase (coerência de fase), permitindo assim, a detecção do sinal através de antenas (o VME induz uma corrente na antena de acordo com a Lei de Faraday). Após cessar o

pulso de RF, a amplitude do sinal após ser detectado não permanece constante declinando rapidamente até zero (Figura 25).

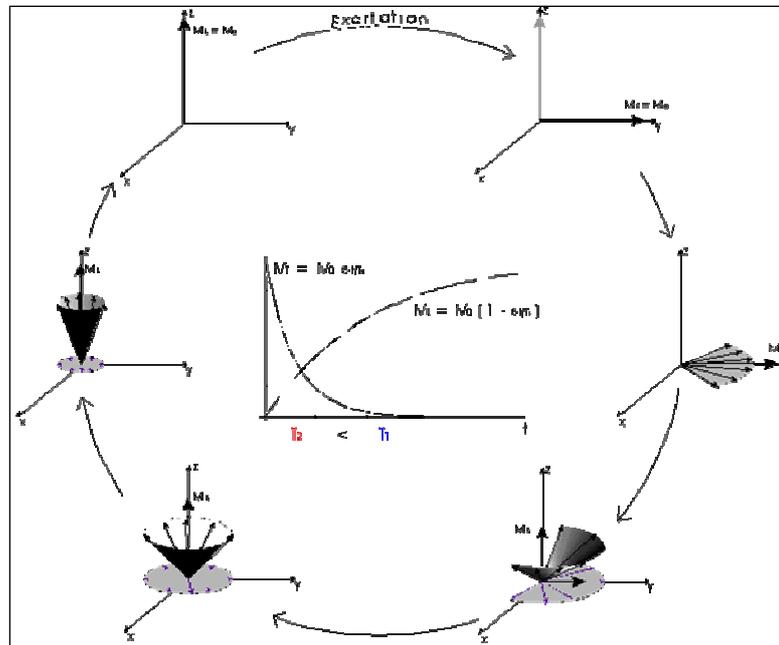


Figura 25. Esquema representativo da obtenção do sinal de IRM.

Fonte: <http://mri.if.sc.usp.br/paginas/port/tutorial/parte-04.html>.

Esta oscilação amortecida é conhecida por Free Induction Decay (FID). A amplitude do sinal FID é proporcional ao número de prótons presentes na amostra. O sinal FID é dado pelas diferenças de relaxação dos tecidos (que ocorre de forma natural), podendo ser de dois tipos:

- Relaxação longitudinal ou recuperação T1 – resulta da libertação de energia dos prótons excitados para o meio circundante, e depende da estrutura molecular, do estado da matéria e da potência de campo magnético. Moléculas com recuperação rápida têm um T1 curto (ex. gordura), e moléculas com recuperação lenta têm um T1 longo (ex. líquido céfalo-raquidiano). O tempo que medeia dois pulsos de excitação de RF é responsável pela quantidade de sinal em T1, e denomina-se de tempo de repetição (TR).
- Relaxação transversal ou recuperação T2 – resulta da troca de energia entre os núcleos, que devido aos seus pequenos campos magnéticos individuais, perdem a sua coerência de fase, irreversivelmente. Moléculas que têm um decaimento rápido, têm um T2 curto (ex. gordura), e moléculas com decaimento lento têm um T2 longo (ex. líquido céfalo-raquidiano). O tempo entre um pulso de excitação e a leitura do sinal controla a quantidade de decaimento T2, e denomina-se de tempo de eco (TE).

O contraste da IRM vai ser definido pelo TR e pelo TE.

### 6.1.1 Sequências em IRM

As sequências em IRM definem um conjunto de parâmetros de aquisição, que por variação dos mesmos permitem a obtenção de diferentes imagens mediante as estruturas que se pretendem estudar. No estudo do tracto vocal não existe um protocolo *standard* pelo que apenas serão descritas as sequências mais usadas em investigação, revistas na literatura (Quadro 6).

**Quadro 6.** Sequências usadas em IRM.

Sequência	Descrição sumária
<b>Spin-eco</b>	Esta sequência é obtida com um pulso de excitação inicial de 90°, seguido de um ou mais pulsos de 180° para refasamento dos prótons. Cada pulso de refasamento origina a produção de um eco (sinal), sendo que a produção de multi-ecos destina-se apenas a avaliar a evolução do sinal das estruturas em diferentes TE (ponderações). Aplicam-se a quase todo o tipo de exames, fornecendo imagens com elevado detalhe anatómico nas imagens ponderadas em T1 (pois favorecem os tecidos que recuperam rapidamente, como a gordura), e também detalhe fisiológico nas imagens ponderadas em T2 (pois favorecem os tecidos que decaem lentamente, como os líquidos).
<b>Turbo spin-eco ou fast spin-eco</b>	Esta sequência tem o mesmo princípio que a anterior, diferenciando-se desta última, pela aplicação de um conjunto de pulsos de refasamento de 180°, de forma a produzir o correspondente número de ecos. A este conjunto de pulsos dá-se o nome de <i>ecotrain</i> ou <i>factor turbo</i> (ETL). Habitualmente são produzidos quatro a dezasseis ecos, o que permite uma redução do tempo de sequência. Estas sequências têm vindo a substituir as sequências spin-eco convencionais, devido aos tempos de aquisição serem substancialmente reduzidos. Têm principal aplicação nos estudos cerebrais e da coluna vertebral, pela sua rapidez e elevada ponderação T2.
<b>Gradiente-eco</b>	Estas sequências utilizam pulsos de excitação com <i>flip angles</i> (FA) inferiores a 90°, o que condiciona a qualidade de imagem (menor amplitude de sinal). A restituição de fase (ou refasamento) é efectuada pela aplicação de um grande de campo magnético (variação linear da força de campo), ao contrário das sequências spin-eco nas quais é efectuada o refasamento pela aplicação de pulsos de RF de 180°. A aplicação de gradientes não compensa os distúrbios na homogeneidade do campo magnético, podendo desta forma surgir artefactos de susceptibilidade magnética (perda de sinal nos tecidos). A sequência turboflash em gradiente-eco baseia-se na redução do TE/TR dos gradientes, permitindo tempos de aquisição de imagens na ordem de um segundo. Como desta redução resulta uma perda considerável do contraste da imagem, torna-se necessário a aplicação de pré-pulsos de RF antes da aquisição, preparando assim os tecidos. Dependendo dos fabricantes pode assumir diferentes terminologias GRASS, FISP ou FLASH. Devido ao seu carácter ultra-rápido, tem como principais aplicações clínicas, o estudo dinâmico das estruturas (como o coração, e o tracto vocal durante a fala) com elevada resolução temporal, e estudos em apneia de forma a superar os artefactos respiratórios e a obtenção de imagens de estruturas muito pequenas. Tem maior aplicação ao nível do estudo cardíaco, permitindo obter cine-imagens do coração.
<b>Sequências ultra-rápidas (imagens ecoplanares)</b>	Este tipo de sequências utiliza apenas parte do pulso de RF para a excitação dos tecidos, e apenas parte da leitura do eco. Consequentemente, a rapidez aumenta possibilitando a obtenção de imagens em menos de 100ms (50 a 100 x 10 <sup>-3</sup> s). Devido ao grande esforço dos gradientes, provocam um intenso ruído acústico, por comparação às outras sequências. Tem como principais aplicações clínicas, o estudo da difusão e perfusão cerebral, o estudo funcional cerebral (o aumento do consumo de glicose durante a execução de uma tarefa motora, ou outra, conduz a um aumento do consumo de oxigénio, consequentemente a um aumento do sinal na região cerebral correspondente a essa actividade), estudos abdominais, a avaliação intra-uterina do feto (IRM fetal), e também constitui um bom recurso para estudo simultâneo da morfologia e função cardíaca.

## 6.2 INSTRUMENTOS EM RESSONÂNCIA MAGNÉTICA

A IRM tem como principais componentes:

- Campo magnético principal;
- Gradientes magnéticos geradores de campo;
- Transmissor de RF e/ou receptor do sinal (antenas);
- Sistema de suporte eléctrico (fornece todo o suprimento eléctrico ao sistema);
- Sistema de refrigeração;
- Consola e workstation de trabalho.

O campo magnético principal tem como função principal, a produção de um campo magnético estático, uniforme, sobre o qual se sobrepõem os gradientes e os pulsos de RF necessários para a obtenção de imagens. O campo magnético pode ser caracterizado segundo o seu tipo (permanente, resistivo ou supercondutor), potência, custo, uniformidade temporal, tamanho físico e produção de campos marginais.

O magnete supercondutor necessita de bobinas para gerar o seu campo magnético. A refrigeração é mantida com hélio líquido, para se manter o estado supercondutor. As linhas de campos magnéticos marginais são significativas, acarretando efeitos significativos de risco de projecteis em relação a objectos ferromagnéticos. Os magnetes supercondutores têm custos operacionais elevados, e são geralmente mais caros que os outros tipos. A vantagem efectiva é que permitem a obtenção de imagens de elevada intensidade (campos acima de 1,0T).

Os gradientes de campo têm como finalidade a produção de variações lineares no campo magnético principal e a localização espacial do sinal de RM. Através da combinação de três sistemas de gradientes (gradiente de selecção do corte, codificação de fase e codificação de frequência) é possível a obtenção de imagens em três planos ortogonais. Caracterizam-se pela sua amplitude (mT/m), tempo de elevação, *slew rate* (razão entre a amplitude máxima e tempo de elevação – T/m/s), e ciclo de tarefa (tempo de activação do gradiente durante um período TR).

As antenas permitem a transmissão de pulsos de RF e/ou a recepção dos ecos. Podem ser divididas em três tipos, mediante o seu campo de visão e relação sinal-ruído:

- Antenas de superfície (têm um campo de visão reduzido, mas como se adaptam à região em estudo possuem elevada relação sinal-ruído);
- Antenas de volume (têm um campo de visão grande, e uma menor relação sinal-ruído);

- Antenas Phased Array (série de antenas montadas num conjunto para a geração de imagens em volume);

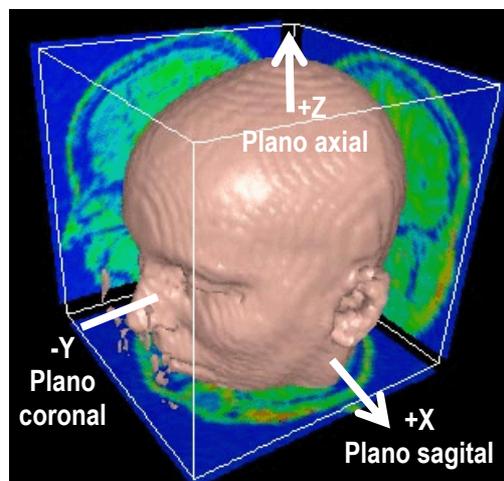
As antenas Phased Array associam o conceito de antenas de superfície (elevada relação sinal-ruído) e elevado campo de visão das antenas de volume. Como se encontram montadas por um conjunto de sub-antenas ligadas independentemente ao sistema de aquisição, podem ser activadas por segmentos.

### 6.2.1 Formação da imagem

O sinal de ressonância magnética, captado pelas antenas, vai ser localizado e codificado através dos gradientes, no espaço K (representação dos dados brutos); sendo posteriormente reconstruído através da transformada de Fourier em imagem.

A acção dos gradientes é feita em três etapas distintas, cada uma das quais codifica um eixo espacial diferentes do corte: (1) gradiente de selecção do corte (activado durante a aplicação do pulso de  $90^\circ$ ), (2) codificação de fase (activado e desactivado na ausência de qualquer estimulação de RF), e (3) codificação de frequência (activado durante a leitura do eco).

O gradiente de selecção do corte permite também determinar a orientação do plano de exame. Existem três gradientes físicos em IRM designados por  $G_x$ ,  $G_y$  e  $G_z$ , orientados ortogonalmente no espaço (Figura 26). Assim, para obter cortes sagitais, o gradiente de selecção de corte corresponderá à activação do gradiente  $G_x$ , sendo os gradientes  $G_y$  e  $G_z$  seleccionados para a codificação de fase e de frequência, respectivamente.



**Figura 26.** Localização espacial do sinal de RM, e sua correlação com os planos anatómicos.

**Fonte:** Imagem modificada de <http://www.mathforum.jp/recommend/microavs/head.gif>

### **6.3 IMPORTÂNCIA DA IRM EM SÍNTESE ARTICULATÓRIA E EM MEDICINA**

A IRM tem vindo a impor-se como a técnica de imagem mais promissora no estudo do tracto vocal. Para além da sua inocuidade, permite um estudo anatómico (morfológico), e multiplanar de todo o tracto vocal. Constitui uma ferramenta inigualável no estudo imagiológico do tracto vocal, pois permite aliar a informação anatómica (forma e relação dos órgãos) ao mecanismo da fala possibilitando a medição correcta da forma do tracto vocal durante a fonação. Este conhecimento, parte fundamental no estudo da produção da fala e do seu controlo, é útil na síntese da fala para construção de modelos articulatórios (Teixeira et al., 2001).

Vários estudos têm demonstrado a importância da IRM neste âmbito, destacando-se:

- Possibilidade de realizar medições nas imagens sagitais para obtenção de funções de área;
- Contribuição para a reavaliação da teoria acústica;
- Elucidação dos factores envolvidos nas individualidades da fala;
- Obtenção de modelos a três dimensões.

A IRM permite o estudo estático (através da sustentação de sons) com uma boa visualização da língua, véu do palatino, lábios e laringe e, actualmente, dadas as novas sequências de aquisição ultra-rápidas, é possível o estudo dinâmico (mediante sincronização).

Em Medicina, esta técnica permite avaliar défices articulatórios pré e pós-cirurgia de pacientes (para tratamento de tumores), assim como o estudo de processos compensatórios (Mády et al., 2001; Mády et al., 2001). A IRM permite também o estudo da cavidade faríngea para despiste do síndrome da apneia do sono (Abbott et al., 2003).

No entanto, residem algumas limitações subjacentes à técnica, como a impossibilidade de visualização dos dentes, e algumas dificuldades na identificação da fronteira ar-tecido. Os maiores desafios que se impõem na IRM são:

- O uso de sequências cada vez mais rápidas, de forma a evitar artefactos provocados por movimentos respiratórios, deglutição e movimento durante a fala, e a obtenção de imagens dinâmicas em pacientes pouco colaborantes;
- Diminuição dos artefactos de susceptibilidade magnética (tipicamente ocorrem nas sequências gradiente-eco, pois o refasamento dos prótons é feito à custa dos gradientes de campo magnético).

## **PARTE 2**

# **ESTUDO EXPERIMENTAL**

## CAPÍTULO 1. INTRODUÇÃO

Após revisão da literatura pode-se constatar que não existe uniformidade entre os diferentes estudos, quer no que diz respeito à técnica de IRM, quer às técnicas de processamento de imagem, o que é justificável dado que a fala é um mecanismo complexo e, acima de tudo, individual de cada um. No entanto, quanto à técnica de IRM é inequívoca a necessidade de utilização de sequências cada vez mais rápidas, de forma a possibilitar não só o estudo estático (com menor esforço dos sujeitos) como o estudo dinâmico. Para estes fins, é crucial a utilização de equipamentos de ressonância magnética com, pelo menos, um campo magnético de 1,5T ou acima, mas cuja acessibilidade está condicionada ao número de equipamentos existentes. Quanto às técnicas de processamento de imagem, a sua utilização tem vindo a ser, no sentido de, tornar o processo de segmentação, o mais automático possível, contribuindo assim para um menor tempo e esforço dos investigadores.

Assim, o presente estudo experimental foi desenvolvido simultaneamente em duas vertentes: (1) numa fase de treino e exploração da IRM aplicada ao estudo da fala, e (2) numa fase de pesquisa e exploração das técnicas de processamento de imagem mais adequadas aos objectivos propostos. A fase de treino e exploração da IRM decorreu na unidade de ressonância magnética, do serviço de Radiologia do Hospital S. João, no Porto, após autorização prévia da Directora de Serviço, e com a sua colaboração e de toda a equipa técnica. Consistiu em duas fases, inicialmente procedeu-se à realização de um estudo estático dos sons do Português, e posteriormente, à realização de um estudo dinâmico de três sílabas. Quanto à fase de pesquisa e exploração das técnicas de processamento de imagem desenvolveu-se, fundamentalmente, com base na pesquisa bibliográfica (documental e electrónica).

De seguida, apresentam-se os objectivos deste estudo experimental, bem como o material e métodos utilizados.

### 1.1 OBJECTIVOS

Dada a escassez de informação morfológico-dinâmica da fala, o principal objectivo deste estudo foi o de reunir a informação necessária para a caracterização da maioria dos sons da fala do PE com base na IRM. Para tal, delinearam-se os seguintes objectivos específicos:

- Caracterizar a fala, inicialmente estudando estaticamente as vogais e a maioria das consoantes do Português, para a criação de modelos tridimensionais dos diferentes sons com a combinação dos diversos cortes, organizados numa base de dados de natureza morfológica;
- Caracterizar dinamicamente a fala por IRM, observando as variações temporais dos diferentes articuladores em diferentes sílabas.

## **1.2 MATERIAL E MÉTODOS**

### 1.2.1 Caracterização da amostra

A amostra deste estudo foi constituída por cinco sujeitos jovens adultos (dois do sexo masculino e três do sexo feminino), voluntários, saudáveis e sem história prévia de patologia orofacial, com idades compreendidas entre os 20 e 26 anos (média de  $22 \pm 2,35$  anos) e um peso médio de  $63,8 \pm 10,03$  Kg. Intencionalmente, foram incluídos na amostra sujeitos com formação na área de Terapia da Fala (quatro dos sujeitos), não só pelo interesse e familiaridade destes com o tema, mas porque o tempo dispendido no treino e explicação do *corpus* da fala é menor. Apenas um sujeito (do sexo feminino), sem formação na área, foi previamente treinado para o efeito.

Para avaliação da “normalidade” das estruturas orofaciais dos sujeitos da amostra em estudo, foi utilizado um protocolo de avaliação orofacial (PAOF), que é um instrumento que permite a avaliação de estruturas anatómicas e funções fisiológicas que estão potencialmente relacionadas com perturbações da comunicação oral (Anexo C). Este protocolo foi criado fundamentalmente para terapeutas da fala com o objectivo de ser um contributo para o diagnóstico e um guia de orientação para o planeamento da intervenção, e está organizado em duas grandes áreas: a morfologia e a função. Neste estudo em particular, aplicou-se apenas parte do PAOF, tendo em conta apenas os parâmetros de avaliação morfológica, adaptando-se a folha de registo deste protocolo. O recurso a este instrumento de avaliação teve como finalidade a obtenção de informação relativa a qualquer alteração ou observação que pudesse influenciar a informação morfológica obtida por IRM, ou até complementar.

Assim, adoptou-se a seguinte sequência de avaliação:

- Observação global directa da face de frente e de perfil de acordo com os diferentes planos, para avaliação da morfologia da face (p.2 e 3 do PAOF, anexo 4);
- Observação da aparência global dos lábios, bem como do freio labial superior, para avaliação da morfologia dos lábios (p.4 do PAOF, anexo 4);
- Observação da coloração, tamanho e comprimento da língua, e observação do freio lingual, para avaliação da morfologia da língua (p.10 do PAOF, anexo 4);
- Registo da ausência de dentes, para avaliação da dentição (p.8 do PAOF, anexo 4).

Existe uma grande variabilidade na função normal dos articuladores, não existindo por isso, um método padronizado para determinar o valor adequado da funcionalidade, pelo que a validade da cotação depende fortemente do avaliador. Neste sentido, esta avaliação foi efectuada com o auxílio de um Terapeuta da Fala com interesse no tema.

Os resultados são cotados com o valor de um, desde que estejam de acordo com os itens pontuáveis conforme quadro 7.

**Quadro 7.** Avaliação orofacial aplicada aos sujeitos da amostra.

Avaliação:	Itens pontuáveis	Cotação	
		Esperada	Obtida
1- Morfologia da face – frontal	Ortognatia	2	
Morfologia da face – perfil	Ortognatia		
2- Morfologia dos lábios		3	
Lábio superior	Adequado		
Freio do lábio superior	Adequado		
Lábio inferior	Adequado		
3- Morfologia da língua		4	
Coloração	Rosada		
Tamanho	Adequado		
Comprimento	Adequado		
Freio lingual	Adequado		
	Total	9	

Relativamente à adequação /alteração da morfologia das estruturas orofaciais, os resultados são apresentados pela correspondência entre o valor esperado e o valor obtido. Para a amostra em estudo obtiveram-se os seguintes resultados, três dos sujeitos obtiveram uma cotação de oito (corresponde a 88,8% do valor esperado); e dois indivíduos obtiveram a cotação máxima esperada. Nenhum dos sujeitos apresentava distúrbios articulatorios (critério de exclusão), pelo que as alterações observadas nalguns sujeitos eram mínimas não afectando, por isso, a produção dos sons.

Desta amostra, apenas dois sujeitos (do sexo masculino e feminino) participaram no estudo estático, enquanto que para o estudo dinâmico participaram quatro sujeitos (dois do sexo masculino e dois do sexo feminino).

### 1.2.2 *Corpus* da fala

O *corpus* da fala consistiu, inicialmente, num conjunto de vinte e cinco sons do PE, constituído pelas vogais orais e nasais (pela facilidade na sua produção) e, pelas consoantes fricativas, consoantes laterais e oclusivas nasais (pois é possível a manutenção prolongada do seu ponto e modo de articulação). Estes sons foram sustentados por dois sujeitos, durante a aquisição das imagens do estudo estático.

No estudo dinâmico, foi solicitado a quatro dos sujeitos da amostra, a repetição sucessiva, durante cada aquisição de imagens, de três sílabas consoante-vogal (/tu/, /ma/ e /pa/). As razões subjacentes à escolha deste *corpus* para o estudo dinâmico estiveram relacionadas com:

- No caso da sílaba /pa/ – a possibilidade de estudar mais uma consoante (/p/) e porque a vogal /a/ não influencia ou modifica esta consoante (por coarticulação);
- No caso da sílaba /ma/ – para permitir a comparação com a sílaba anterior, em contextos articulatorios semelhantes, dado que a consoante /m/ é uma consoante oclusiva nasal, enquanto que o /p/ é uma consoante oclusiva oral;
- No caso da sílaba /tu/ – não só para estudo dessa consoante, como também porque estes sons assumem posições articulatorias completamente diferentes, tornando, no estudo dinâmico, a modificação do tracto vocal mais evidente.

Com este *corpus* tentou-se abranger o maior leque possível de posições articulatorias.

### 1.2.3 Equipamento utilizado e procedimentos gerais

Para a recolha das imagens do tracto vocal foi utilizado um equipamento de ressonância magnética Siemens Magnetom Symphony ® com um campo magnético de 1,5T e, foi usada uma antena Phased Array de cabeça para recepção do sinal. Este equipamento de ressonância magnética possui um eficiente sistema de gradientes com uma amplitude acima de 30 mT/m, um *slew rate* acima de 125T/m/s, e que, segundo o fabricante, permite a aquisição de um FOV máximo de 50cm. Apresenta várias funcionalidades e permite a aquisição de sequências ultra-rápidas possibilitando a realização de estudos cardíacos, angiográficos, tridimensionais, assim como outras aplicações clínicas como a espectroscopia, aplicações neurológicas, quantificação de fluxo, entre outras.

Antes de qualquer ensaio foi garantido o consentimento informado de cada sujeito relativamente ao estudo a realizar, bem como foi solicitado um compromisso escrito de cada um quanto à sua disponibilização e aceitação (Anexo D).

Os sujeitos foram posicionados em decúbito dorsal, com a cabeça devidamente imobilizada, e foram colocados auscultadores, de forma a facilitar também a intercomunicação com os mesmos, devido ao elevado ruído provocado pelo equipamento.

O registo e inserção do peso de cada sujeito, na workstation do equipamento de IRM, é mandatório, para permitir o cálculo do limite de absorção de RF (SAR) pelos tecidos. No final de cada ensaio procedeu-se ao registo destes valores para cada sujeito, de acordo com os limites impostos pela Food and Drug Administration (FDA).

Também foi garantida a confidencialidade de cada sujeito, bem como a sua segurança (através da pesquisa de eventuais contra-indicações da IRM) e acompanhamento, não tendo sido administrados quaisquer fármacos. Os ensaios, com base na IRM, foram efectuados mediante a supervisão do Médico Radiologista e seguindo as normas operacionais do serviço.

Durante os ensaios, todos os procedimentos (como o registo do ensaio, o tipo de estudo, o tipo e parâmetros de aquisição e sons estudados) e principais dificuldades ou conclusões observadas foram registadas em documento próprio, elaborado pela autora para o efeito (Anexo E).

A recolha do sinal acústico da fala não foi possível devido ao intenso ruído presente na sala, provocado, não só pelo sistema de refrigeração da bobina do campo magnético, mas principalmente pela activação e desactivação dos gradientes durante a obtenção das imagens. De acordo com a sequência de imagem utilizada a intensidade de ruído é variável, sendo que as sequências ultra-rápidas tendem a originar um ruído mais intenso. Ainda assim, foi realizada a tentativa de gravação

do sinal da fala através do intercomunicador, que se revelou pouco efectiva ou mesmo inútil. Ponderou-se ainda a colocação de um microfone específico ou outro material (por exemplo um tubo) não ferromagnético (devido à presença de campo magnético intenso) o mais próximo possível do tracto vocal dentro da antena, mas que, por questões de segurança (pois obrigava à colocação do gravador fora da sala mantendo a porta da sala aberta) e/ou dificuldades na aquisição de material com essas características ficou inviabilizada.

#### 1.2.4 Técnicas de análise de imagem utilizadas

A análise das imagens consistiu em duas etapas distintas, primeiro, na segmentação das imagens usando o *plugin* 3D editing do programa de processamento de imagem Image J <sup>1</sup>, e posteriormente, na sua reconstrução tridimensional. A representação gráfica e combinação dos modelos tridimensionais de cada *stack* foram possíveis graças a um software para criação de gráficos e animações com aplicações interactivas 3D, denominado Blender <sup>2</sup>.

A funcionalidade e interpretação destes softwares de processamento de imagem, é descrita mais adiante no Capítulo 3.

---

1 Desenvolvido no National Institutes of Health. Disponível em: <http://rsb.info.nih.gov/ij>.

2 Disponível em: <http://www.blender.org/cms/Home.2.0.htm>.

## CAPÍTULO 2. FASE DE TREINO E EXPLORAÇÃO DA IRM

Esta fase de treino e exploração da IRM esteve condicionada, quer pelas características inerentes ao próprio equipamento, quer porque, até à data, não existia nenhum protocolo de exame dirigido especificamente ao estudo do tracto vocal. Desta forma, teve-se em atenção, para a definição do protocolo, algumas das recomendações sugeridas na literatura revista, quanto à utilização da IRM para estudo do tracto vocal:

- Deve existir um compromisso entre a resolução de imagem e o tempo de aquisição (Demolin et al., 2000; Engwall, 2004);
- Quanto à articulação dos sons: pode ser obtida de duas formas: (1) através da repetição sucessiva de um som durante a aquisição, e posterior reconstrução (Shadle et al., 1999; Mathiak et al., 2000); (2) através da sustentação de um som durante toda a aquisição com ou sem paragens para respiração do indivíduo (Demolin et al., 1998; Engwall, 2000; Soquet et al., 2002). A paragem para respiração do indivíduo pode levar à produção de artefactos e à modificação da articulação. Sem paragens para respiração corre-se também o risco de ocorrer artefactos por fadiga do indivíduo;
- O número e a espessura de cortes resultam do compromisso entre a resolução de imagem e o tempo de aquisição. Quanto maior o número de cortes maior é a resolução, mas em contrapartida o tempo de aquisição aumenta. Para além disso, quer a espessura quer o número de cortes dependem da região em estudo;
- Quanto à orientação dos cortes: a exactidão na extracção de contornos é maior se o corte for perpendicular à região em estudo;
- Quanto ao *corpus* da fala: resulta do compromisso entre o investigador e o conforto /bem-estar do indivíduo. Quanto maior o número de sons estudados maior é a fadiga do indivíduo;
- Relativamente ao plano em estudo: deve ter em conta a região em estudo. Os cortes sagitais fornecem informação completa de todo o tracto vocal, da configuração de todos os articuladores envolvidos na produção da fala e da sua extensão. Constituem, segundo Di Girolamo et al. (1996) a melhor modalidade no estudo dos padrões durante a fala. Os cortes sagitais têm sido amplamente usados em investigação. Os cortes coronais permitem não só complementar a informação dimensional dos cortes sagitais, como

também permitem um estudo dos lábios (posição e forma), e da língua. Os cortes axiais têm apenas utilidade no estudo da faringe, para além de que o processo de segmentação é mais árduo pela subjectividade na interpretação dos mesmos;

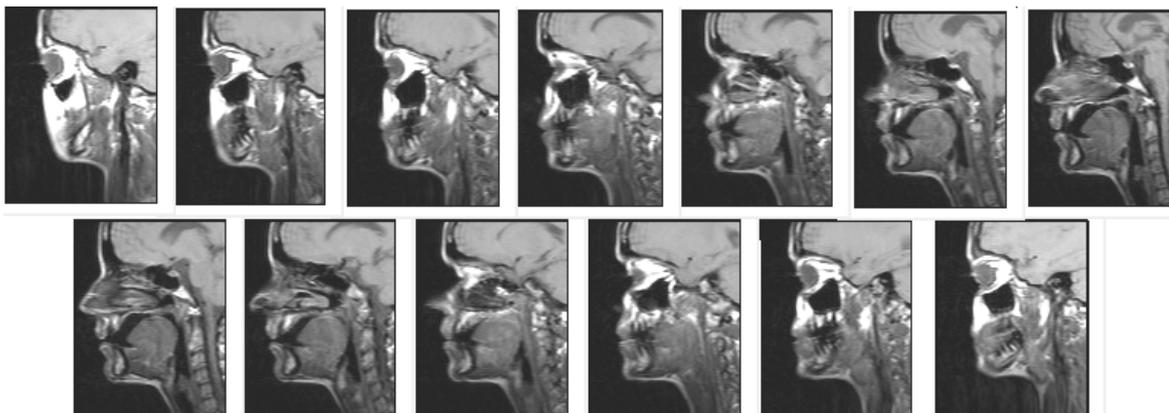
- Relativamente ao tempo de sustentação dos sons: segundo Engwall (2003) devem usar-se tempos de aquisição mais curtos nos estudos estáticos, ou se possível, deve optar-se preferencialmente, pela IRM em tempo real;
- Quanto à postura corporal: vários estudos (Engwall, 2003; Kitamura et al., 2005) demonstram que a postura corporal influencia alterando ou deformando os tecidos moles e estruturas ósseas de acordo com a força da gravidade, principalmente se os sons forem sustentados. No entanto, a posição de decúbito dorsal dos indivíduos, exigida nos equipamentos de ressonância magnética actuais com campos magnéticos elevados, tem sido amplamente usada, com algumas vantagens, como uma maior estabilidade e imobilização da região em estudo;
- Quanto ao *field of view* (FOV): deve ser ajustado de acordo com a região em estudo, normalmente entre 15 a 20cm. No entanto, de forma a diminuir os artefactos (de aliasing e movimento) nos estudos dinâmicos recomenda-se o uso de FOV maiores (30cm) (Narayanan et al., 2004);
- Sequências ultra-rápidas: apresentam mais artefactos que as ditas convencionais. Podem ser obtidas com base no refasamento por gradientes (estão desaconselhadas devido ao elevado número de artefactos pois são mais susceptíveis), ou no refasamento por RF pois apresentam maior resolução (no entanto, contribuem para uma maior absorção de RF pelos tecidos).

## **2.1 ESTUDO ESTÁTICO**

Neste estudo procedeu-se, inicialmente, à realização de experiências com apenas um dos sujeitos para a definição e adequação de um protocolo. No entanto, de forma a permitir o ajuste do tempo de aquisição das sequências, procedeu-se à gravação do tempo médio de sustentação, de alguns dos sons do *corpus* para cada um dos sujeitos (valor médio medido de  $12,5 \pm 0,71$ s).

A primeira experiência realizada consistiu na obtenção de treze cortes sagitais com 5mm de espessura (sem intervalo entre cortes), de forma a cobrir todo o tracto vocal, durante a sustentação prolongada da vogal /u/ com uma duração total de 33s. O maior inconveniente encontrado, para

além do elevado esforço exigido ao sujeito, observou-se durante o processo de segmentação; isto porque apesar do número elevado de cortes obtidos apenas foi possível a delimitação de contornos em seis desses cortes (Figura 27). Além disso, após reconstrução tridimensional a dimensão lateral apresentada, como o arredondamento dos lábios e a dimensão da cavidade oral não corresponde de facto, à realidade morfológica do tracto vocal (descrito mais adiante no Capítulo 4).



**Figura 27.** Estudo sagital completo da vogal /u/ sustentada por um sujeito (sexo feminino).

Para além de parte das imagens obtidas não apresentarem qualquer utilidade para segmentação do tracto vocal, verificou-se a necessidade de inclusão da imagem dos dentes no processo de segmentação, para permitir uma determinação mais fiel dos contornos; dada a baixa intensidade de sinal da região dentária (similar à região de interesse) não era possível a sua separação do tracto vocal (como vários estudos já haviam documentado). Assim, optou-se pela aquisição de uma imagem médiosagital de referência com o tracto vocal em posição de repouso (lábios e língua em contacto com os dentes), antes de iniciar o estudo estático (para cada um dos sujeitos), com maior detalhe anatómico, para permitir a identificação e extracção da imagem dos dentes e, posteriormente a sua adição (Figura 28).



**Figura 28.** Imagem médiosagital de referência, para identificação da imagem dos dentes (contornos a branco).

Dadas as limitações do estudo anteriormente descrito, com um número elevado de cortes num só plano, optou-se então pela combinação de cortes representativos nos três principais planos anatómicos, de forma a abranger as principais estruturas do tracto vocal, obtendo-se:

- Três cortes no plano sagital – para visualização de todo o tracto vocal;
- Quatro cortes no plano coronal – para visualização dos lábios e, principalmente da língua;
- Quatro cortes no plano axial – para visualização da faringe.

No entanto, a combinação e o registo espacial dos diferentes cortes, revelou-se um processo bastante complicado, mesmo tendo em conta a interpretação dos dados das imagens obtidas, em formato DICOM (abordado mais adiante no ponto 2.3). Posteriormente, os cortes axiais foram retirados do protocolo, principalmente pelas dificuldades verificadas no processo de segmentação.

Assim, foram obtidos dois conjuntos de imagens que representam o volume em estudo – *stacks*, primeiro no plano sagital (três cortes contíguos com 5mm de espessura) e depois no plano coronal (quatro cortes espaçados 10mm entre si e com 6mm de espessura), durante 9s e 9,9s para cada som, respectivamente.

Antes do início de cada aquisição, o sujeito era informado acerca do som pretendido através do intercomunicador, sendo também dadas instruções para inspirar fundo, e para que após uma contagem crescente de um a três pela autora, o mesmo iniciasse a sustentação no número dois; o arranque da aquisição era efectuado pelo técnico, na contagem do número três. Este procedimento para além de servir para orientar os sujeitos, permitiu eliminar o tempo de reacção dos sujeitos no início de cada sustentação do som, pois aquando do arranque da aquisição o som já estava a ser produzido.

As imagens sagitais ponderadas em T1 foram obtidas usando uma sequência Turbo Spin-eco, com os seguintes parâmetros de aquisição: TR = 443ms, TE = 17ms, FA = 150°, ETL = 7, matriz de 128x128, relação sinal-ruído = 1 e um FOV = 150mm. As imagens coronais ponderadas em T1 obtiveram-se usando uma sequência Turbo Spin-eco, com os seguintes parâmetros de aquisição: TR = 470ms, TE = 15ms, FA = 150°, ETL = 7, matriz de 128x128, relação sinal-ruído = 1,03 e um FOV = 150mm (Anexo F).

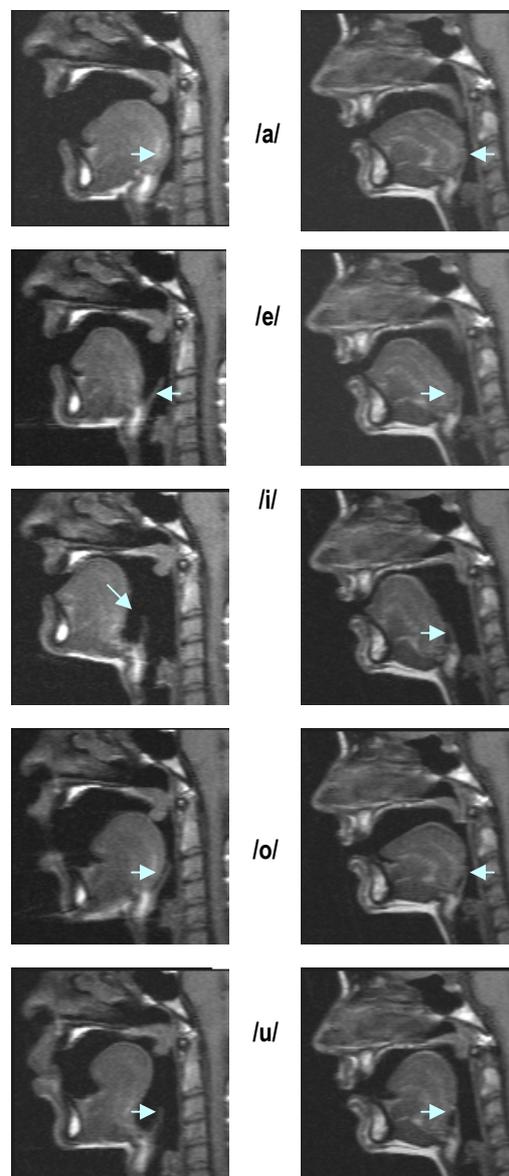
Para a correcta combinação das *stacks* (plano sagital e coronal) teve-se em atenção um conjunto de parâmetros na fase de programação das sequências para os dois planos, nomeadamente:

- Usar matrizes iguais;
- Não usar interpolação entre cortes;
- Usar um FOV com as mesmas dimensões e sem redução na codificação de fase;

- O centro de cada corte é definido por três coordenadas, cujos eixos foram similares no que diz respeito à orientação de cada conjunto de cortes e coincidentes entre planos.

De seguida, apresentam-se algumas das imagens obtidas neste estudo estático, nos dois planos de imagem, realçando-se também qual a sua relevância no estudo da produção da fala.

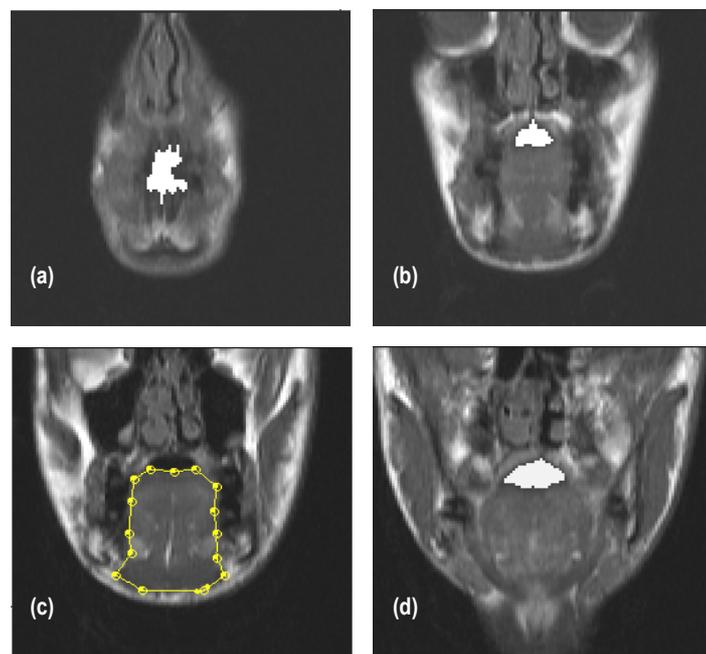
Como se pode verificar na figura 29, as imagens no plano sagital têm especial interesse, pois permitem o estudo anatómico completo de todo o tracto vocal, da forma e posição da maioria dos articuladores envolvidos na produção da fala, como a língua, véu do palato e lábios.



**Figura 29.** Imagens médiosagittais das vogais orais do Português sustentadas por um sujeito do sexo masculino (esquerda) e um do sexo feminino (direita).

Nestas imagens da figura 29 podem observar-se as diferenças nas dimensões do tracto vocal na sustentação de diferentes sons entre sujeitos de sexos diferentes. Apesar destas diferenças, os movimentos dos diferentes articuladores são similares, como por exemplo a posição da língua; esta, assume posições mais recuadas, centrais ou mais avançadas consoante a vogal. As projecções dos lábios e os graus de abertura da boca diferem, como pode ser observado, de vogal para vogal. A epiglote também surge bem individualizada, de acordo com a posição e forma da língua (setas).

As imagens coronais fornecem especificamente informação lateral do tracto vocal, sendo principalmente úteis no estudo da forma dos lábios e língua (Figura 30).

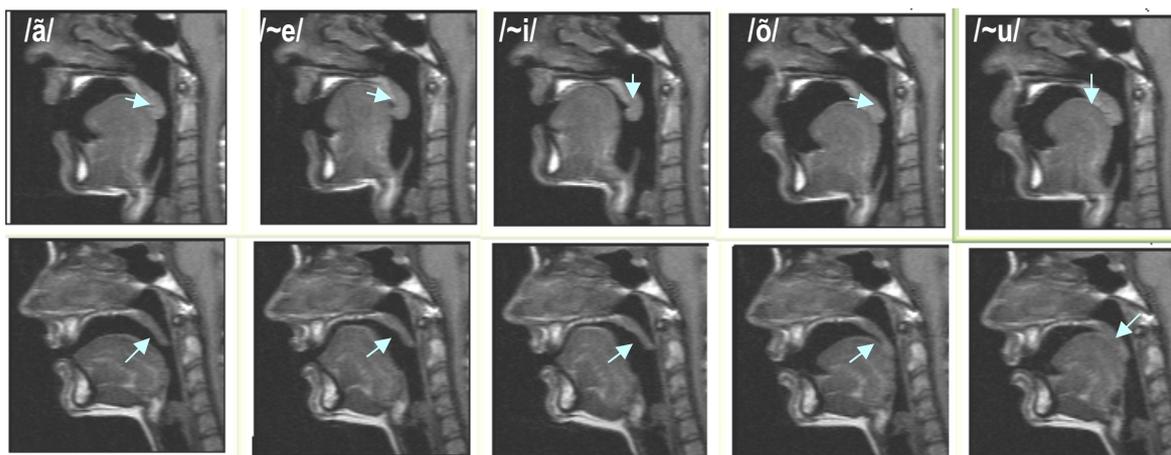


**Figura 30.** Stack de imagens coronais da consoante /ch/.

Nas imagens da figura 30, a área preenchida a branco pretende demonstrar o arredondamento dos lábios (30a) e a relação entre a língua e o palato (30b, 30d). A forma da língua é facilmente identificável, estando sinalizada o seu contorno com uma curva *bézier* (30c).

A nasalidade é produzida pelo abaixamento do véu do palato, permitindo a passagem de parte do ar expirado pela cavidade nasal, que funciona como uma cavidade anti-ressonante. A introdução desta característica faz com que o som seja alterado significativamente, tornando-o mais complexo. No Português, o conhecimento da nasalidade tem um interesse acrescido, dado o seu frequente uso na fala corrente.

A figura 31 ilustra um conjunto de imagens no plano médiosagital das vogais nasais do Português, obtidas no presente estudo estático, onde é visível o abaixamento do véu do palato e a sua relação com a língua (setas) nas sustentação dos diferentes sons.



**Figura 31.** Imagens médiosagittais das vogais nasais do Português, sustentadas por um sujeito do sexo masculino (acima) e um do sexo feminino (abaixo).

## **2.2 ESTUDO DINÂMICO**

Para a realização de qualquer estudo dinâmico é fulcral o recurso a técnicas de sincronização; no entanto, em IRM estas são apenas utilizadas, protocolarmente, no estudo do coração (por sincronização fisiológica com a actividade eléctrica do coração), ou para permitir a redução de artefactos provocados pela respiração (monitorização dos movimentos do tórax durante a respiração). Não existe na prática clínica, nenhum dispositivo específico que possa ser utilizado como “*trigger*” no estudo de outras estruturas, e que seja compatível com o equipamento de IRM, pelo que se registaram grandes dificuldades para a realização deste estudo dinâmico. O apoio técnico prestado pelos fabricantes do equipamento foi exíguo, exigindo o recurso a uma série de adaptações, quer por limitação de recursos, quer por limitações na programação das sequências existentes.

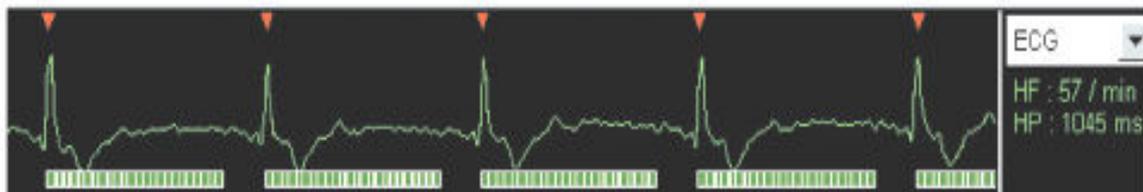
Na revisão da literatura encontram-se alguns estudos dinâmicos, com descrições muito sumárias relativamente às técnicas de sincronismo usadas. Alguns dos estudos dinâmicos encontrados foram realizados recorrendo à aquisição directa e dinâmica das imagens pelo próprio equipamento (usando técnicas adaptadas de sincronismo por sub-amostragem). A título de exemplo, usando o mesmo princípio da estroboscopia, Mathiak et al. (2000), descreveram uma aquisição dinâmica na qual os sujeitos tentavam sincronizar a fala com o ruído provocado pelo equipamento. Para além de requerer muita concentração por parte dos sujeitos, esta técnica obrigava a um elevado número de repetições dos sons, para além de não existir um controlo directo relativamente ao ruído provocado pelo equipamento. Outros estudos recorreram à aquisição das imagens através da associação das mesmas à informação proveniente da segmentação do sinal

áudio recolhido através do intercomunicador (Shadle et al., 1999; Demolin et al., 2002). Nesta técnica, os sujeitos iniciavam a repetição alguns segundos antes do arranque das sequências, durante a qual depois, simultaneamente era efectuada a gravação do sinal da fala. Esta sincronização é difícil de realizar, principalmente devido ao elevado ruído durante a aquisição, dificultando a análise da fala.

Para além de técnicas de sincronismo adequadas, nos estudos dinâmicos é aconselhável a utilização de sequências ultra-rápidas, mas que, no caso particular do tracto vocal, e após várias experiências, têm aplicações muito restritas; pois o número de artefactos é muito elevado, prejudicando a qualidade das imagens (impedindo mesmo a sua segmentação). No tracto vocal, os artefactos são abundantes pois resultam não só do movimento dos articuladores como também do próprio fluxo de ar.

Assim, para a realização deste estudo dinâmico utilizou-se uma sequência rápida FLASH 2D (usada normalmente em estudos cardíacos), e um trigger cardíaco com monitorização por electrocardiograma (ECG). Esta sequência foi adaptada para o estudo do tracto vocal, através do uso de uma antena de cabeça e da tentativa de sincronização do sinal da fala com o próprio batimento cardíaco dos sujeitos, que recebiam informação acústica do mesmo via auscultadores.

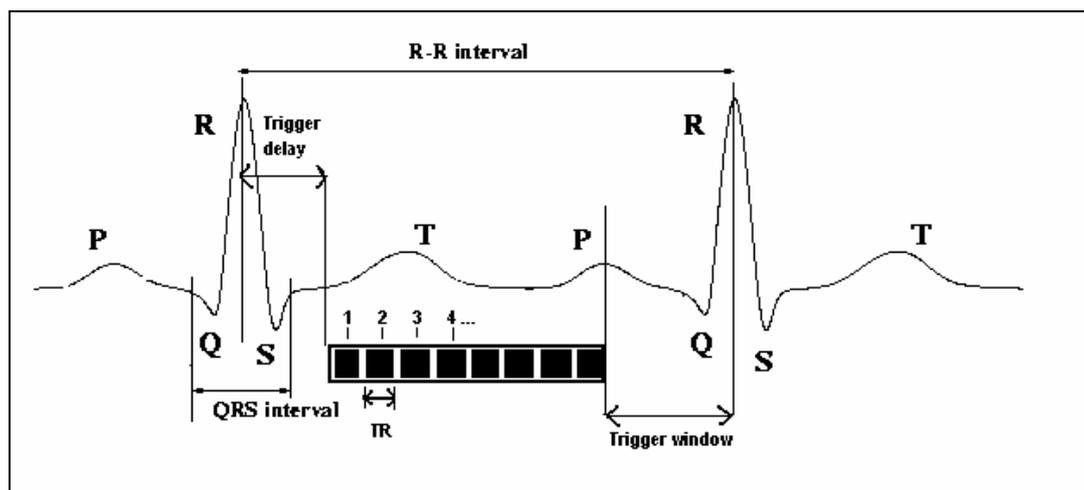
Neste tipo de sincronização, o pico da onda R do ECG determina o início da aquisição, sendo obtida uma imagem por cada intervalo R-R, numa determinada fase (deslizante) do ciclo cardíaco. Esta sequência é continuamente repetida, através da monitorização do intervalo R-R, sendo as sequências de imagens reconstruídas quando a aquisição está completa (Figura 32).



**Figura 32.** Gráfico de monitorização por ECG e respectivo registo do conjunto de imagens a adquirir mediante cada intervalo R-R (caixas a verde por baixo do traçado).

**Fonte:** [http://www.blackwellpublishing.com/content/BPL/Images/Content\\_store/Sample\\_chapter/1405124474/1405124474\\_4\\_002.pdf#search=%22MRI%20cardiovascular%20methods%20techniques%22](http://www.blackwellpublishing.com/content/BPL/Images/Content_store/Sample_chapter/1405124474/1405124474_4_002.pdf#search=%22MRI%20cardiovascular%20methods%20techniques%22).

A rapidez das sequências em IRM, está principalmente relacionada com o TR, no entanto, usando este tipo de sincronismo, o TR está condicionado ao intervalo R-R (duração do ciclo cardíaco) de cada sujeito, sendo cada onda R determinada pela frequência cardíaca. A frequência cardíaca vai condicionar também o tipo de ponderação da imagem e o número de imagens a adquirir durante o TR (Figura 33).



**Figura 33.** Representação esquemática da técnica de sincronização por ECG em IRM para sub-amostragem de seqüências.

O trigger window (TW) corresponde ao período de espera antes de cada onda R (atraso temporal) e, deve ser escolhido de forma a compensar qualquer aumento da frequência cardíaca, estando por isso dependente da duração do ciclo cardíaco. O período de espera após cada onda R denomina-se de trigger delay (TD), e corresponde ao tempo necessário para a excitação por RF do corte, que neste caso específico foi de 100ms. Quanto maior for o intervalo R-R (duração de cada ciclo cardíaco) maior é o número de fases cardíacas obtidas (maior número de imagens adquiridas).

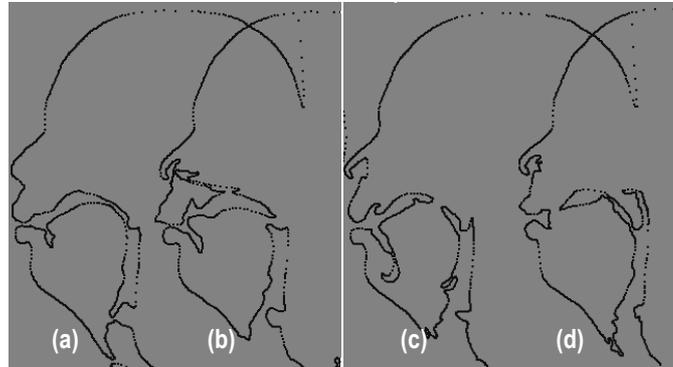
Neste estudo dinâmico, pediu-se os sujeitos, que após audição, durante alguns minutos, do próprio batimento cardíaco (através dos auscultadores do equipamento) iniciassem a repetição sucessiva de uma sílaba, de forma síncrona com o sinal sonoro produzido, sem exageros de articulação ou prolongamento dos sons, e respirando de forma calma e normal.

A seqüência era iniciada alguns segundos após verificação do sincronismo, obtendo-se um conjunto de cortes médiosagittais com 6mm de espessura, em número e com uma duração total variável de acordo a duração do ciclo cardíaco (Anexo G).

Foram usados os seguintes parâmetros de aquisição: TR = 60ms ou 86ms (para a sílaba /pa/ em três dos sujeitos), TE = 4,4ms, FA = 30°, e um FOV = 300mm; o tempo de aquisição total entre sujeitos variou entre 12s e 22s.

A figura 34 apresenta uma seqüência de alguns dos contornos extraídos das imagens, relativas à repetição da sílaba /tu/ para um dos sujeitos. Como se pode observar, a forma do tracto vocal tem variações pequenas, no entanto, são perceptíveis as modificações temporais da posição dos articuladores. A consoante /t/ inicia-se pelo fecho dos lábios (34a) seguido pela aproximação do ápex da língua à região alveolar (34b). Posteriormente, a língua recua e o seu dorso eleva-se,

dando-se a abertura da boca (34c) para a produção da vogal /u/, repetindo-se o processo para a repetição seguinte (34d).



**Figura 34.** Extracção de contornos das imagens médiosagittais obtidas no estudo dinâmico, pela repetição da sílaba /u/.

Quando comparadas as cine-imagens das três sílabas do estudo dinâmico entre sujeitos, comprova-se a variabilidade inter-oradores descrita na literatura, não só porque os aspectos anatómicos são diferentes, mas também, porque cada sujeito adopta estratégias diferentes no controlo dos movimentos dos articuladores (Anexo H). Os estudos dinâmicos permitem o estudo dos movimentos articulatorios envolvidos na produção da fala e, também uma maior compreensão dos fenómenos de coarticulação.

### **2.3 INTERPRETAÇÃO DAS IMAGENS OBTIDAS**

A combinação dos diferentes cortes, nos planos sagital e coronal, revelou-se uma tarefa de difícil concretização, exigindo a exploração e análise das imagens obtidas quanto à sua referenciação e localização espacial. Pretendia-se assim, obter o alinhamento das imagens num sistema de coordenadas comum – aqui denominado registo – usando um processo, o mais automático possível, e cuja localização espacial correspondesse de facto à sua posição de origem no sujeito.

Como foi descrito no capítulo 6 da parte I, a localização e codificação do sinal recebido pela antena é realizada pelos gradientes de campo magnético. No entanto, não está explícito inicialmente quais os parâmetros ou indicadores nas imagens que fornecem informação acerca da sua localização espacial relativa e absoluta, sendo mesmo inexistentes nas imagens fornecidas pelo sistema em formato .bmp. Após pesquisa bibliográfica, e porque inicialmente, as imagens foram

gravadas em formato *bitmap* (bmp), não se verificando nenhum indicador nesse sentido, optou-se pela gravação das imagens no seu formato de origem – Digital Imaging and Communications in Medicine (DICOM).

As imagens médicas das diferentes modalidades em Radiologia são apresentadas no formato DICOM, que é o *standard* global de comunicação mais usado actualmente. Este *standard* de transmissão de imagens digitais possibilita a recuperação de imagens e de toda a informação associada (metadados), assim como a partilha de informação entre equipamentos.

O DICOM contém todos os metadados referentes à gestão da imagem, impressão, transferência e arquivo (como os parâmetros de aquisição, características de imagem, e outros objectos relativos à orientação do paciente) e, também informação da *stack* (como a resolução, a matriz, e a distância de separação entre dois cortes, em milímetros – *depth*). Os atributos (e “*tags*”) são usados para descrever objectos de informação, e em conjunto formam elementos de dados.

Graças ao software utilizado na segmentação das imagens foi possível aceder à informação DICOM, mas pela sua especificidade, pois os atributos não apresentam unidades de grandezas normalizadas, foi necessário recorrer ao apoio técnico dos fabricantes do equipamento, para a sua interpretação. No entanto, além de tardia, a exígua ajuda revelou-se desajustada. A interpretação dos dados foi realizada, mas não foi possível ultrapassar, a despeito dos diversos esforços desenvolvidos, a existência de um aparente desajuste de pequena monta no alinhamento (registo) entre os cortes coronais e os cortes sagitais.

O DICOM Conformance Statement é um documento que descreve a funcionalidade e sintaxe, para cada fabricante, mostrando o cumprimento deste *standard*, no qual todos os atributos usados nesse equipamento específico são descritos, assim como alguns elementos de dados. Com base neste documento, foi possível, após extensa pesquisa bibliográfica e interpretação geométrica, realizar a interpretação dos atributos relacionados com a localização das imagens. Assim, o alinhamento dos cortes foi obtido com base na interpretação dos seguintes atributos (e “*tags*”):

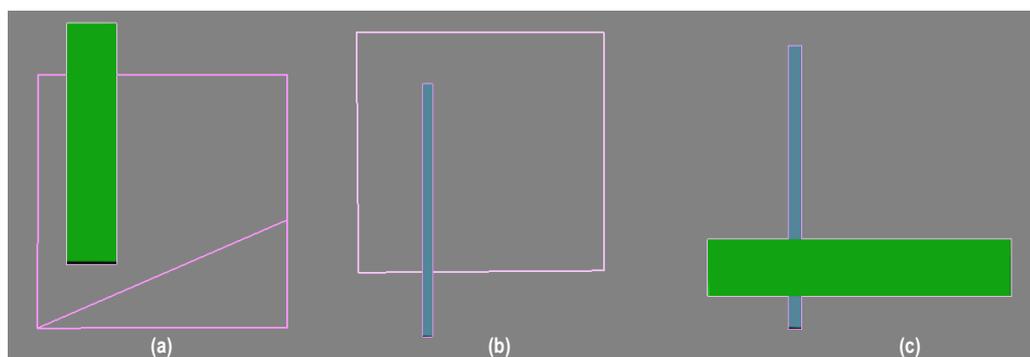
- *Image orientation* (0020,0037) – corresponde à direcção dos cosenos da primeira linha e da primeira coluna relativamente ao paciente;
- *Image position* (0020,0032) – corresponde às coordenadas x, y e z do canto superior esquerdo (centro do primeiro voxel transmitido) da imagem, em milímetros;
- *Slice thickness* (0018,0050) – espessura de corte em milímetros;
- *Pixel spacing* (0028,0030) – distância física no paciente entre o centro de cada *pixel*, em milímetros.

Os valores dos atributos das *stacks* obtidas, no presente estudo estático, encontram-se descritos no quadro 8.

**Quadro 8.** Atributos usados para a representação tridimensional no espaço de cada *stack*.

<i>Stack</i>	Indivíduo 1 (sexo masculino)		Indivíduo 2 (sexo feminino)	
	Sagital	Coronal	Sagital	Coronal
<i>Image orientation</i>	0\1\0\0\0\0\1	1\0\0\0\0\0\1	0\1\0\0\0\0\1	1\0\0\0\0\0\1
<i>Image position (mm)</i>	5; -145,8; 41,02	-75; -110; 41	5; -142,2; 7,05	-77,79; -91,99; 25,89
<i>Slice thickness</i>	5mm	6mm	5mm	6mm
<i>Depth</i>	5mm	16mm	5mm	16mm
<i>Pixel spacing</i>	1,17/1,17mm			
<i>Resolução</i>	0,853 px/mm			

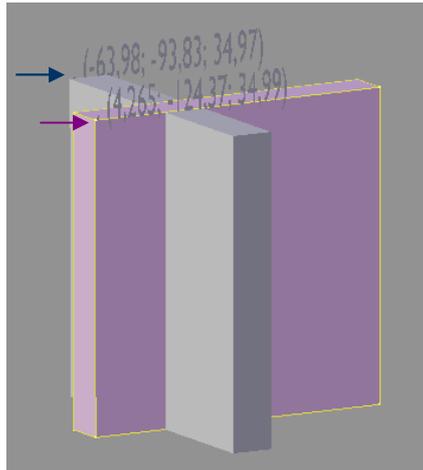
O sistema de coordenadas e os valores do atributo *image orientation* especificam a orientação das linhas e colunas da matriz de imagem. Antes de conseguir a interpretação correcta relativamente ao atributo do *image position*, realizaram-se várias tentativas de alinhamento das *stacks*, no ambiente 3D do Blender, usando directamente esses valores, mas o desajuste era injustificável como se pode verificar na figura 35.



**Figura 35.** Representação esquemática com caixas da posição das *stacks* com base apenas nos valores do atributo *image position*, na vista lateral (a), vista de frente (b) e vista de cima (c). A caixa verde pretende representar a *stack* coronal, e a caixa azul representa a *stack* sagital.

De seguida, na tentativa de encontrar justificação para este desajuste, realizaram-se ensaios recorrendo ao uso de *fantomas*, para excluir a possibilidade de que esta situação poderia resultar de uma eventual mudança de posição dos sujeitos. No entanto, os *fantomas* de calibração do equipamento eram simplesmente invólucros esféricos repletos de água, e que não contribuíram com qualquer informação útil ou esclarecedora.

Posteriormente, e após muita persistência na procura de explicação para os factos, a correcta localização espacial do *pixel* no sistema de coordenadas, em ambiente Blender, foi obtida através do produto dos valores do atributo *image position* pelo valor da resolução de imagem, permitindo assim o correcto alinhamento e combinação das *stacks* quando observadas no sistema de representação 3D (Figura 36).



**Figura 36.** Representação esquemática da localização espacial de cada plano.

Uma nota se regista que é a necessidade de existência de um tutorial para formação dos técnicos envolvidos apoiado num léxico específico relacionado com o DICOM.

### CAPÍTULO 3. FASE DE PESQUISA E EXPLORAÇÃO DAS TÉCNICAS DE PROCESSAMENTO DE IMAGEM

A análise da geometria do tracto vocal requer o recurso a técnicas de processamento de imagem ajustadas à finalidade de cada estudo, de forma a permitir a delimitação das diferentes cavidades deste tubo de passagem do ar dos restantes tecidos que o circundam. Para tal, é necessário realizar duas tarefas fundamentais: (1) a localização e segmentação da região de interesse (ROI), e (2) a reconstrução tridimensional desse objecto; a primeira etapa é determinante para o sucesso da segunda.

Nalguns dos estudos encontrados na revisão da literatura, a concretização destas tarefas era obtida através da extracção manual de contornos (usando uma transparência ou curvas *bézier*) e, por reconstrução através da colocação de uma grelha, ou usando o Matlab (Demolin et al., 1998; Badin et al., 2000; Soquet et al., 2002; Serrurier et al., 2005; Badin et al., 2006). Estas tarefas, para além de muito árduas e morosas, são muito dependentes do investigador (aumentando a incerteza dos resultados) dada a edição manual das imagens. Outros porém, referiram a utilização da técnica de segmentação semi-automática *threshold*, que é menos morosa e de mais fácil execução (Baer et al., 1991; Demolin et al., 1996; Badin et al., 1998; Apostol et al., 1999; Engwall, 2000b; Kroger et al., 2000; Engwall, 2004). Apenas Behrends et al. (2001) e Narayanan et al. (2004) recorreram ao uso de técnicas automáticas, as *Kalman snakes* e a segmentação de crescimento por regiões, respectivamente.

Apesar do menor ou maior automatismo na realização destas tarefas, este processo continua a ser complexo, dados os vários problemas que ainda se verificam (mencionados em diversos estudos), relacionados, por um lado com a técnica de IRM, e por outro pelas características anatómicas do tracto vocal:

- Ainda se desconhece qual a acção da epiglote e dos seios piriformes na produção da fala, existindo, por isso, dúvidas quanto à importância da sua delimitação e inclusão na segmentação do tracto vocal;
- A não identificação dos dentes em IRM;
- O término do tracto vocal nos lábios ainda não está bem definido;
- Existem estruturas, que pela similaridade de intensidade de sinal em IRM com o tracto vocal (como os discos intervertebrais da coluna, a cavidade nasal, a porção óssea cortical do palato, e o ar circundante na face e pescoço) dificultam a correcta delimitação no caso de

utilização de técnicas semi-automáticas ou automáticas, devido à *fuga* desses contornos (Figura 37);

- O processo de localização do tracto vocal é também complexo e exigente (pela necessidade de conhecimentos anatómicos) consoante o plano em estudo, principalmente nos cortes axiais e coronais, pela dificuldade na identificação dos dentes e alvéolos e, na interpretação das próprias imagens.



Figura 37. Fuga de contornos usando a técnica semi-automática de *threshold*.

Num estudo comparativo entre os três métodos de segmentação mais usados em IRM para o tracto vocal (método manual, elástico e *threshold*), Soquet et al. (1998) constataram que apesar da distribuição dos erros ser muito similar para os três métodos, o método *threshold* é o que apresenta menor dispersão, e que independentemente do método usado, qualquer um dos seus parâmetros (contraste, tipo de curva, ou limiar) tem influência na área resultante.

Neste estudo experimental em particular, optou-se pelo método de *threshold*, por vários motivos:

- Pelo seu automatismo – a delimitação é quase automática, dado o limiar da ROI;
- Pela sua flexibilidade e adaptabilidade – o contorno é facilmente ajustável, e perfeitamente adaptável à morfologia do tracto vocal (ao contrário das curvas *bézier* em que a adaptação é mais discutível);
- Pela menor dependência do investigador, comparativamente a métodos manuais.

### 3.1 PROGRAMAS DE PROCESSAMENTO DE IMAGEM

Nos vários estudos revistos, a referência a programas de processamento de imagem é quase inexistente, isto porque a delimitação dos contornos é frequentemente manual, encontrando-se apenas dois programas <sup>3</sup> (um apenas compatível com Linux e outro que exigia licença). Por estes motivos, foi necessário realizar uma vasta pesquisa e exploração de programas de processamento de imagem, das suas funcionalidades, e principalmente da sua adequação a este estudo em particular.

O programa CTMRedit <sup>4</sup> foi o primeiro software encontrado nesta fase de pesquisa, pela sua divulgação em alguns sites ligados à investigação da fala, dado que foi desenvolvido para utilização, principalmente, na análise da produção da fala em IRM. Este programa permite a visualização, edição e reconstrução tridimensional de imagens de tomografia computadorizada e ressonância magnética e opera em paralelo com o Matlab. A segmentação das imagens pode ser realizada de forma manual ou automaticamente (*threshold*), e permite também a interpolação e reconstrução da ROI a três dimensões (Figura 38).

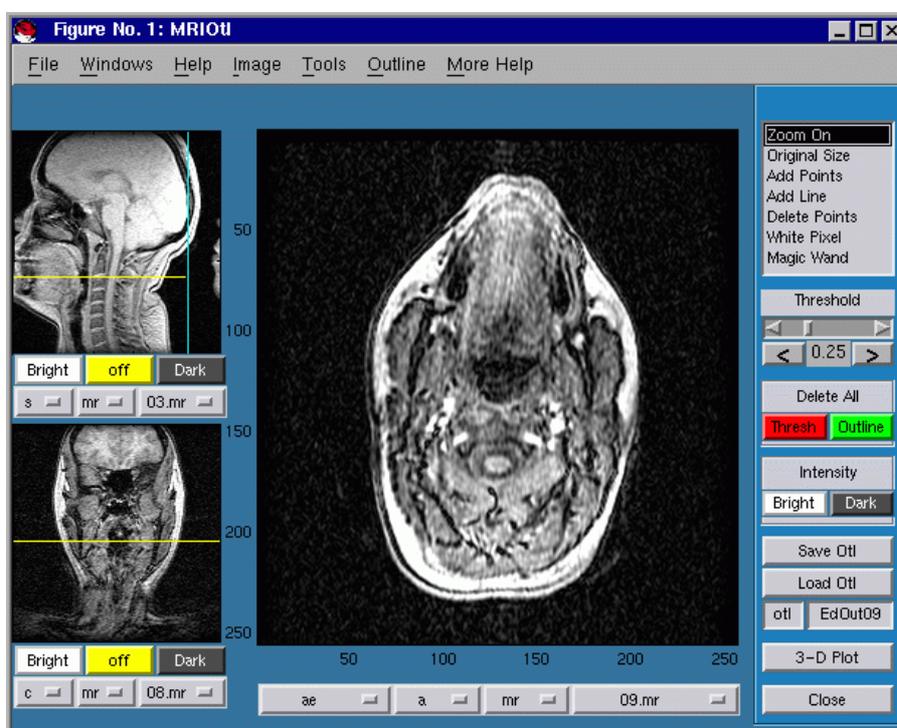


Figura 38. Janela de imagem e menu do software CTMRedit.

Fonte: <http://www.icsl.ucla.edu/~spapl/CTMRedit/tutorial.html>.

<sup>3</sup> VIDA (disponível em: <http://dpi.radiology.uiowa.edu/vida/vidahome.html>) e INRIA (disponível em: <http://cgal.inria.fr/Reconstruction/download.html>.)

<sup>4</sup> Disponível em: <http://www.icsl.ucla.edu/~spapl/CTMRedit/index.html>.

Na exploração deste software encontraram-se documentados três problemas (bugs de codificação, falhas de comunicação e falhas funcionais), bem como várias limitações, nomeadamente:

- O processo de segmentação é muito lento;
- A reconstrução tridimensional só é possível através da obtenção de imagens nos três planos;
- Os menus são pouco esclarecedores e apresentam várias limitações em termos de ferramentas de processamento de imagem (apenas é possível a realização de zoom, e manipulação do contraste da imagem).

Por todos estes motivos, optou-se por não o utilizar.

Apesar do vasto leque de programas de reconstrução tridimensional encontrados (Quadro 9), as suas características são variadas, bem como a sua complexidade de utilização e interacção. Para além disso, nem todos estão disponíveis de forma gratuita.

**Quadro 9.** Lista de programas de reconstrução tridimensional.

Programas de reconstrução tridimensional	Plataforma	Formatos de ficheiros compatível
Amira	Windows/HP-UX/Solaris	Vários formatos de ficheiros /DICOM
Analyse	Linux	Vários formatos de ficheiros
Etdips	Windows	Formato Analyze/DICOM/TIFF
Julius framework & JuliusLight	Linux/Windows/Macintosh	DICOM
Mimics (trial)	Windows	Vários formatos de ficheiros /DICOM
VGDtudio & VGStudio Max	Linux/Windows/Macintosh	Vários formatos de ficheiros /DICOM
Volview (trial)	Linux/Windows/Macintosh	Todos os formatos de ficheiros/DICOM
3D Doctor (trial)	Windows	Todos os formatos de ficheiros/DICOM
MRlcro	Linux/Windows/Solaris	Formato Analyze
ImageJ	Linux/Windows/Macintosh	Todos os formatos de ficheiros/DICOM

O programa MRlcro <sup>5</sup> permite a conversão de imagens médicas (em formato DICOM) provenientes da IRM, em formato Analyze, sem perda de dados, fazendo a sua correcta orientação, automaticamente. Possibilita a visualização, edição, segmentação de imagens, e a criação de outras imagens em diferentes planos. O seu ambiente de trabalho (Figura 39) é mais atractivo e funcional que o CTMRedit, para além do processo de segmentação ser mais rápido.

<sup>5</sup> Disponível em: <http://www.icsl.ucla.edu/~spapl/CTMRedit/tutorial.html>.

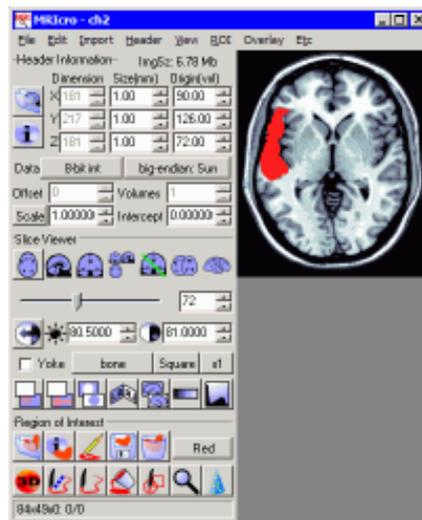


Figura 39. Ambiente de trabalho e menu do software MRIcro.

Fonte: <http://www.sph.sc.edu/comd/rorden/micro.html>.

Para além disso, possui mais operações geométricas, como rotação, translação, zoom, funções não lineares, e também permite o registo das imagens. A principal limitação deste programa encontra-se no mau resultado final da reconstrução tridimensional da ROI, que é inadequada para os objectivos deste estudo, o que motivou o abandono do mesmo.

Assim, o programa escolhido para este estudo experimental foi o ImageJ, pelos aspectos que a seguir se apresentam:

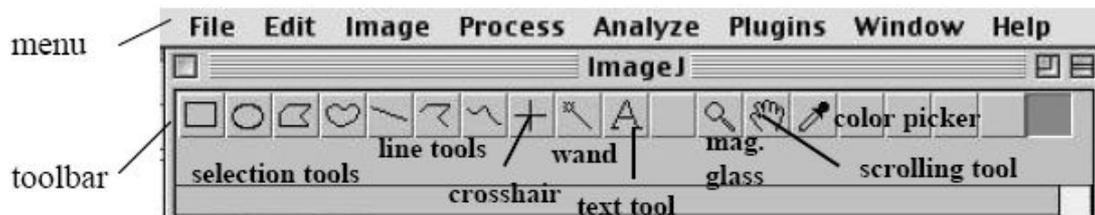
- Finalidade de utilização do programa;
- Objectivo do estudo para a reconstrução tridimensional;
- Tipo de imagens a analisar (formatos dos ficheiros, modalidade de imagem);
- Forma de arquivo após análise para ambos os estudos (estático e dinâmico);
- Requisitos técnicos necessários (sistema operativo, velocidade);
- Ferramentas de processamento disponíveis.

### 3.1.1 ImageJ

O ImageJ é um programa de processamento de imagem e análise de domínio público, compatível com qualquer sistema operativo, de fácil instalação, permitindo abrir todos os formatos de ficheiros como JPEG, BMP, PNG, TIFF, GIF e DICOM. Este programa permite a gravação das imagens também em qualquer formato, nomeadamente em ficheiros de vídeo (AVI, QuickTime),

adequando-se à forma de arquivo necessária para o estudo dinâmico efectuado. Para além das várias funcionalidades de origem, permite obter uma série de funcionalidades extras opcionais (“*plugins*”).

O menu é bastante acessível, e é constituído por uma barra de tarefas com ferramentas de selecção, de texto, de magnificação, de localização e dimensão de *pixels* (Figura 40).



**Figura 40.** Ambiente de trabalho e menu do programa ImageJ.

Fonte <http://rsbweb.nih.gov/ij/docs/pdfs/ImageJ.pdf#search=%22IMAGEJ%20pdf%22>.

Com este programa é possível efectuar a análise das imagens, como medir áreas, comprimentos, ângulos, ou a determinação do valor máximo e mínimo, a visualização do histograma, para além de permitir também a realização de operações geométricas (rotação, escala, redimensionar objectos). Outras funcionalidades importantes, especificamente para este estudo, são a possibilidade de visualizar os diferentes cortes obtidos para o mesmo plano num só conjunto de imagens – *stack*, e de visualizar os metadados associados às mesmas, através das teclas CTRL+ I; esta informação é útil para efectuar posteriormente o registo das *stacks*.

A segmentação das imagens do estudo estático foi realizada através do *plugin 3D editing tool*, que a seguir se descreve.

#### 3.1.1.1 *Plugin 3D editing tool*

O *plugin 3D editing tool* é uma ferramenta de modelação tridimensional, que permite a delimitação de contornos, com base em seis ferramentas de segmentação diferentes (curvas *bézier*, esferas (*ball tool*), *splines*, curvas *bézier* abertas (*pipe tool*), pincel (*brush tool*), e *threshold* (*wand tool*)) e que podem ser usadas para delimitar diferentes ROI na mesma *stack*. À excepção da ferramenta *wand*, nas restantes ferramentas, os pontos dos contornos têm que ser definidos pelo utilizador. Este *plugin* também permite associar outra ferramenta auxiliar de segmentação – o *SegmentingAssistant*, que se baseia na técnica *threshold*, realizando a detecção quase automática da ROI.

A principal vantagem do *3D editing tool* é a possibilidade de agrupar os contornos de uma determinada ROI da *stack*, num único objecto tridimensional (*skin*), podendo este ser posteriormente representado a três dimensões noutras aplicações gráficas como o Blender, Autocad, ou outros.

Assim, a segmentação das imagens do presente estudo, foi efectuada de acordo com os seguintes procedimentos:

- (1) Conversão dos cortes de cada plano em *stacks*, para cada som;
- (2) Identificação da região a segmentar e adição manual de objectos opacos para permitir o fecho da mesma, em cada um dos cortes;
- (3) Adição da imagem dos dentes apenas nas imagens das diferentes *stacks* sagitais, obtida pela extracção dos contornos dos mesmos do estudo anatómico feito inicialmente;
- (4) Delimitação dos contornos (fronteira ar-tecidos moles), de cada imagem bidimensional, para cada um dos sons, usando a técnica semi-automática de *threshold*, através da ferramenta SegmentingAssistant (Figura 41).

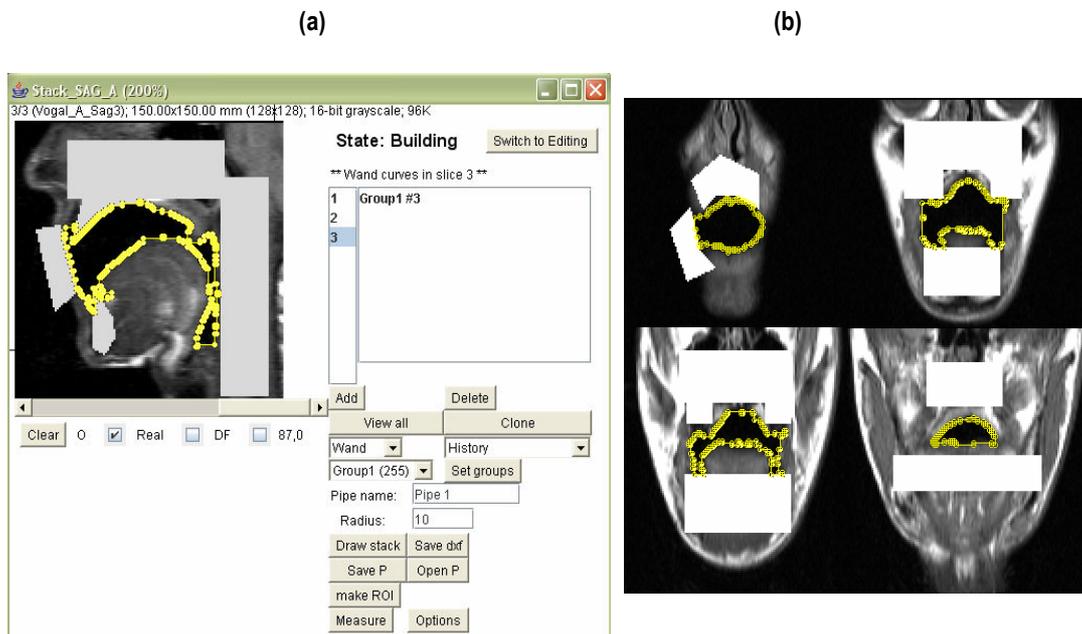


Figura 41. Outlines da *stack* sagital (a) e coronal (b) da vogal aberta /a/.

Com o auxílio da ferramenta SegmentingAssistant, após selecção de uma ROI, os contornos são automaticamente definidos (através do limiar), podendo ser ajustados de forma a permitir que o mesmo coincida com a estrutura a segmentar. Através da janela desta ferramenta é possível a

visualização e ajuste dos valores máximo e mínimo dessa ROI, bem como da localização do seu centro (posições horizontal e vertical) – centróide (Figura 42).

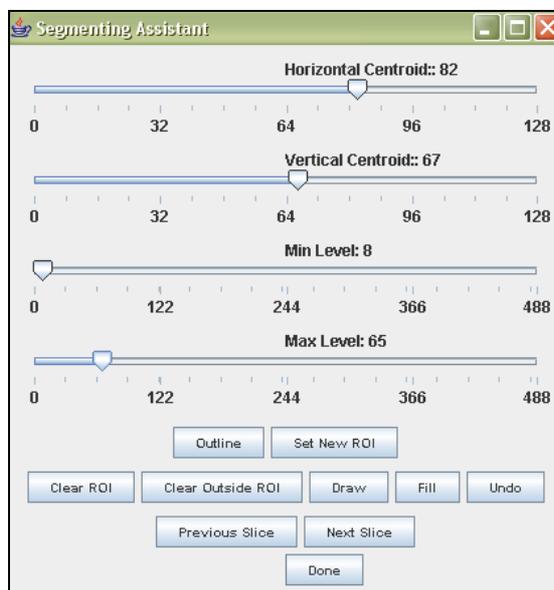


Figura 42. Janela da ferramenta SegmentingAssistant.

Ao conjunto de contornos bidimensionais obtidos pela delimitação em cada imagem, para cada som, dá-se o nome de *outlines*, e definem um objecto. Cada *outline* gerada pelo SegmentingAssistant está restrita a um número máximo de 1000 pontos.

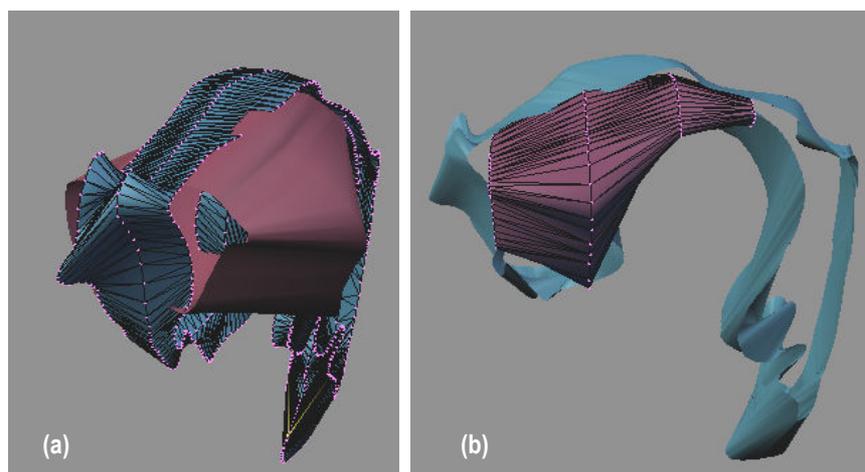
As *outlines* geradas em formato *.shapes* foram posteriormente usadas para a construção dos modelos tridimensionais, e depois importadas para o software Blender.

### 3.2 BLENDER

O Blender é um software para modelação tridimensional, animação, representação, criação interactiva e de efeitos visuais, e criação de jogos, de domínio público compatível com qualquer sistema operativo. Tem várias aplicações, nomeadamente em arquitectura, desenho gráfico, animações gráficas, e na modelação 3D de objectos (aplicação usada neste estudo em particular).

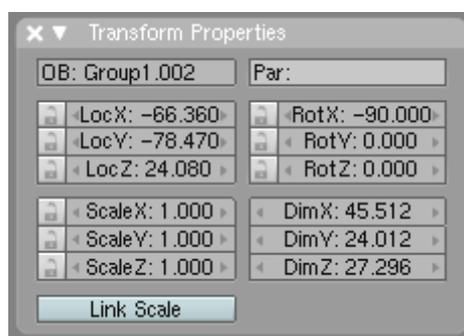
As *outlines* foram importadas para o ambiente Blender, utilizando uma escala de um (significa que o objecto será representado graficamente no seu tamanho real) e, um factor *depth* de 4,265 (*stack sagital*) e de 13,65 (*stack coronal*) que corresponde à distância de separação, em *pixels*, entre dois cortes consecutivos de uma *stack*. As imagens que se seguem (Figura 43) representam

duas perspectivas diferentes do modelo tridimensional obtido para a vogal /u/, após importação das *outlines* para o espaço gráfico deste software. A skin (objecto tridimensional) representada a azul (43a) foi gerada pela união das três *outlines* sagitais extraídas das imagens bidimensionais nesse plano; a skin representada em tons de vermelho (43b) resultou da união das quatro *outlines* extraídas da *stack* coronal.



**Figura 43.** Modelo tridimensional da vogal /u/ obtido pela união das *outlines* sagitais e coronais.

A combinação e representação tridimensional das duas *stacks* (sagital e coronal) foram realizadas através de dois procedimentos distintos: (1) a re-orientação das *outlines* no espaço gráfico do Blender, e (2) a translação das mesmas para a posição real do sujeito tendo em conta as coordenadas de origem das imagens obtidas; através do menu *Transform Properties* da figura 44 (conforme fluxograma de etapas, Anexo I).



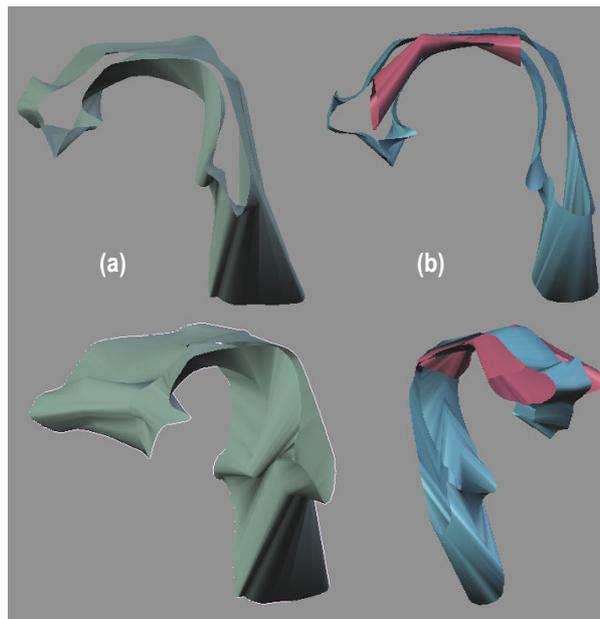
**Figura 44.** Menu *Transform Properties* do Blender usado na combinação e representação tridimensional das *stacks*.

## CAPÍTULO 4. BASE DE DADOS DE NATUREZA MORFOLÓGICA PARA CARACTERIZAÇÃO DO PORTUGUÊS EUROPEU

Esta base de dados de natureza morfológica pretende ilustrar a forma e posição dos articuladores envolvidos na produção dos diferentes sons do *corpus* da fala, com base nos modelos tridimensionais criados (Anexo H).

### 4.1 MODELOS TRIDIMENSIONAIS

Nos modelos tridimensionais a seguir apresentados (Figura 45) faz-se a comparação entre, o que neste trabalho denominamos de, modelos convencionais e modelos não-convencionais. De facto, a forma mais convencional para criação de modelos tridimensionais consiste na obtenção de um elevado número de cortes da estrutura, num só plano (normalmente, o axial), e sua posterior reconstrução. No estudo do tracto vocal, pelas suas características anatómicas (comprimento e dimensão lateral e transversal), o plano eleito é o sagital (como descrito anteriormente). Os modelos não-convencionais foram obtidos pela combinação de diferentes cortes, cuja informação é característica desse plano, constituindo por isso uma mais-valia para a caracterização dessa estrutura anatómica.



**Figura 45.** Modelo 3D convencional (a) e modelo 3D não convencional (b) em perspectivas diferentes, da vogal /u/ sustentada pelo mesmo sujeito (sexo feminino).

Estes modelos tridimensionais da figura 45, obtidos de forma convencional (no primeiro estudo estático atrás descrito), e de forma não convencional (estudo estático final) para a mesma vogal, demonstram claramente que: (1) no modelo 3D convencional, apesar do elevado número de cortes num só plano, a informação lateral não é claramente demonstrada, como o arredondamento dos lábios, e a dimensão lateral da cavidade oral; (2) a informação sagital obtida de forma não convencional é suficiente para a caracterização deste plano, demonstrando mais claramente a forma da língua e a sua relação com o palato; (3) a informação coronal permite complementar a informação sagital; (4) no caso do modelo convencional a informação lateral é pouco perceptível não representado a forma morfológica real do tracto vocal.

Para melhor identificação dos diferentes modelos 3D, a seguir apresentados, optou-se pela utilização de uma nomenclatura mais simplificada, ao invés dos símbolos do IPA para a classificação dos sons do PE, e que se apresenta no quadro 10.

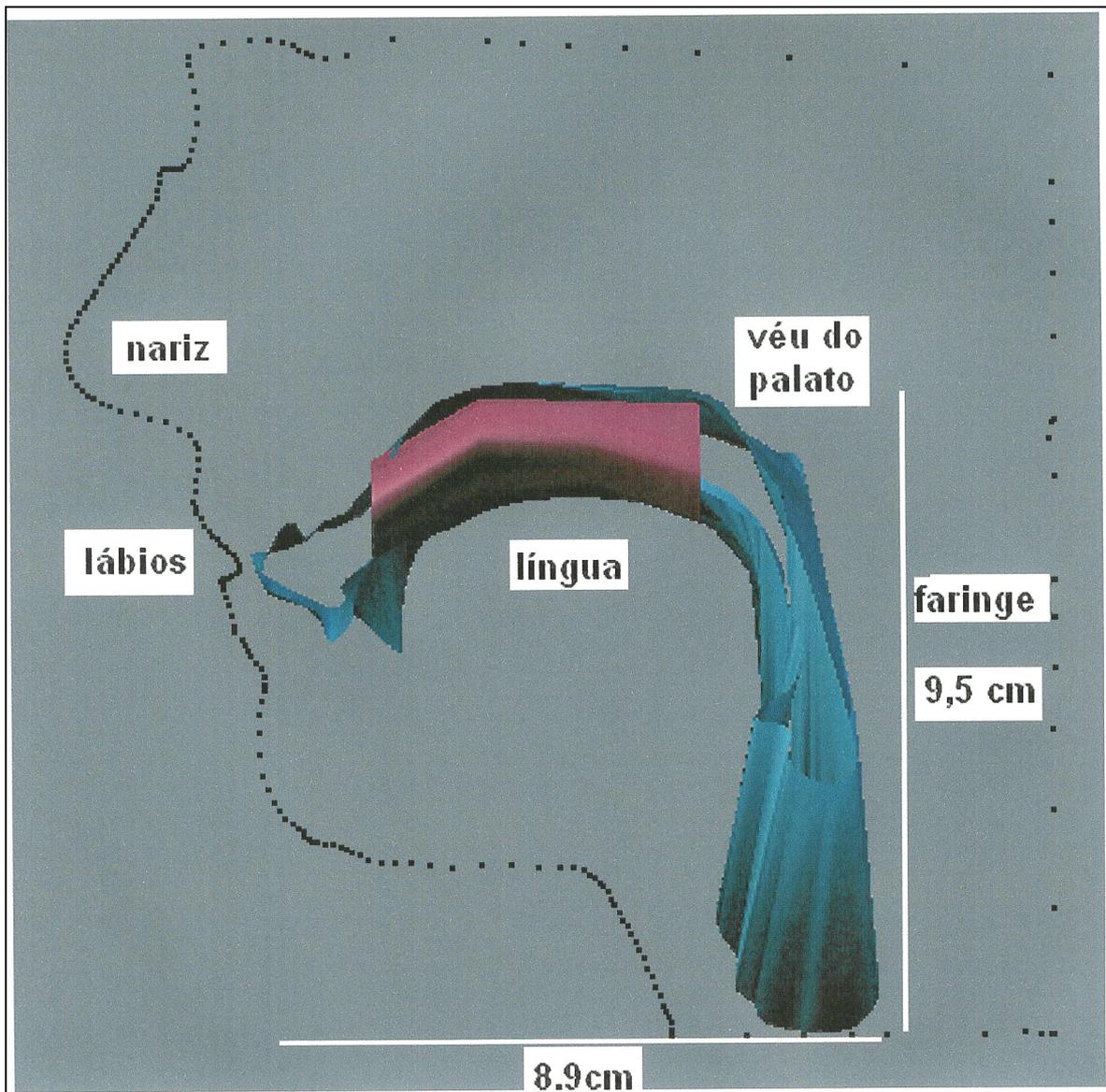
**Quadro 10.** Nomenclatura utilizada nos modelos 3D para a caracterização dos sons do PE.

Vogais orais		Vogais nasais		Consoantes	
Nomenclatura	IPA	Nomenclatura	IPA	Nomenclatura	IPA
Vogal I	i	Vogal A nasal	ã	Cons_F	f
Vogal (e)	e	Vogal E nasal	~e	Cons_V	v
Vogal E	ɛ	Vogal I nasal	~i	Cons_S	s
Vogal A	a	Vogal O nasal	õ	Cons_Z	z
Vogal (a)	ɶ	Vogal U nasal	~u	Cons_CH	ʃ
Vogal (t)	ɨ			Cons_J	ʒ
Vogal O	ɔ			Cons_L	l
Vogal (o)	o			Cons_LH	ʎ
Vogal U	u			Cons_M	m
				Cons_N	n
				Cons_NH	ɲ

Estes modelos tridimensionais (Figuras 47 a 51), não-convencionais, do *corpus* da fala, criados a partir dos cortes provenientes do estudo estático, para a caracterização do PE, pretendem demonstrar importantes características dos eventos articulatórios envolvidos na produção dos diferentes sons, nomeadamente:

- Forma da língua e a sua relação com a epiglote;
- Posição central da língua na cavidade oral e sua relação com o palato;
- Grau de abertura da boca e forma dos lábios;
- Comprimento de todo o tracto vocal;
- Posição do véu do palato (abertura ou fecho).

Para melhor compreensão quanto à forma de apresentação e dimensões dos modelos tridimensionais obtidos, apresenta-se na figura 46, uma imagem de um dos modelos tridimensionais, para a vogal (a) sustentada por um sujeito do sexo feminino, em tamanho real, e com indicação das suas principais dimensões e, da sua relação com outras estruturas (para melhor localização anatômica).



**Figura 46.** Imagem em tamanho real, do modelo tridimensional da vogal (a), de um sujeito do sexo feminino, com indicação das suas dimensões e correlação anatômica.

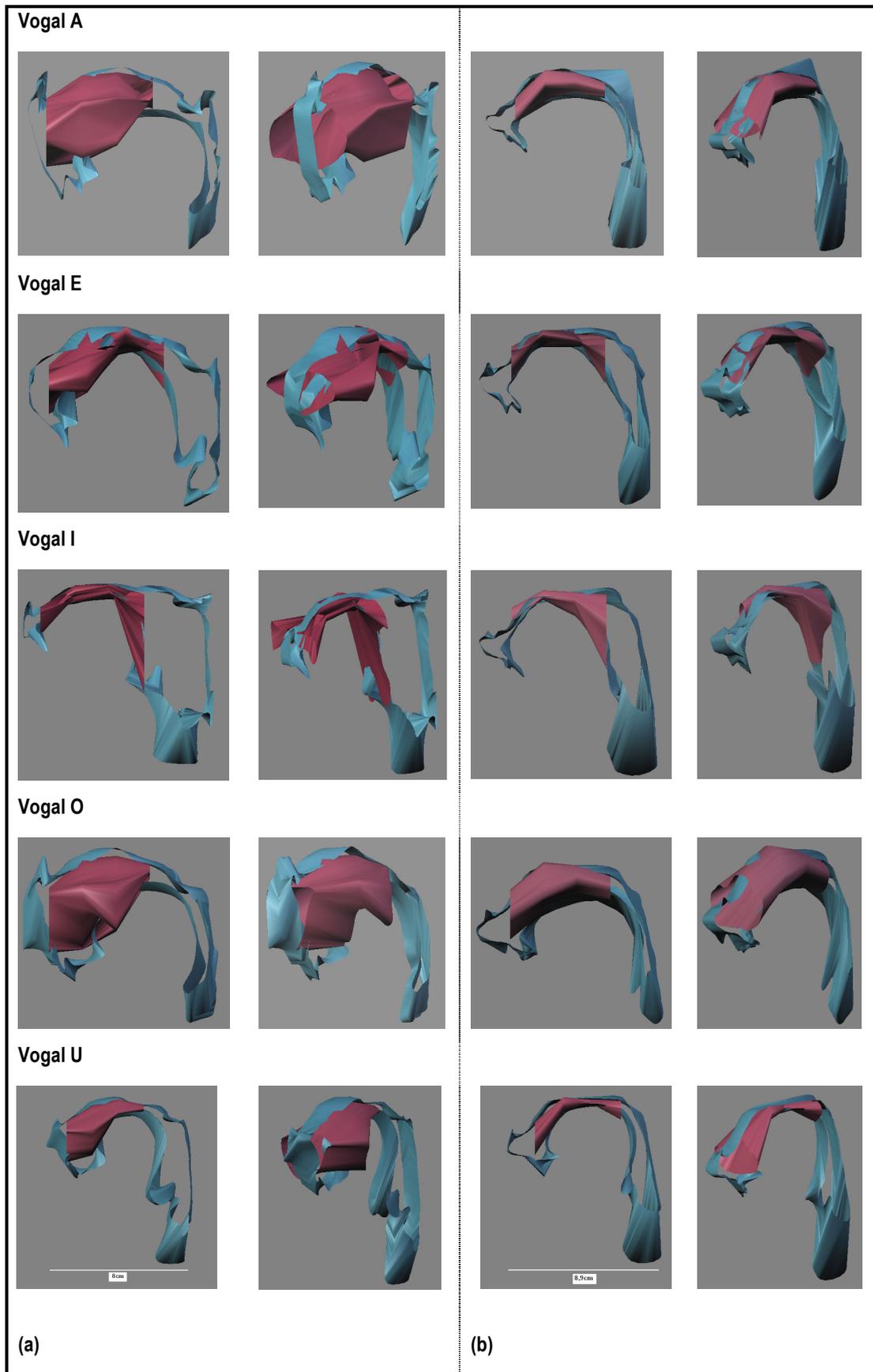
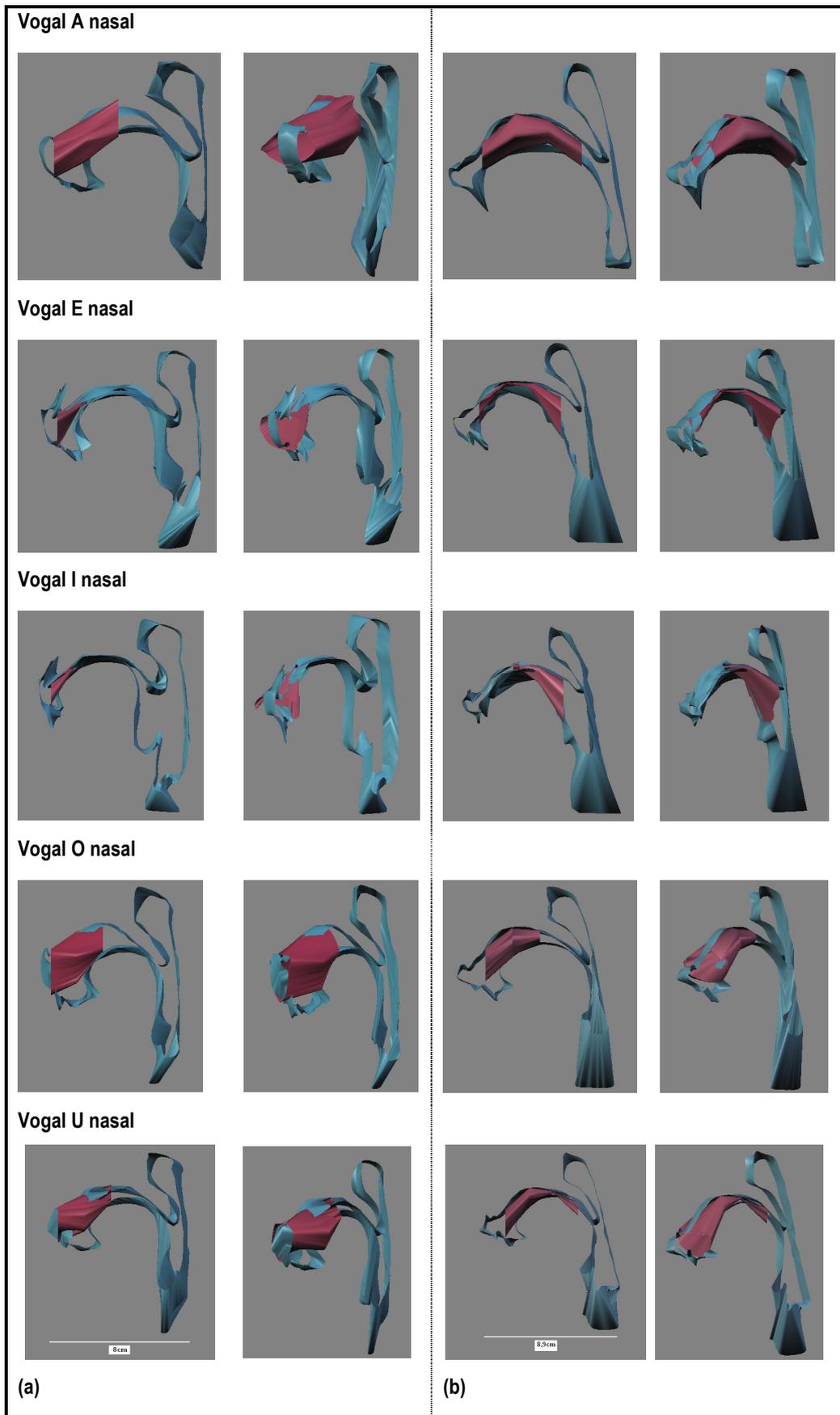
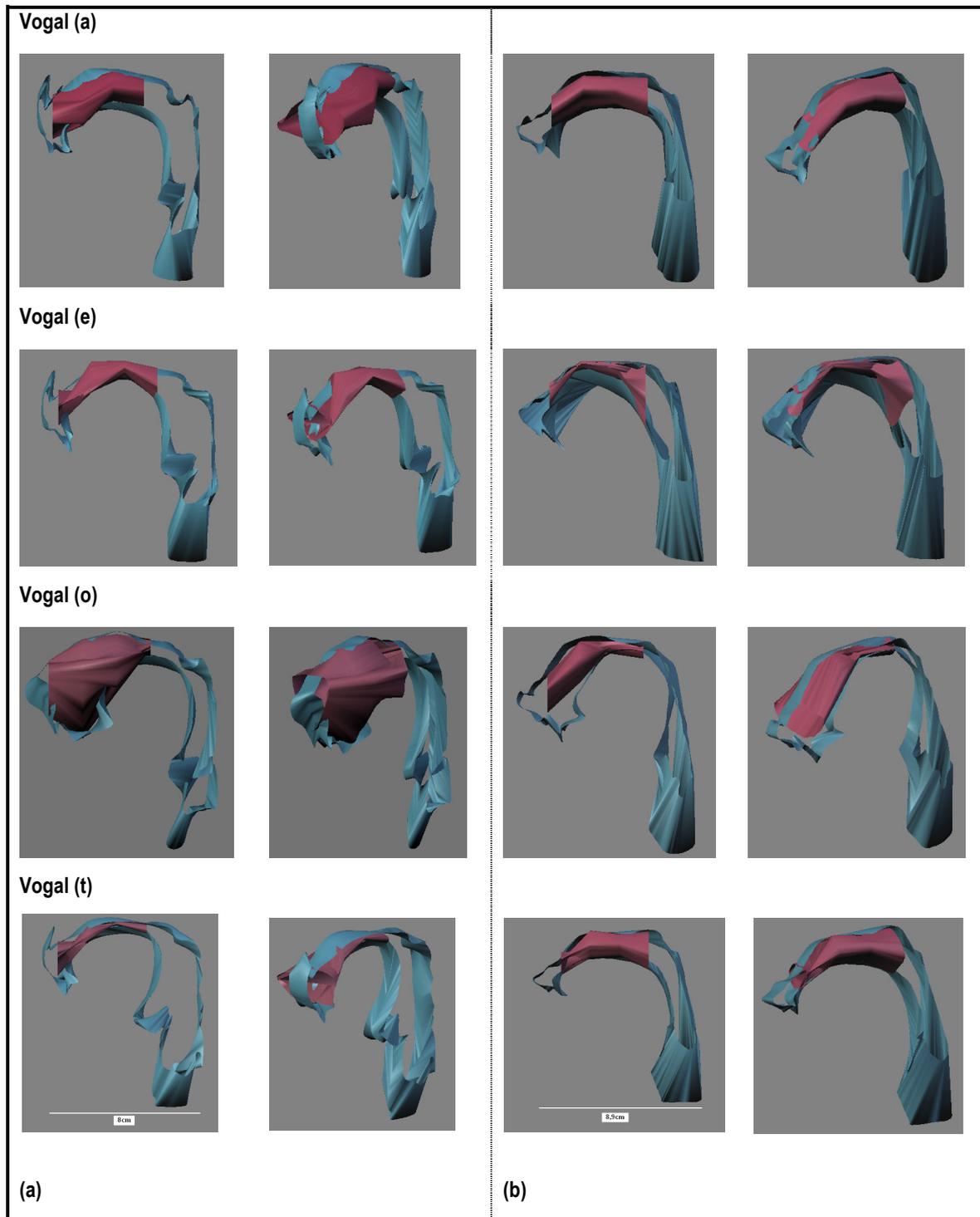


Figura 47. Modelos tridimensionais das cinco vogais orais do Português produzidas por um sujeito do sexo masculino (a) e do sexo feminino (b).



**Figura 48.** Modelos tridimensionais das cinco vogais nasais do Português produzidas por um sujeito do sexo masculino (a) e do sexo feminino (b).

Por comparação dos modelos tridimensionais da figura 46 e da figura 47, é bem demonstrada a principal diferença entre as vogais orais e nasais, observando-se a abertura do véu do palato (com aproximação em direcção à língua) e a formação da cavidade nasal, na figura 47.



**Figura 49.** Modelos tridimensionais das quatro vogais orais do Português produzidas por um sujeito do sexo masculino (a) e do sexo feminino (b).

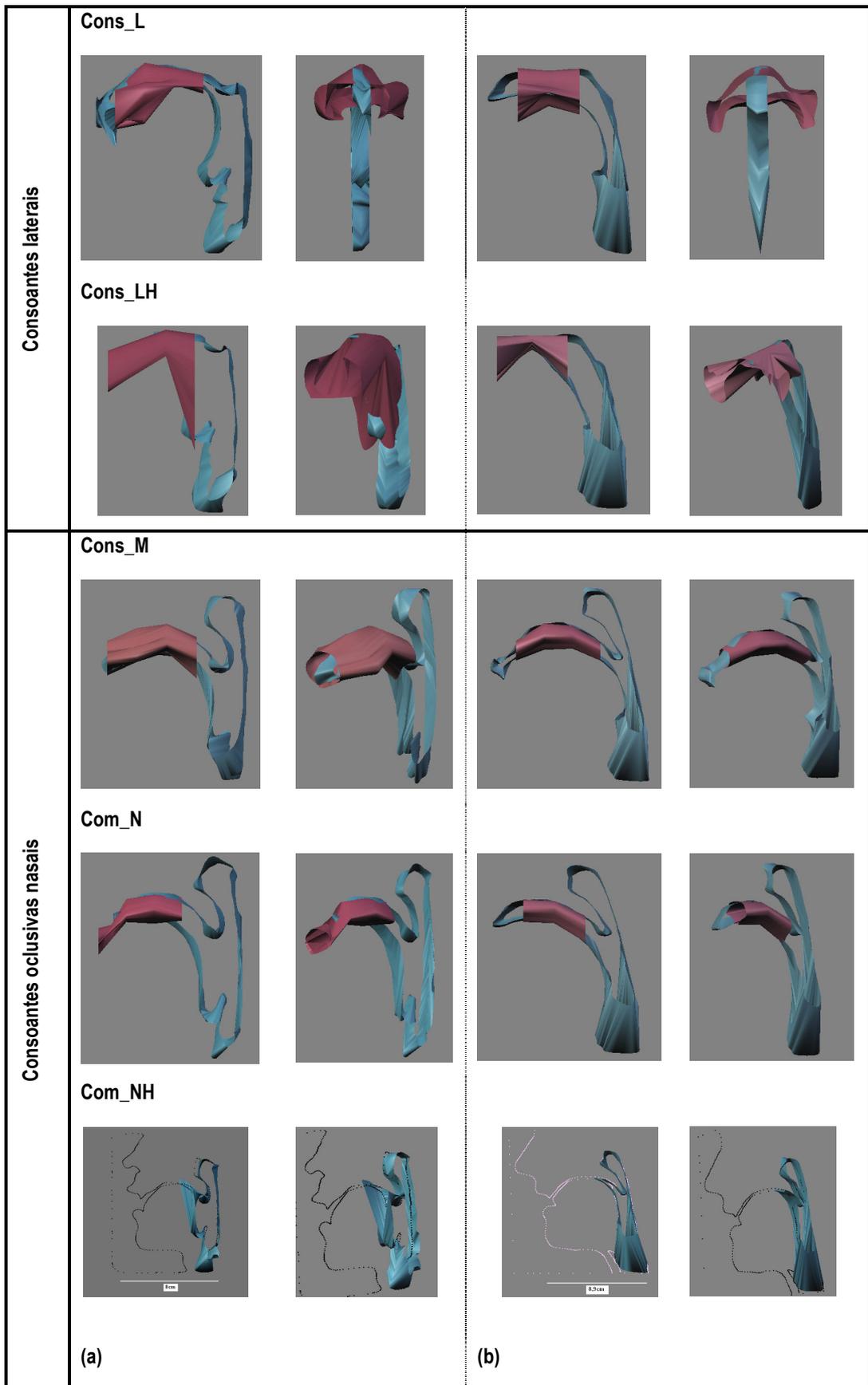
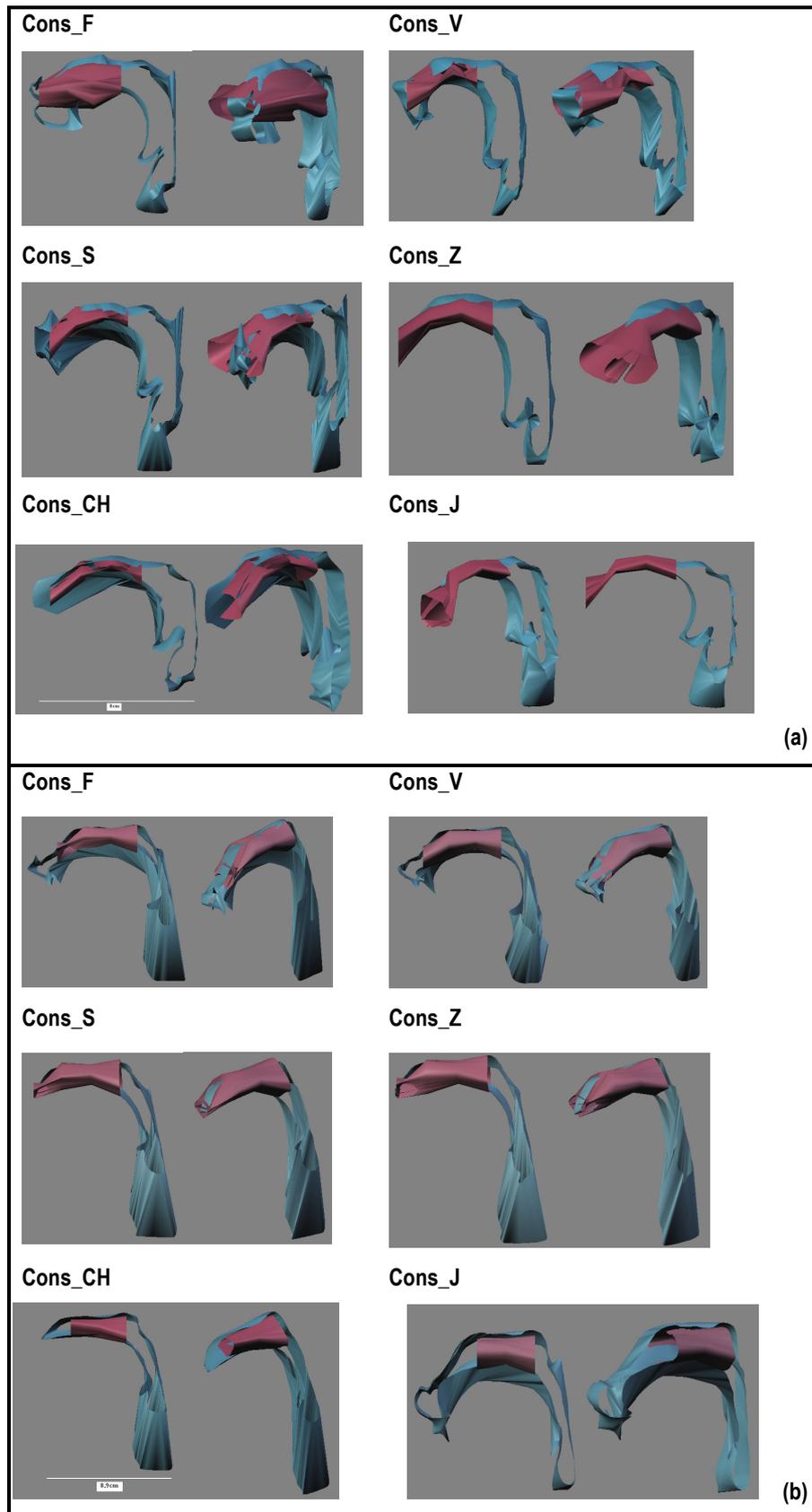


Figura 50. Modelos tridimensionais das consoantes laterais e oclusivas nasais do Português produzidas por um sujeito do sexo masculino (a) e do sexo feminino (b).



**Figura 51.** Modelos tridimensionais das consoantes fricativas do Português produzidas por um sujeito do sexo masculino (a) e do sexo feminino (b).

## 4.2 DISCUSSÃO DOS RESULTADOS

Os modelos tridimensionais apresentados nas figuras anteriores demonstram a posição e forma dos diferentes articuladores envolvidos na produção dos diferentes sons, em diferentes perspectivas. Como se verificaram algumas diferenças posicionais entre as duas *stacks* (sagital e coronal), para alguns sons, como resultado de alguns erros no registo das mesmas (como mencionado anteriormente), procedeu-se ao ajuste manual da posição de algumas *stacks*.

Na maioria dos modelos tridimensionais obtidos para caracterização dos diferentes sons do PE, é possível constatar a importância quer da informação sagital, quer da informação coronal para uma melhor compreensão da morfologia do tracto vocal durante produção da fala. As modificações posicionais dos diferentes articuladores são particularmente bem demonstradas no plano sagital, enquanto que a informação coronal possibilita uma melhor observação das dimensões laterais do tracto vocal, e em particular da cavidade oral.

Nas imagens apresentadas verificam-se alguns desajustes dimensionais, que nos parece poder estar relacionado, por um lado com a variabilidade na produção dos sons, e por outro com método de segmentação utilizado. Para reduzir o esforço dos sujeitos durante a sustentação dos sons dado o *corpus* em estudo, utilizaram-se tempos de aquisição mais curtos (ajustados de acordo com o tempo de sustentação médio dos sujeitos), e optou-se pela obtenção das *stacks* separadamente para cada som. Estas aquisições distintas justificam a variabilidade, que é uma característica importante e que influenciou de alguma forma os resultados obtidos.

Na produção das vogais, as principais diferenças encontradas entre os diversos sons decorrem da posição da língua e lábios, bem demonstrados nos modelos tridimensionais obtidos (Figura 47). Apesar das diferenças anatómicas entre os dois sujeitos, as posições articatórias são similares, mas são mais exuberantes, na opinião da autora, para o sujeito do sexo masculino.

Da apreciação destes modelos, também se constata que as dimensões das *stacks* são para uma grande maioria dos sons, coincidentes, principalmente nos sons cujas posições articatórias são mais “rígidas”, como nas consoantes (Figura 50 e 51), comparativamente a sons mais abertos, com verificado nalgumas vogais.

No entanto, durante a segmentação das imagens, devido à particularidade de algumas posições articulatórias, nem todos os cortes coronais foram usados na construção dos modelos, como se pode observar para as vogais nasais (Figura 48).

A informação coronal é particularmente relevante, no caso das consoantes laterais (Figura 48), pois a obstrução do fluxo de ar provocada pela língua num determinado ponto da cavidade oral (junto dos alvéolos na consoante /l/ e junto do palato na consoante /lh/) causa a passagem do ar pelos lados da língua. Apenas na consoante /nh/, a informação coronal se mostrou pouco proveitosa, uma vez que o contacto entre a língua e o palato é total e o som sai pelas cavidades nasais (Figura 50).

Quanto à utilidade destes modelos tridimensionais, estes representam a forma do tracto vocal para uma determinada posição articulatória, e podem por isso, vir a servir por um lado, para o cálculo das funções de área (estas funções são obtidas pela medição de áreas seccionais, e servem de *input* para o cálculo das formantes, validando a teoria acústica), e por outro, para a construção de modelos articulatórios tridimensionais (com base nos parâmetros articulatórios) para síntese da fala.

Para além disso, com base nas imagens bidimensionais obtidas em diferentes cortes, no estudo estático inicial deste trabalho, foi possível a construção de um molde tridimensional físico do tracto vocal (que a seguir se descreve).

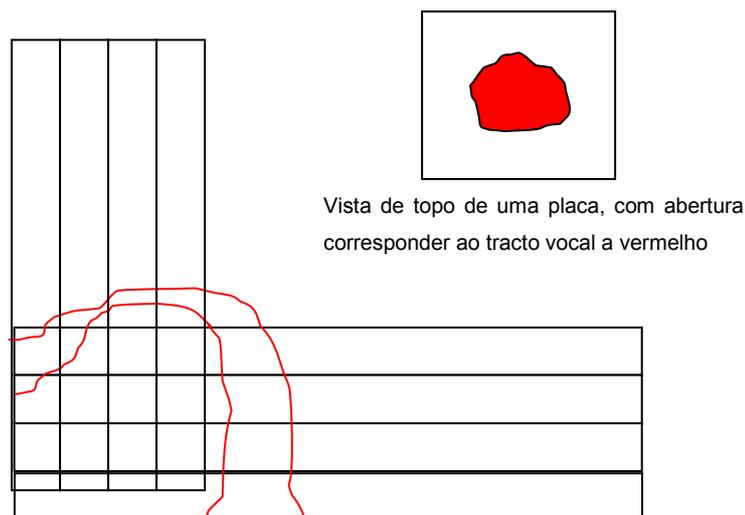
### **4.3 SIMULAÇÃO ACÚSTICA REAL DO TRACTO VOCAL BASEADA NA RECOLHA DE DADOS MORFOLÓGICOS DE IRM**

Num trabalho prático desenvolvido por outro investigador, no âmbito do Mestrado de Engenharia Biomédica (Anexo J), foi possível a construção de um molde tridimensional físico, para a simulação acústica real do tracto vocal, com base na informação morfológica obtida no estudo estático inicial nos três planos anatómicos (sagital, coronal e axial) para a vogal /i/.

Para além da autora ter fornecido as imagens obtidas para a vogal em causa, foram também fornecidas informações posicionais e dimensionais das mesmas, para que fosse possível a reprodução, o mais fiel possível, deste som.

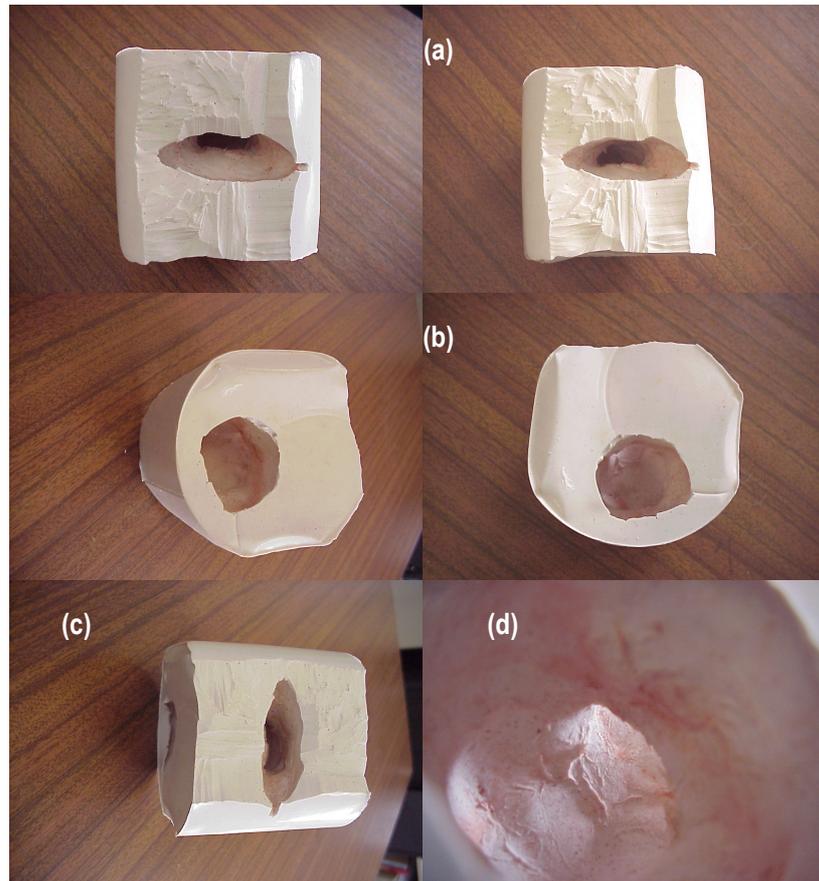
Assim, a construção do molde foi efectuada de acordo com os seguintes procedimentos:

- (1) Através de impressões das imagens coronais e axiais em tamanho real, e depois de decalcadas em placas de poliuterano com uma espessura igual à de cada corte, foi possível recortar a forma do tracto vocal;
- (2) As placas foram depois sobrepostas e alinhadas, com base em pontos comuns de cada imagem, de modo a formar uma abertura contínua (Figura 52);
- (3) A abertura contínua foi preenchida por resina de poliuretano, produzindo um bloco sólido com a forma do tracto vocal;
- (4) Este bloco serviu depois para molde positivo do tracto vocal artificial, em silicone.



**Figura 52.** Vista sagital do alinhamento das placas, para posterior construção do molde físico real do tracto vocal para a vogal /i/.

Nas imagens que se seguem (Figura 53) apresentam-se várias vistas do molde do tracto vocal construído. Apesar do aspecto grosseiro do molde, repare-se que nas duas vistas frontais (53a) é possível visualizar o ângulo e grau de abertura da boca, e nas restantes vistas a forma e abertura da porção inferior do tracto vocal.



**Figura 53.** Vistas frontais (a), vistas da base (b), vista frontal e da base (c) e vista interior (d) do molde do tracto vocal para a vogal /i/.

## CONCLUSÃO E PERSPECTIVAS PARA O FUTURO

O estudo da produção da fala, pela complexidade anatómica e funcional do tracto vocal, tem despertado um enorme interesse multidisciplinar. A utilização e evolução da IRM têm contribuído fortemente para o seu grande desenvolvimento. Apesar das vantagens inequívocas da IRM comparativamente a outras técnicas, persistem algumas limitações fortes face às necessidades já conhecidas, relacionadas em parte com a pouca acessibilidade destes equipamentos, e com as limitações ainda existentes de rapidez dos mesmos e com a inexistência de protocolos específicos para as recolhas.

Com base na IRM, e mediante disponibilização de um equipamento no Hospital S. João, realizaram-se vários ensaios e recolhas, que permitiram a caracterização morfológica e dinâmica do tracto vocal para a maioria dos sons do PE de um número de sujeitos adequado ao presente trabalho de dissertação. Obtiveram-se imagens em diferentes planos de um grande número e diversidade de posições articulatórias, primeiro através de um estudo estático (com sustentação dos sons), e posteriormente de sequências articulatórias através de um estudo dinâmico de três sílabas (mediante adaptação de um sequência ultra-rápida e usando como técnica de sincronização da produção vocal a monitorização e audição do ECG do sujeito).

As imagens do estudo estático, depois de analisadas e processadas, culminaram num conjunto de modelos tridimensionais, não-convencionais, através da combinação das informações sagital e coronal, permitindo a sua organização numa base de dados morfológica para caracterização dos eventos articulatórios (forma e posição dos articuladores) do *corpus* da fala.

Na definição do protocolo do estudo estático, e dado tratar-se de um estudo inicial, optou-se pela não realização de cortes oblíquos. A realização de cortes ortogonais e perpendiculares entre si permitiu obter a exactidão necessária para a extracção dos contornos, bem como facilitou o registo geométrico das *stacks*, que mais tarde, se revelou uma tarefa exigente e árdua.

No nosso entender, este estudo morfológico permite obter, não só, uma melhor compreensão do mecanismo da produção da fala, como também permite, em estudos comparativos, o estudo da variabilidade da fala inter-sujeitos.

Este conhecimento é reputado de muito útil não só para estudos de fonética articulatória, como também para o desenvolvimento de modelos articulatórios para síntese da fala, através do cálculo de funções de área e de parâmetros de posição dinâmica dos articuladores.

Apesar das várias dificuldades e limitações encontradas para realização do estudo dinâmico, foi possível a obtenção de sequências de imagens representativas de três sílabas, e que demonstram as variações temporais observáveis dos movimentos dos articuladores durante a produção da fala, bem como possibilitam um melhor entendimento dos mecanismos de coarticulação.

Durante a concretização deste trabalho verificaram-se algumas dificuldades, que influenciaram de forma não desprezável os resultados obtidos, nomeadamente:

- (1) Dificuldades na delimitação dos contornos, em parte devido à não identificação dos dentes em IRM, informação esta que é particularmente importante nos cortes coronais;
- (2) Limitações no número de imagens obtidas, condicionado quer pelo tempo de aquisição, quer pelo compromisso entre tempo e resolução do equipamento, e também quer no sentido de minimizar o esforço e variabilidade da articulação dos sujeitos durante a sustentação dos sons;
- (3) O tempo de aquisição, nos estudos dinâmicos, foi condicionado por limitações protocolares e de funcionamento do equipamento, bem como pela técnica de sincronização usada;
- (4) Dificuldades no registo geométrico das *stacks*, e cuja interpretação segundo a norma DICOM ainda não foi totalmente concretizada, por falta de apoio técnico, e informação existente incompleta;
- (5) A impossibilidade de registo acústico da fala, pelo elevado ruído provocado pelo equipamento durante as aquisições.

Por aperfeiçoar está ainda o moroso e difícil processo de segmentação das imagens, apesar do método de segmentação *threshold* ser quase automático. Este aspecto é importante devido significativamente, em parte, ao elevado número de imagens obtidas nestes estudos, e também aos problemas que ainda se verificam relacionados, por um lado com a própria técnica de IRM, e por outro com características anatómicas do tracto vocal.

O registo geométrico das imagens de diferentes tipos de cortes foi um factor crítico e com implicações na representação 3D dos modelos apresentados. Persistem ainda pequenas incertezas a reduzir mediante uma comunicação mais aprofundada com o sector técnico de apoio à utilização ou mesmo concepção do equipamento e do seu software. Verificaram-se alguns erros facilmente detectáveis, obrigando ao ajuste manual de alguns sons, devidos, provavelmente, às dificuldades relacionadas com a norma DICOM. Um melhor entendimento poderia ser conseguido com o desenvolvimento de *fantomas* apropriados, de forma a permitir obter informações posicionais ou dimensionais rigorosas e assim o “staff” ligado ao equipamento realizar uma calibração mais credível especificamente para os tipos de recolhas em causa neste trabalho. Deve esclarecer-se que as exigências das actuais recolhas não se verificam nas actividades para fins médicos, pelo que os fabricantes não devem ser muito solicitados por necessidades de calibração inter-planos de corte.

Os modelos tridimensionais criados ainda não estão totalmente fechados lateralmente por inexistência dessa possibilidade nas aplicações de processamento conhecidas, pelo que, em investigações futuras será necessário completar estes modelos híbridos, de forma a permitir a construção de um modelo anatómico 3D digital do tracto vocal que permita, entre muitas outras coisas extrair informação completa da superfície delimitadora e da sua dinâmica.

Quanto à IRM dinâmica, apesar de estar condicionada aos equipamentos existentes, o seu futuro passará pela extensão de outras sequências mais rápidas, como pelo desenvolvimento de um dispositivo para sincronização do sinal da fala, específico para estudo do tracto vocal, e viável (dada a presença do elevado campo magnético), de forma a possibilitar uma melhor adequação e fiabilidade destes estudos.

O aperfeiçoamento do processo de segmentação terá que passar necessariamente pelo desenvolvimento de técnicas ou dispositivos (como sejam próteses, pequenos marcadores, etc) que auxiliem na identificação dos dentes por IRM, sem necessidade de recorrer a outras técnicas, como a tomografia computadorizada, pelo seu carácter ionizante, e que não interfiram com a produção da fala.

Para além das limitações da IRM acima referidas (não identificação dos dentes e ruído provocado), acresce também a influência da postura corporal (decúbito dorsal) exigida na produção dos sons, pelo que, no futuro, estas poderão ser parcialmente compensadas pela

combinação com outras técnicas, nomeadamente a articulografia electromagnética, a electropalatografia ou a ultrassonografia.

Na fase de projecto deste estudo, ponderou-se a construção de uma maquete realista para um dos sons estudados, com base na informação morfológica obtida. Embora o tempo disponível não permitisse que a autora o realizasse plenamente, o objectivo não ficou frustrado devido à colaboração de outro investigador, também no âmbito do Mestrado de Engenharia Biomédica, que realizou uma primeira construção de uma maquete física do tracto vocal para experimentação acústica da vogal /i/.

Para o futuro perspectiva-se, por um lado, a implementação destes estudos na prática clínica para o diagnóstico e orientação clínica na avaliação de doentes com *deficits* articulatorios, e por outro, a possibilidade de utilização destes dados para o ensino de línguas, neste caso do PE.

Esta base de dados morfológica-dinâmica poderá também, num futuro próximo, servir como uma ferramenta auxiliar quer para a formação dos Terapeutas da Fala, como para a utilização da mesma por estes profissionais na avaliação de perturbações da fala.

Como já havia sido mencionado anteriormente, com a finalização destes modelos tridimensionais, será possível também o cálculo de funções de área e dos parâmetros articulatorios para síntese da fala, e apesar de estar fora do âmbito desta dissertação, a autora pretende manter-se ligada à continuidade da respectiva investigação.

Como resultado deste trabalho, a autora submeteu um artigo com alguns dos resultados obtidos, e que será apresentado no simpósio COMPIImage a realizar de 20 a 21 de Outubro de 2006, em Coimbra.

## **REFERÊNCIAS**

ABBOTT, M.B., DARDZINSKI, B.J., DONNELLY, L.F. Using volume segmentation of cine MR data to evaluate dynamic motion of the airway in pediatric patients. *AJR Am.J. Roentgenol.* 2003, 181 (3), pp. 857-9.

APOSTOL, L., PERRIER, P., RAYBAUDI, M., SEGEBARTH, C. 3D Geometry of the Vocal Tract and Interspeaker Variability. *Proceed. 14th International Congress of Phonetic Sciences (ICPhS), San Francisco, USA.* 1999, pp. 443-446.

AVILA-GARCÍA, M.S., CARTER, J.N., DAMPER, R.I. Extracting Tongue Shape Dynamics from Magnetic Resonance Image Sequences, *Transactions on Engineering, Computing and Technology V2.* 2004, p. 288-291.

BADIN, P., BOREL, P., BAILLY, G., REVÉRET, L., BACIU, M., SEGEBARTH, C. Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images. *Proceed. 5th Speech Production Seminar, München, Germany.* 2000, pp.261-264.

BADIN, P., POUCHOY, L., BAILLY, G., RAYBAUDI, M., SEGEBARTH, C., LEBAS, J.F., TIEDE, M.K., VATIKIOTIS-BATESON, E., TOHKURA, Y. Un modèle articulatoire tridimensionnel du conduit vocal basé sur des données IRM. *Actes des 22èmes Journées d'Etude sur la Parole, Martigny, Suíça.* 1998, pp. 283-286.

BADIN, P., SERRURIER, A. Three-dimensional Modeling of Speech Organs: Articulatory Data and Models, *IEICE Technical Committee on Speech, Kanazawa, Japan.* 2006, pp (In print).

BAER, T., GORE, J.C., GRACCO, L.C., NYE, P.W. Analysis of Vocal Tract Shape and Dimensions using Magnetic Resonance Imaging: Vowels. *J. Acoust. Soc. Am.* 1991, 90, pp. 799-828.

BARBOSA, J.M. *Introdução ao estudo da Fonologia e Morfologia do Português.* Coimbra: Livraria Almedina, 1994.

BARROS, M.J. *Estudo Comparativo e Técnicas de Geração de Sinal para a Síntese da Fala.* Porto: Faculdade de Engenharia do Porto, 2002. Tese de Mestrado.

BEHRENDTS J., WISMULLER A. A Segmentation and Analysis Method for MRI data of the Human Vocal Tract, *FIPKM-37.* 2001, pp. 179-189.

DANG, J., HONDA, K. An improved vocal tract model of vowel production implementing piriform fossa resonance and transvelar nasal coupling. *ICSLP, Philadelphia, USA.* 1996, pp. 965-968.

DEMOLIN D., LECUIT V., METENS T., NAZARIAN B., SOQUET A. Magnetic Resonance Measurements of the Velum Port Opening. Proceed. 5th International Conference on Spoken Language Processing (ICSLP), Sydney, Australia. 1998, pp. 425-429.

DEMOLIN, D., GEORGE, M., LECUIT, V., METENS, T., SOQUET, A., RAYMAEKERS, H. Coarticulation and articulatory compensation studied by dynamic MRI. Proceed. 5th Eurospeech 97 Conference, Rhodos, Greece. 1997, pp. 31-34.

DEMOLIN, D., HASSID, S., METENS, T., SOQUET, A. Real-time MRI and articulatory coordination in speech. C R Biol. 2002, 325(4), pp. 547-56.

DEMOLIN, D., METENS, T., SOQUET, A. Real time MRI and articulatory coordinations in vowels. Proceed.. 5 th Speech Production Seminar, München, Germany. 2000, pp. 86-93.

DEMOLIN, D., METENS, T., SOQUET, A. Three-dimensional Measurement of the Vocal Tract by MRI. Proceed. 4th International Conference on Spoken Language Processing (ICSLP), Philadelphia, USA. 1996, pp. 272-275.

DENES, P.B., PINSON, E.N. *The Speech Chain – The Physics and Biology of Spoken Language*. 2<sup>nd</sup> Edition. New York: W. H. Freeman and Company, 1993.

DI GIROLAMO, M., CORSETTI, A., LAGHI, A., FERONE, E., IANNICELLI, E., ROSSI, M., PAVONE, P., PASSARIELLO, R. Assessment with magnetic resonance of laryngeal and oropharyngeal movements during phonation. Radiol Med (Torino). 1996, 92 (1-2), pp. 33-40.

ENGWALL O. Are static MRI representative of dynamic speech? Results from a comparative study using MRI, EPG and EMA. Proceed. of 6th International Conference on Spoken Language Processing (ICSLP), Beijing, China. 2000a, pp. 17-20.

ENGWALL, O. A 3D Tongue Model based on MRI data. Proceed. of 6th International Conference on Spoken Language Processing (ICSLP), Beijing, China. 2000b, pp. 901-904.

ENGWALL, O. A revisit to the Application of MRI to the Analysis of Speech Production - Testing our assumptions. Proceed. 6th Int. Seminar on Speech Production, Sydney. 2003.

ENGWALL, O. From real-time MRI to 3D tongue movements. Proceed. of ICSLP, Jeju Island, Korea. 2004, Vol. II: 1109-1112., pp. 4-8.

ENGWALL, O. Making the Tongue Model Talk: Merging MRI & EMA Measurements. Proceed. 7th EuroSpeech Conference, Aalborg, Germany. 2001a, pp. 261-264.

ENGWALL, O. Synthesising static vowels and dynamic sounds using a 3D vocal tract model. Proceed. of 4th ISCA Tutorial and Research Workshop on Speech Synthesis. 2001b, pp. 81-86.

ENGWALL, O. Vocal tract modeling in 3D. *TMH-QPSR1-2*. 1999, pp. 31-38.

ENGWALL, O., BADIN, P. An MRI study of Swedish fricatives: Coarticulatory effects. Proceed. 5th Seminar On Speech Production: Models And Data, München, Germany. 2000, pp. 297-300.

ETTEMA, S.L., KUEHN, D.P., PERLMAN, A.L., ALPERIN, N. Magnetic resonance imaging of the levator veli palatini muscle during speech. *Cleft Palate Craniofacial Journal*. 2002, 39(2), pp. 130-44.

FANT, G. (1960). *The Acoustic Theory of Speech Production*. Mouton, The Hague.

FARIA, I.H., PEDRO, E.R., DUARTE, I., GOUVEIA, C.A.M. *Introdução à Linguística Geral e Portuguesa*. Lisboa: Caminho, 1996.

GARDNER, E., GRAY, D.J., O'RAHILLY, R. *Anatomia*. 4ª Edição. Rio de Janeiro: Guanabara Koogan, 1988.

HOOLE, P., NGUYEN, N. Electromagnetic articulography in coarticulation research. In: W.H Hardcastle and N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques*, Cambridge University Press. 1999, pp. 260-269.

KANE, A.A., BUTMAN, J.A., MULLICK, R., SKOPEC, M., CHOYKE, P. A new method for the study of velopharyngeal function using gated magnetic resonance imaging. *Plast Reconstr Surg*. 2002, 109(2), pp. 472-81.

KENNETH T., INGO T., HOFFMAN E., STORY B. Three-dimensional vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness. *The Journal of the Acoustical Society of America*. 2001, Vol. 109, Issue 2, pp. 742-747.

KRÖGER, B.J., WINKLER, R., MOOSHAMMER, C., POMPINO-MARSCHALL, B. Estimation of Vocal Tract Area Function from Magnetic Resonance Imaging: Preliminary Results. Proceed. 5th Seminar On Speech Production: Models And Data, München, Germany. 2000, pp. 333-336.

LUFKIN, R.B. *Manual de Ressonância Magnética*. 2ª Edição. Rio de Janeiro: Guanabara Koogan, 1999.

MÁDY, K., SADER, R., ZIMMERMANN, A., HOOLE, P., BEER, A., ZEILHOFER, H., HANNIG, C. Assessment of Consonant Articulation in Glossectomee Speech by Dynamic MRI. Proceed. of 7th International Conference on Spoken Language Processing (ICSLP), Denver, USA. 2002.

MÁDY, K., SADER, R., ZIMMERMANN, A., HOOLE, P., BEER, A., ZEILHOFER, H., HANNIG, C. Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction. Proceed. 4th International Speech Motor Conference, Nijmegen, Netherlands. 2001, pp.142-145.

MALMBERG, B. Tradução de Oliveira Figueiredo. *A Fonética – no Mundo dos Sons da Linguagem*. Lisboa: Edição Livros do Brasil, 1954.

MARTINS, M.R. *Ouvir falar – Introdução à Fonética do Português*. 3ª Edição. Lisboa: Editorial Caminho, 1988.

MATEUS, M.H., ANDRADE, A., VIANA, M.C., VILLALVA, A. *Fonética, Fonologia e Morfologia do Português*. Lisboa: Universidade Aberta, 1990.

MATHIAK, K., KLOSE, U., ACKERMANN, H., HERTRICH, I., KINCSES, W.E., GRODD, W. Stroboscopic articulography using fast magnetic resonance imaging, *Internat. J. of Lang. and Comm. Disorders*. 2000, 35(3), pp.419-425.

MOHAMMAD, M., MOORE, E., CARTERR, J.N., SHADLE, C., GUNN, S.J. Using MRI to image the moving vocal tract during speech. Proceed. 5th Eurospeech 97 Conference, Rhodes, Greece. 1997, (4), pp. 2027-2030.

NARAYANAN, S., NAYAK, K., LEE, S., SETHY, A., BYRD, D. An Approach to Real-time Magnetic Resonance Imaging for Speech Production. *Journal Acoustical Society of America*. 2004, 115(4), pp. 1771-76.

ROWDEN, C. *Speech Processing*. London: McGraw-Hill, 1992.

SERRURIER, A., BADIN, P. A Three-dimensional Linear Articulatory Model of Velum based on MRI data. *InterSpeech*, Lisbon. 2005, pp. 2161-2163.

SHADLE, C.H., MOHAMMAD, M., CARTER, J.N., JACKSON, P.J.B. Multi-planar dynamic magnetic resonance imaging: new tools for speech research. Proceed. 14th International Congress of Phonetic Sciences (ICPhS), São Francisco, USA. 1999, pp. 623-626.

SOQUET, A., LECUIT, V., METENS, T., DEMOLIN, D. Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI. *Speech Communication*. 2002, 36, pp.169-180.

SOQUET, A., LECUIT, V., METENS, T., NAZARIAN, B., DEMOLIN, D. Segmentation of the Airway from the Surrounding Tissues on Magnetic Resonance Images: A Comparative Study. *Proceed. ICSLP, Sydney*. 1998, pp. 3083-3086.

STONE, M., DAVIS, E.P., DOUGLAS, A.S., NESSAIVER, M., GULLAPALLI, R., LEVINE, W.S., LUNDBERG, A. Modeling the Motion of the Internal Tongue from tagged Cine-MRI images. *J. Acoust. Soc. Am.* 2001, 109(6), pp. 2974-2982.

TAKEMOTO, H., HONDA, K. Measurement of Temporal Changes in Vocal Tract Area Function during a continuous vowel sequence using a 3D Cine-MRI Technique. *Proceed. 6th Int. Seminar on Speech Production, Sydney*. 2003, pp. 284-289.

TEIXEIRA, A., VAZ, F., MARTINHO, L., COIMBRA, R.L. Articulatory Synthesis of Portuguese. *III Encontro do Forum Internacional de Investigadores Portugueses, Institut EETA, Aveiro, Portugal*. 2001.

WRENCH, A.A., MCINTOSH, A.D., HARDCASTLE, W.J. Optopalatograph (OPG): a new apparatus for speech production analysis. *Conference on Spoken Language Processing in Philadelphia, USA*. 1996, Vol.3, pp. 1589-1592.

## **ANEXOS**

**ANEXO A. INTERNATIONAL PHONETIC ALPHABET**



**ANEXO B. SPEECH ASSESSMENT METHODS PHONETIC ALPHABET**

## PORTUGUESE

### Consonants

	Symbol	Word	Transcription
plosives	p	pai	paj
	b	barco	"barku
	t	tenho	"teJu
	d	doce	"dos@
	k	com	ko~
	g	grande	"gr6nd@
fricatives	f	falo	"falu
	v	verde	"verd@
	s	céu	sEw
	z	casa	"kaz6
	S	chapéu	S6"pEw
	Z	jóia	"ZOj6
nasals	m	mar	mar
	n	nada	"nad6
	J	vinho	"viJu
liquids	l	lanche	"l6nS@
	L	trabalho	tr6"baLu
	r	caro	"karu
	R	rua	"Ru6
Vowels and diphthongs	i	vinte	"vint@
		lápiz	"lapiS
	e	fazer	f6"zer
	E	belo	"bElu
	a	falo	"falu
	6	cama	"k6m6
		madeira	m6"d6jr6
	O	ontem	"Ont6~j~
	o	lobo	"lobu
	u	jus	ZuS
		futuro	fu"turu
	@	felizes	f@"liz@S
	i~	fim	fi~
	e~	emprego	e~"pregu (or em-)
	6~	irmã	ir"m6~
	o~	bom	bo~
	u~	um	u~
	aw	mau	maw etc.: iw, ew, Ew, (ow)
	aj	mais	majS etc.: ej, Ej, Oj, oj,
	6~j~	têm	t6~j~ etc.: e~j~, o~j~, u~j~

SAMPA home page, disponível em: <http://www.phon.ucl.ac.uk/home/sampa/portug.htm>.

## ANEXO C. PROTOCOLO DE AVALIAÇÃO OROFACIAL



## Protocolo de Avaliação Orofacial Folha de Registro

NOME \_\_\_\_\_ Idade \_\_\_\_ A \_\_\_\_ M

DATA \_\_\_\_\_

TERAPEUTA DA FALA \_\_\_\_\_

SISTEMA DE REGISTO:



Adequado

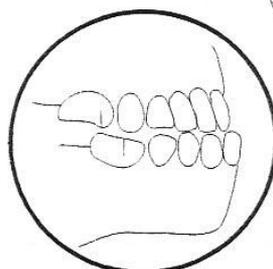
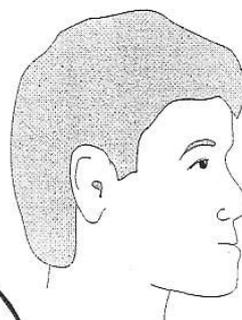
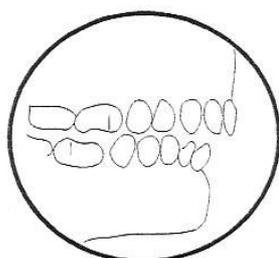
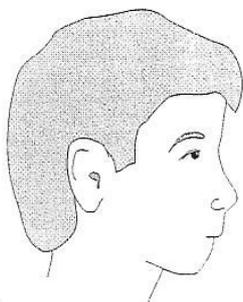
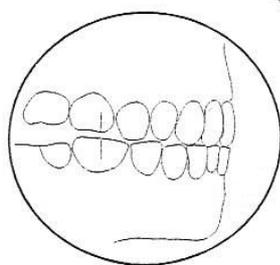
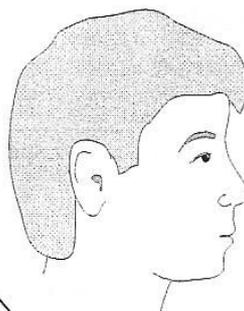
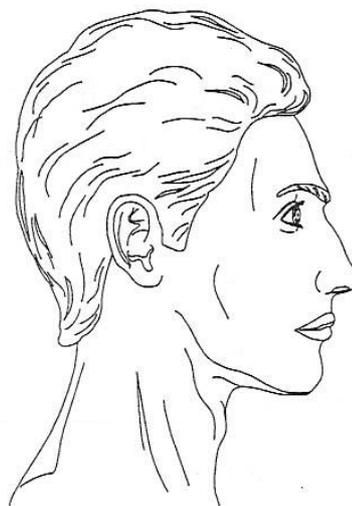
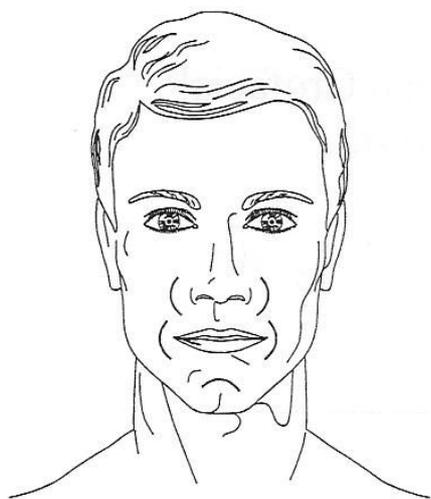


Alterado

### COTAÇÃO FINAL

	Esperada	Cotação Final Obtida
Morfologia	24 <small>consultar pág 15</small>	
Função	44 <small>consultar pág 15</small>	
Diadococinésia	<small>consultar pág. 16</small>	

OBSERVAÇÕES/RECOMENDAÇÕES





## FACE

### Morfologia

<b>Frontal</b>	Ortognatia	<input type="checkbox"/>	Endognatia	<input type="checkbox"/>	Exognatia	<input type="checkbox"/>
<b>Perfil</b>	Ortognatia	<input type="checkbox"/>	Retrognatia	<input type="checkbox"/>	Prognatia	<input type="checkbox"/>
<b>Anomalias</b> (tipo e localização) _____						
_____						

### Postura em Repouso

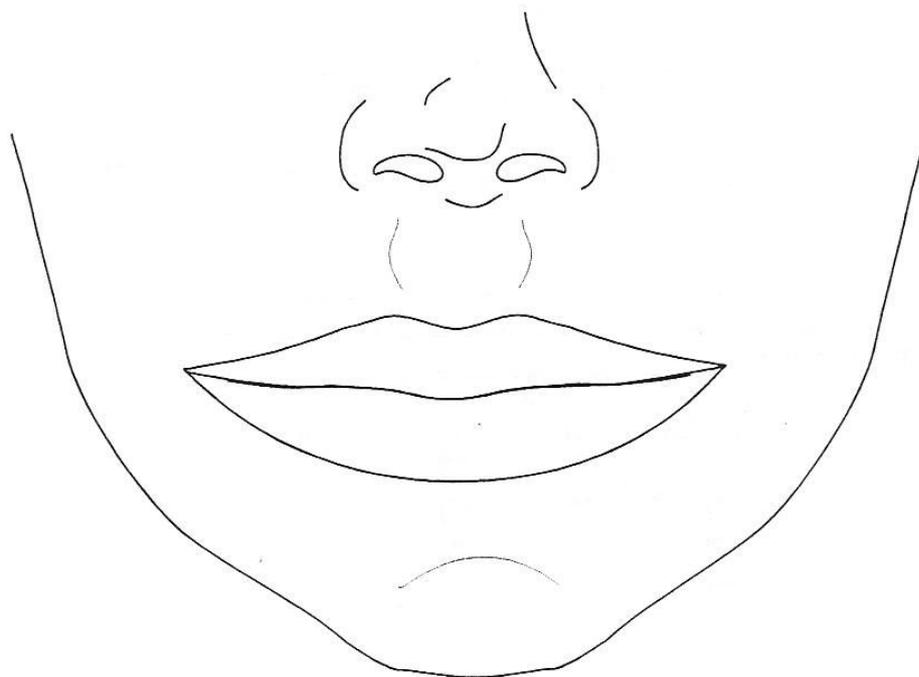
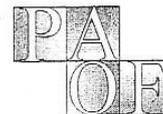
	Desnivelamento			Observações
	Simétrico	Esq	Dto	
Rugas/Testa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Sobranceiras	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Sulco Nasolabial	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Sulco Labiomarginal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Ausentes		Presentes	
Movimentos involuntários	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Praxia Facial

	É visível movimento adequado a olho nu	É visível movimento a olho nu com alterações	Não é visível movimento a olho nu	Observações -amplitude -simetria -movimentos involuntários e associados
Enrugar a TESTA				
Fechar os OLHOS				
Franzir o NARIZ				

### Tónus

Adequado	Alterado	Observações
<input type="checkbox"/>	<input type="checkbox"/>	



## LÁBIOS

### Morfologia

	Adequado	Alterado	Anomalias (tipo e localização)
Lábio Superior	<input type="checkbox"/>	<input type="checkbox"/>	
Freio (láb. superior)	<input type="checkbox"/>	<input type="checkbox"/>	
Lábio Inferior	<input type="checkbox"/>	<input type="checkbox"/>	



### Postura em Repouso

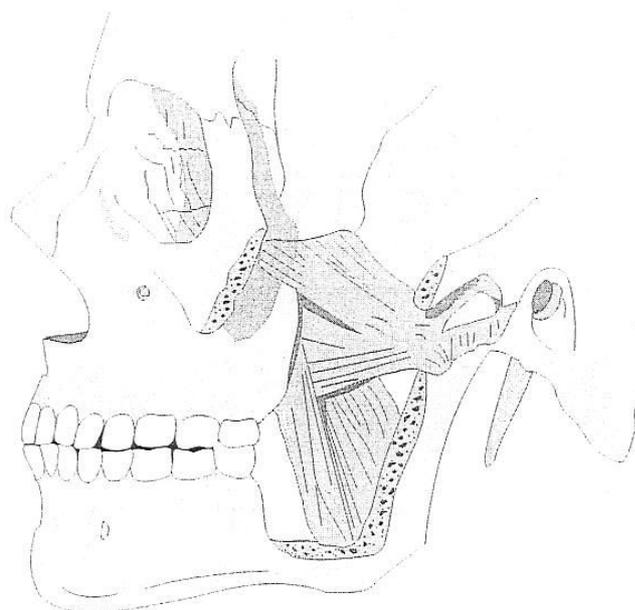
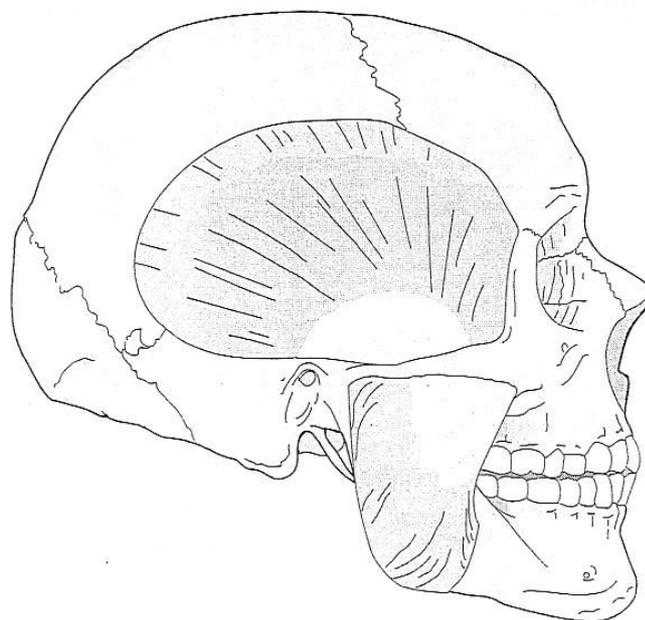
Adequada	Alterada	Observações
<input type="checkbox"/>	<input type="checkbox"/>	

### Praxia Labial

	É visível movimento adequado a olho nu	É visível movimento a olho nu com alterações	Não é visível movimento a olho nu	Observações -amplitude -simetria -movimentos involuntários e associados
Protrusão				
Retracção				
Estiramento				
Lateralização Esquerda				
Lateralização Direita				

### Tónus

Adequado	Alterado	Observações
<input type="checkbox"/>	<input type="checkbox"/>	





### Postura em repouso

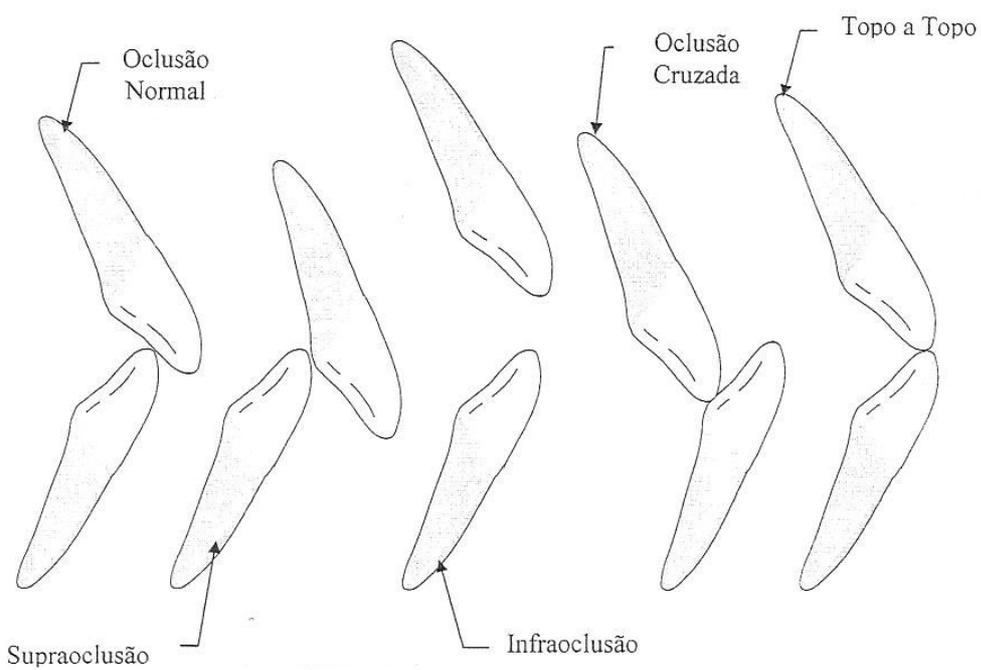
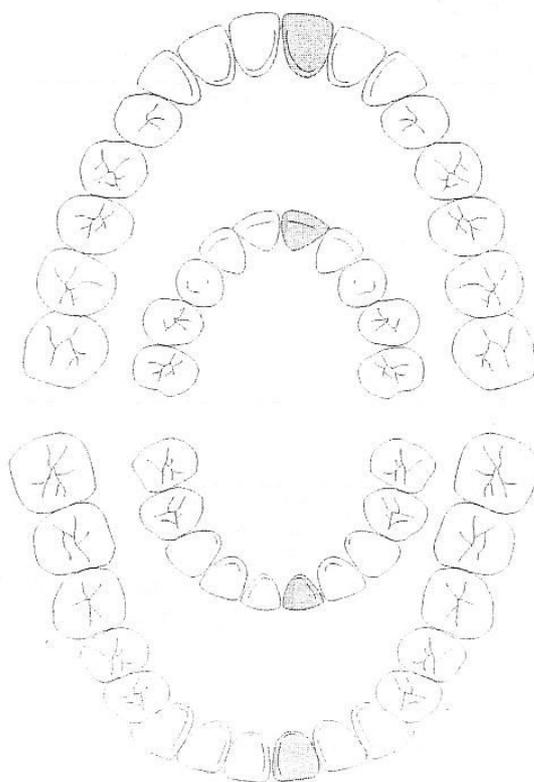
Adequada	Alterada	Observações
<input type="checkbox"/>	<input type="checkbox"/>	

### Praxia Mandibular

	É visível movimento adequado a olho nu	É visível movimento a olho nu com alterações	Não é visível movimento a olho nu	Observações -amplitude -simetria -movimentos involuntários e associados
Depressão (Abertura)				
Elevação (Encerramento)				
Protrusão				
Retracção				
Lateralização Esquerda				
Lateralização Direita				

### Tónus

Adequado	Alterado	Observações
<input type="checkbox"/>	<input type="checkbox"/>	





## DENTIÇÃO

### Implantação Dentária

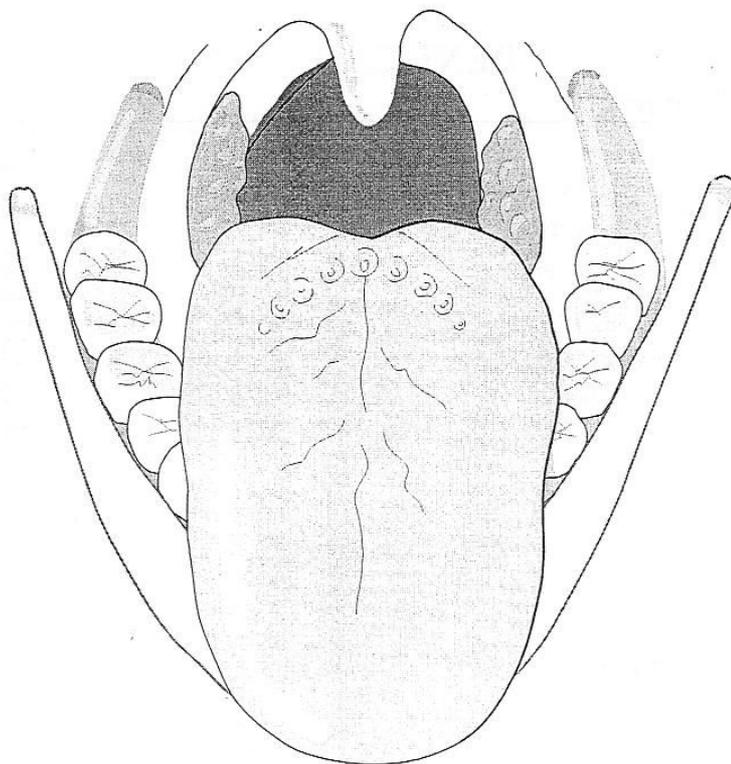
	SUP	INF
Regular	<input type="checkbox"/>	<input type="checkbox"/>
Labioversão	<input type="checkbox"/>	<input type="checkbox"/>
Línguoversão	<input type="checkbox"/>	<input type="checkbox"/>
Rotação	<input type="checkbox"/>	<input type="checkbox"/>
Sobreposições	<input type="checkbox"/>	<input type="checkbox"/>
Diastemas	<input type="checkbox"/>	<input type="checkbox"/>
Ausência de dentes	<input type="checkbox"/>	<input type="checkbox"/>
Observações:		

### Coloração das Gengivas

	SUP	INF
Rosada	<input type="checkbox"/>	<input type="checkbox"/>
Avermelhada	<input type="checkbox"/>	<input type="checkbox"/>
Hialina	<input type="checkbox"/>	<input type="checkbox"/>
Cianosada	<input type="checkbox"/>	<input type="checkbox"/>
Anomalias (tipo/localização)		

## OCLUSÃO

	Anterior	Esq	Dto
Normal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supraoclusão	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Infraoclusão	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cruzada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Topo a topo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Observações:			



## LÍNGUA

### Morfologia

<b>Coloração</b>	Rosada <input type="checkbox"/>	Avermelhada <input type="checkbox"/>	Hialina <input type="checkbox"/>	Cianosada <input type="checkbox"/>
<b>Tamanho</b>	Adequado <input type="checkbox"/>	Macro <input type="checkbox"/>	Micro <input type="checkbox"/>	
<b>Comprimento</b>	Adequado <input type="checkbox"/>	Curta <input type="checkbox"/>	Comprida <input type="checkbox"/>	
<b>Freio</b>	Adequado <input type="checkbox"/>	Curto <input type="checkbox"/>		
<b>Anomalias</b> (tipo e localização)				



### Postura em repouso

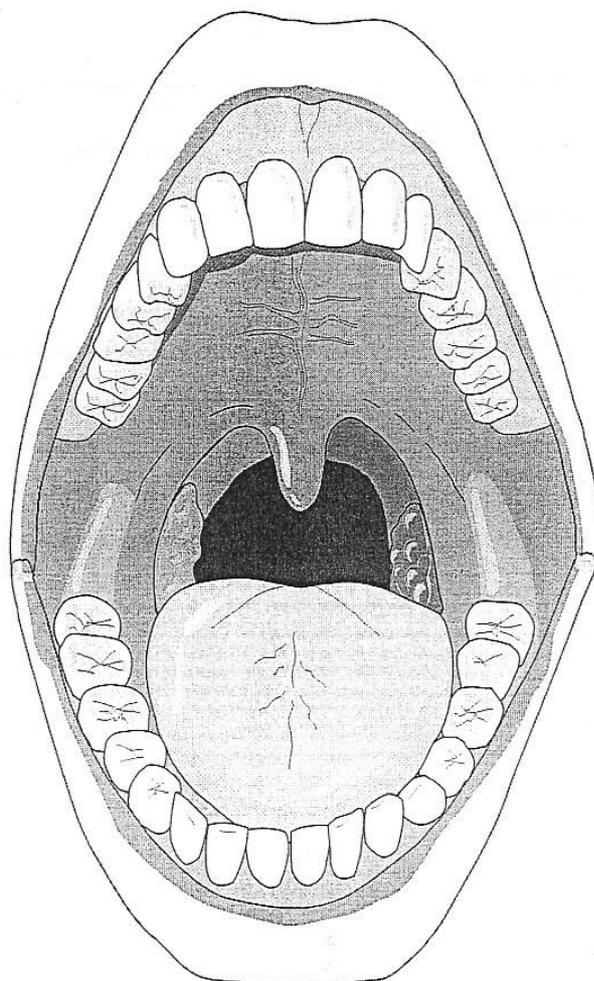
Adequada	<input type="checkbox"/>	Alterada	<input type="checkbox"/>	Observações
Movimentos involuntários		Ausentes	Presentes	
		<input type="checkbox"/>	<input type="checkbox"/>	

### Praxia Lingual

	É visível movimento adequado a olho nu	É visível movimento a olho nu com alterações	Não é visível movimento a olho nu	Observações -amplitude -simetria -movimentos involuntários e associados
Protrusão				
Retracção				
Supraversão interna				
Supraversão externa				
Infraversão interna				
Infraversão externa				
Lateralização interna Esq				
Lateralização interna Dta				
Lateralização externa Esq				
Lateralização externa Dta				

### Tónus

Adequado	Alterado	Observações
<input type="checkbox"/>	<input type="checkbox"/>	



## PALATO DURO

### Morfologia

<b>Coloração</b>	Rosada <input type="checkbox"/>	Avermelhada <input type="checkbox"/>	Hialina <input type="checkbox"/>	Cianosada <input type="checkbox"/>
<b>Rebordo Alveolar</b>	Adequado <input type="checkbox"/>	Largo <input type="checkbox"/>	Ausente <input type="checkbox"/>	
<b>Papila Incisiva</b>	Adequada <input type="checkbox"/>	Atrésica <input type="checkbox"/>		
<b>Abóboda Palatina</b>	Adequada <input type="checkbox"/>	Ogival <input type="checkbox"/>	Aplanada <input type="checkbox"/>	
<b>Anomalias</b> (tipo e localização)				



## PALATO MOLE

### Morfologia

<b>Coloração</b>	Rosada <input type="checkbox"/>	Avermelhada <input type="checkbox"/>	Hialina <input type="checkbox"/>	Cianosada <input type="checkbox"/>
<b>Úvula</b>	Adequada <input type="checkbox"/>	Ausente <input type="checkbox"/>		
	Curta <input type="checkbox"/>	Longa <input type="checkbox"/>	Bífida <input type="checkbox"/>	
<b>Amígdalas</b>			Esq <input type="checkbox"/>	Dto <input type="checkbox"/>
	Adequadas		<input type="checkbox"/>	<input type="checkbox"/>
	Ausentes		<input type="checkbox"/>	<input type="checkbox"/>
<b>Anomalias</b> (tipo e localização)	Hipertrofiadas		<input type="checkbox"/>	<input type="checkbox"/>

### Postura em repouso

<b>Simetria</b>		Adequada <input type="checkbox"/>	Desnívelamento Esq <input type="checkbox"/>	Dto <input type="checkbox"/>
	dos arcos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	da úvula	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Movimentos involuntários</b>	Ausentes <input type="checkbox"/>	Presentes <input type="checkbox"/>	Observações	

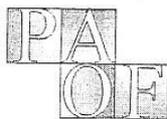
### Praxia Velofaríngea

	É visível movimento adequado a olho nu	É visível movimento a olho nu com alterações	Não é visível movimento a olho nu	Observações -amplitude -simetria -movimentos involuntários e associados
Elevação do palato mole com posteriorização				
Contração dos pilares de fauces com redução do istmo de fauces				



## DIADOCOCINÉSIA

	Nº de repetições	Tempo Obtido em segundos	Observações Ritmo/ Precisão /Velocidade
pλ/pλ/pλ	16		
tλ/tλ/tλ	16		
kλ/kλ/kλ	16		
pλtə/pλtə	12		
pλtəkə/pλtəkə	8		



## COTAÇÃO FINAL

### Morfologia

	COTAÇÃO	
	Esperada	Obtida
Morfologia da face - frontal e perfil	2	
Morfologia do lábio superior, freio do lábio superior e morfologia do lábio inferior	3	
Morfologia da língua (coloração, tamanho, comprimento e freio)	4	
Dentição (implantação superior e inferior, coloração da gengiva superior e inferior)	4	
Oclusão dentária (anterior, lateral esquerda e lateral direita)	3	
Morfologia do palato duro (coloração, rebordo alveolar, papila incisiva, abóboda palatina)	4	
Morfologia do palato mole (coloração, úvula, amígdala esquerda e amígdala direita)	4	
TOTAL	24	

### Função

	COTAÇÃO	
	Esperada	Obtida
Postura facial em repouso (simetria das rugas da testa, sobrancelhas, sulco nasolabial, sulco labiomarginal e ausência de movimentos involuntários)	5	
Praxia facial (3 movimentos)	3	
Tónus facial adequado	1	
Postura labial em repouso (adequada, linha de encerramento adequada, simetria adequada)	3	
Praxia labial (5 movimentos)	5	
Tónus labial adequado	1	
Postura mandibular em repouso adequada	1	
Praxia mandibular (6 movimentos)	6	
Tónus mandibular adequado	1	
Postura lingual em repouso (adequada e ausência de movimentos involuntários)	2	
Praxia lingual (10 movimentos)	10	
Tónus lingual adequado	1	
Postura do palato mole em repouso (simetria dos arcos, da úvula e ausência de movimentos involuntários)	3	
Praxia do esfíncter velofaríngeo (2 movimentos)	2	
TOTAL	44	



## COTAÇÃO FINAL

### Diadococinésia

	Nº de repetições	Tempo Obtido em seg	Número de repetições/ Tempo Obtido em segundos = Número de repetições por segundo
pΛ/pΛ/pΛ	16		
tΛ/tΛ/tΛ	16		
kΛ/kΛ/kΛ	16		
pΛtə/pΛtə	12		
pΛtəkə/pΛtəkə	8		

Valores Normais Esperados - consultar guia de interpretação do PAOF.

**ANEXO D. FORMULÁRIO DE CONSENTIMENTO INFORMADO**

## FORMULÁRIO DE CONSENTIMENTO INFORMADO

Conforme a “Declaração de Helsíquia” da Associação Médica mundial  
(Helsínquia 1694; Tóquio 1975; Veneza 1983; Hong Kong 1989; Somerset West 1996; Edimburgo 2000)

**DESIGNAÇÃO DO ESTUDO:** Estudo Morfológico-dinâmico do Tracto Vocal Humano

**Investigadora:** Sandra Moreira Rua

**E-mail:** [smr@estsp.ipp.pt](mailto:smr@estsp.ipp.pt)

**Orientador:** Professor Diamantino Freitas (DEEC – FEUP)

**Âmbito:** Mestrado de Engenharia Biomédica

### Introdução

Foi solicitado para participar no estudo acima mencionado pois possui conhecimentos na área da Terapia da Fala, facilitando o processo de recolhas imagiológicas do tracto vocal, dado que já possui treino na pronúncia correcta dos diferentes sons.

### A sua participação é voluntária

Este formulário de consentimento informado fornece-lhe as informações necessárias sobre o estudo que serão discutidas consigo. Assim que tiver compreendido o estudo e concordar participar no mesmo, solicita-se que assine o formulário de consentimento. Ser-lhe-á fornecida uma cópia para seu uso pessoal.

Antes de ser informado a respeito do estudo, é importante que tome conhecimento dos seguintes aspectos:

- A sua participação é inteiramente voluntária;
- Poderá decidir não participar ou cancelar a sua participação neste estudo, em qualquer altura.

### Objectivos do estudo

Inicialmente, serão efectuadas recolhas imagiológicas por Ressonância Magnética do tracto vocal, para caracterizar a articulação da fala (informação morfológica), estudando estaticamente as vogais e semivogais do Português bem como algumas consoantes (plosivas e fricativas);

Seguidamente, será realizada a caracterização dinâmica da fala, tanto quanto possível em tempo-real, tendo em conta:

- Áreas seccionais (para estimar a forma da laringe, cavidade nasal e oral);
- Forma da língua, véu do palatino e suas posições;
- Relação entre os diferentes articuladores;
- Limites do tracto vocal nas suas diferentes porções;

Farão parte deste estudo cerca de 5 pessoas, jovens, voluntárias e com formação na área da Terapia da Fala ou previamente treinadas para o efeito.

### Procedimentos

Se resolver participar neste estudo e assinar o respectivo formulário de consentimento, posteriormente ser-lhe-á pedido que preencha um questionário para pesquisa de eventuais contra-indicações para a execução de exames por Ressonância Magnética. Todas as recolhas serão supervisionadas pelo Médico Radiologista envolvido no estudo.

Durante as recolhas:

- Será instruído devidamente de modo a facilitar a sua colaboração;
- Ser-lhe-á solicitado que pronuncie diferentes sons, em sequências devidamente estruturadas;
- Deverá permanecer imóvel durante o exame, tanto quanto possível;
- Não será administrado qualquer tipo de fármaco.

### Riscos e/ou desconforto

A Ressonância Magnética (RM) é um método de imagem, que permite a obtenção de imagens multiplanares de qualquer parte do corpo humano, e que não utiliza radiação ionizante. A imagem por RM tem por base a interacção de um campo magnético e pulsos de radiofrequência com os núcleos do corpo humano.

Não existem estudos científicos que comprovem a existência de danos biológicos resultantes da exposição a campos magnéticos como o utilizado em RM.

Os pulsos de radiofrequência aplicados causam variações de temperatura, devido ao depósito de energia nos tecidos. Será, por isso, respeitado o limite de absorção ( $\leq 0,4$  W/kg) pelos tecidos bem como o aumento da temperatura corporal não será superior a 1°C.

Durante o exame verificam-se níveis de ruído acústico variáveis produzidos pelo equipamento de RM.

Dada a utilização de um campo magnético, existe a possibilidade de interferência eléctrica, risco potencial de projecteis, torção de objectos, logo serão tomadas todas as normas de segurança necessárias.

A necessidade de permanecer imóvel durante as recolhas e o próprio posicionamento poderão causar algum desconforto.

### Benefícios previstos

Dada a escassez de informação morfológico-dinâmica da fala, será possível através da sua participação reunir a informação necessária para a caracterização dos sons da fala. O conhecimento desta informação permitirá uma melhor compreensão dos mecanismos subjacentes às disfunções da fala, assim como possibilitará o desenvolvimento de algoritmos para síntese e reconhecimento da fala.

### Custos

Este estudo não implicará qualquer custo para si.

### Sigilo

Será garantida a confidencialidade dos seus dados pessoais:

- A sua identificação será efectuada através de um código;
- Não será identificado pessoalmente em nenhuma publicação referente a este estudo.

### Declaração de Consentimento

Eu, abaixo-assinado, ..... (*nome completo*),  
compreendi a informação que me foi fornecida acerca do estudo em que irei participar, tendo me sido dada a oportunidade de fazer as perguntas que julguei necessárias.

Por isso, consinto participar no estudo respondendo a todas as questões propostas e permitindo a execução das recolhas imagiológicas por Ressonância Magnética do tracto vocal.

Porto, ..... de ..... de 200.....

\_\_\_\_\_  
(Assinatura)

Pelo Investigador Responsável,

\_\_\_\_\_

**ANEXO E. FOLHA DE REGISTO**

**FOLHA DE REGISTO**

N.º indivíduo _____
Ensaio n.º _____

**DESIGNAÇÃO DO ESTUDO:** Estudo Morfológico-dinâmico do Tracto Vocal Humano**Investigadora:** Sandra Moreira Rua**E-mail:** [smr@estsp.ipp.pt](mailto:smr@estsp.ipp.pt)**Orientador:** Professor Diamantino Freitas (DEEC – FEUP)**Âmbito:** Mestrado de Engenharia Biomédica**Caracterização da amostra:**

Sexo: M \_\_\_\_ F \_\_\_\_

Data exame: \_\_\_\_/\_\_\_\_/\_\_\_\_

Idade: \_\_\_\_ anos

Peso: \_\_\_\_ Kg

**Duração média de sustentação dos sons** \_\_\_\_\_ segundos**Características do equipamento:** Magnete de \_\_\_\_ Tesla**Tipo de estudo:** \_\_\_\_ Estático \_\_\_\_ Dinâmico \_\_\_\_ Tempo-real**Sequência utilizada/ ponderações:** FSE WT1**Protocolo:**

TR - ms

TE - ms

ETL -

N.º cortes -

Orientação -

FOV -

BW -

NEX -

ETL -

Matriz -

Outros:

**SAR -****Procedimentos e sons pronunciados:**

1

**Outras observações:****Principais dificuldades:**

**ANEXO F. PROTOCOLO DO ESTUDO ESTÁTICO**

## PROTOCOLO ESTUDO ESTÁTICO

**DESIGNAÇÃO DO ESTUDO:** Estudo Morfológico-dinâmico do Tracto Vocal Humano

**Investigadora:** Sandra Moreira Rua

**E-mail:** [smr@estsp.jpp.pt](mailto:smr@estsp.jpp.pt)

**Orientador:** Professor Diamantino Freitas (DEEC – FEUP)

**Âmbito:** Mestrado de Engenharia Biomédica

**Médica Radiologista responsável:** Professora Doutora Isabel Ramos

1º Localizer cerebral nos três planos a zero (0,0,0) do isocentro do magnete (centro da antena) para marcação dos cortes.

2º Estudo anatómico no plano sagital com FOV de 150mm à região em estudo.

**Parâmetros:**

Sequência Turbo Spin-eco – imagem ponderada em T1

TR – 443ms; TE – 15ms; ETL – 7; FA 150°;

Cortes: **1** com **5 mm de espessura;** Matriz – 128x128;

**Parâmetros dos cortes sagitais:** Turbo Spin-eco – imagens ponderadas em T1

TR – 443ms; TE – 17ms; ETL – 7; FA 150°;

Cortes: **3** com **5mm de espessura** sem intervalo entre cortes

FOV – 150mm (FOV AP de 100%); Matriz – 128x128; SNR – 1

**Tempo de aquisição/ som – 9s**

**Parâmetros dos cortes coronais:** Turbo Spin-eco – imagens ponderadas em T1

TR – 470ms; TE – 15ms; ETL – 7; FA 150°;

Cortes: **4 cortes espaçados 16mm entre si** com **6mm de espessura cada**

**Tempo de aquisição/ som – 9,9s**

FOV – 150mm (FOV AP de 100%); Matriz – 128x128; SNR – 1,03

**ANEXO G. PROTOCOLO DO ESTUDO DINÂMICO**

## PROTOCOLO ESTUDO DINÂMICO

**DESIGNAÇÃO DO ESTUDO:** Estudo Morfológico-dinâmico do Tracto Vocal Humano

**Investigadora:** Sandra Moreira Rua

**E-mail:** [smr@estsp.ipp.pt](mailto:smr@estsp.ipp.pt)

**Orientador:** Professor Diamantino Freitas (DEEC – FEUP)

**Âmbito:** Mestrado de Engenharia Biomédica

**Médica Radiologista responsável:** Professora Doutora Isabel Ramos

Sincronização:

- Trigger cardíaco com monitorização por ECG;
- A aquisição corresponde ao intervalo R-R, emitindo um sinal sonoro “bip” no pico da onda R;
- Sincronização do sujeito com o sinal sonoro;
- Solicita-se após verificação do sincronismo (pela audição através do intercomunicador e visualização do ECG) que o sujeito inicie a pronúncia do som, sem exageros de articulação ou prolongamento no tempo de articulação, respirando de forma calma e normal.

**Corpus:**

1. /tu/ 2. /ma/ 3. /pa/

**Protocolo:**

1º Localizer cerebral nos três planos a zero (0,0,0) do isocentro.

2º Estudo cine-RM com os seguintes parâmetros de aquisição:

Sequência: Flash 2D

TR – 60ms / 86ms (de acordo com o intervalo R-R e frequência cardíaca de cada sujeito)

TE – 4,4ms

FA – 30°

NEX – 1

Matriz – 208 x 256

Cod. Fase – AP

**1 corte sagital com 6mm de espessura**

FOV – 300mm

Transformada parcial de Fourier – 60%

**Parâmetros fisiológicos: ECG**

Trigger delay – 100ms

**Tempo de aquisição total das repetições: variável de acordo com a duração do ciclo cardíaco.**

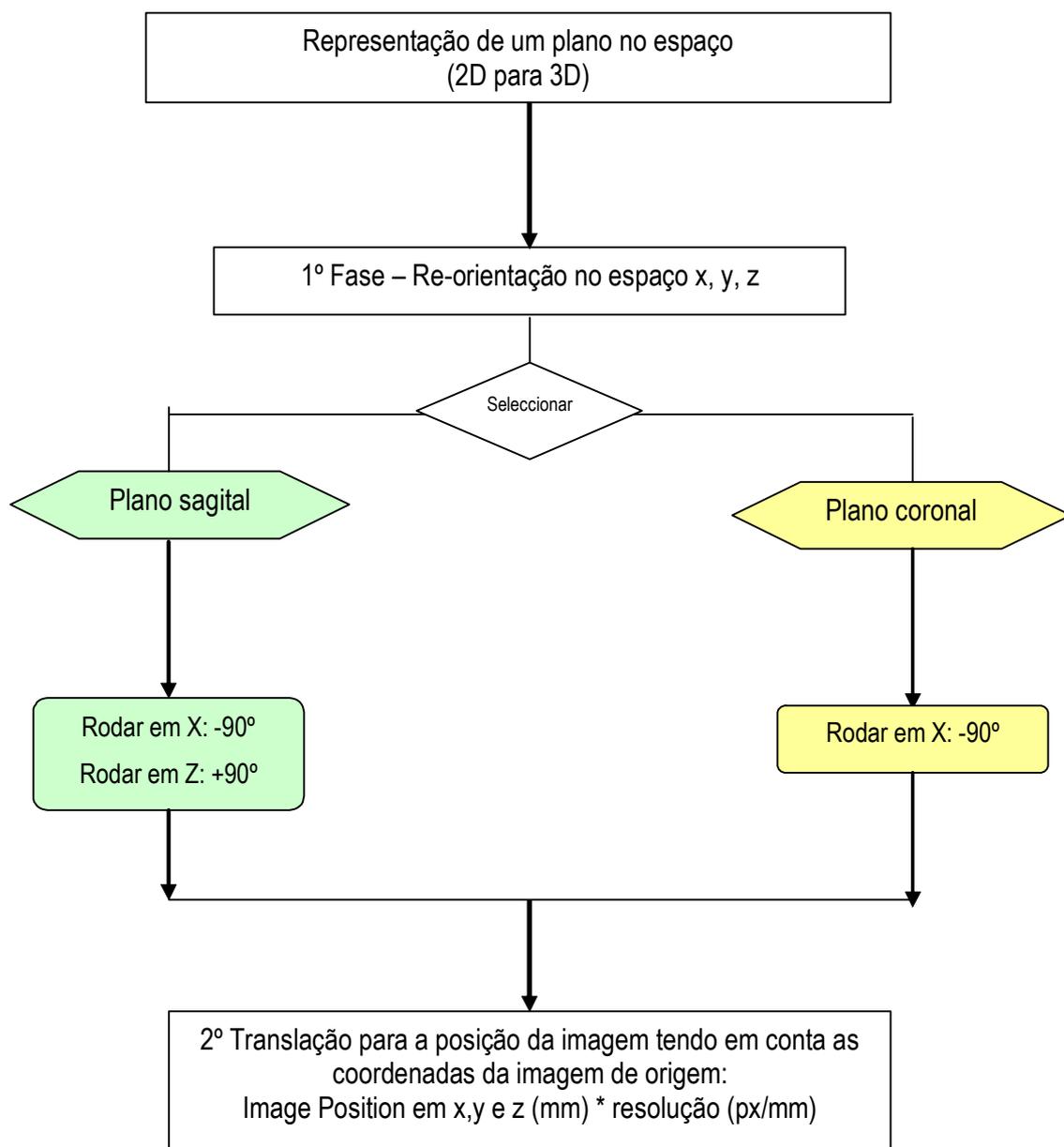
N.º de fases – determina o n.º imagens a adquirir e depende do intervalo R-R

**ANEXO H. CD COM ESTUDO ESTÁTICO E DINÂMICO**

Este anexo contém um CD com os seguintes ficheiros:

- Est\_estático\_Suj\_1: com os modelos tridimensionais dos 25 sons do Português provenientes do estudo estático, para o sujeito do sexo masculino. Modelos em formato PNG.
- Est\_estático\_Suj\_2: com os modelos tridimensionais dos 25 sons do Português provenientes do estudo estático, para o sujeito do sexo feminino, bem como um modelo 3D da vogal U com um varrimento completo de todo o tracto vocal no plano sagital. Modelos em formato PNG.
- Estudo dinâmico: contém quatro pastas separadas para cada sujeito e respectivo sexo, com os seguintes ficheiros "Stack\_MA", "Stack\_PA" e "Stack\_TU". Ficheiros de vídeo que podem ser visualizados com Windows Media Player ou com o QuickTime Player.

**ANEXO I. FLUXOGRAMA CORRESPONDENTE ÀS ETAPAS DE  
COMBINAÇÃO E REPRESENTAÇÃO TRIDIMENSIONAL**



**ANEXO J. RELATÓRIO DE TRABALHOS PRÁTICOS DO INVESTIGADOR LUÍS JACINTO DO MEB**

## “Em Direcção a Uma Prótese Laringe – Produção Acústica do Sinal de Fala e Seus Componentes”

Relatório da Disciplina de Trabalhos Práticos do Mestrado em Engenharia Biomédica da Faculdade de Engenharia da Universidade do Porto, 2006

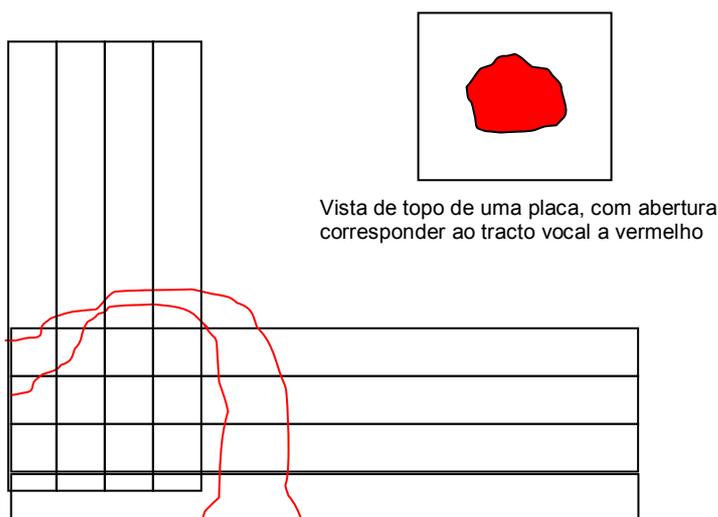
Autor: LUIS JACINTO

### OBJECTIVO

Tendo por base a imagiologia de ressonância magnética recolhida referente à conformação do tracto vocal para a vogal /i/, este trabalho teve como objectivo construir um molde tridimensional físico do tracto vocal que fosse capaz de reproduzir, com o máximo de fidelidade possível, as suas características anatómicas conformativas e acústicas, para posteriores ensaios de fonte-filtro.

### PROCEDIMENTO

Impressões em tamanho real das imagens de ressonância magnética, dos cortes sagitais e coronais para um mesmo indivíduo vocalizando a vogal /i/, foram decalcadas em placas de poliuretano com uma espessura igual à de cada corte. Em cada placa foi recortada a forma do tracto vocal dando lugar a uma abertura. As placas foram depois sobrepostas e alinhadas, com base em pontos comuns a cada imagem, de modo a formar uma abertura contínua.



Vista sagital do alinhamento das placas, para posterior construção do molde físico real do tracto vocal para a vogal /i/.

A cavidade contínua formada pela sobreposição das placas foi preenchida por resina de poliuretano com o objectivo de produzir um bloco sólido com a forma do tracto vocal. Este bloco foi moldado, pelo uso de pasta de moldar e lixa, para suavizar as passagens abruptas entre as placas. Posteriormente foi colocado numa estrutura fechada, a qual se preencheu com borracha de silicone de reticulação por condensação. Após retirar o bloco sólido de resina do interior do silicone estava formado um tubo sólido, oco por dentro, e com a conformação do tracto vocal.

## **RESULTADO**

O resultado obtido foi deveras satisfatório, aproximando-se a conformação do molde à conformação esperada do tracto vogal para a vogal /i/, pela observação das imagens de RM e outros estudos do mesmo âmbito. Contudo convém referir que vários erros poderão ter sido introduzidos nomeadamente: no corte em espessura das placas, no corte das aberturas correspondentes ao tracto vocal para cada placa, no alinhamento das placas, no processo de termoconformação da resina de poliuretano por reacção química com o poliuretano das placas, na moldagem das passagens abruptas do bloco sólido de resina de poliuretano, e no corte do bloco de silicone para remoção do bloco de resina de poliuretano do seu interior.

**ANEXO K. BIBLIOGRAFIA**

- BEHREND, J. Vocal Tract Recovering using MRI. Oral presentation in Face, Speech and Acoustics, 2002.
- BEOLCHI, L., FACCHINETTI, S. *Telemedicine Glossary*. 5<sup>th</sup> Edition. Brussels: European Commission, 2003.
- CRARY, M.A., KOTZUR, I.M., GAUGER, J., GORHAM, M., BURTON, S. Dynamic magnetic resonance imaging in the study of vocal tract configuration. *Journal of Voice*. 1996, 10 (4), pp.378-388.
- DANG, J., HONDA, K. Estimation of vocal tract shape from sounds via a physiological articulatory model. *J. Phonetics*. 2002, 30, pp. 511–532.
- ENGWALL, O. Combining MRI, EMA and EPG measurements in a three-dimensional tongue model. *Speech Communication*. 2003, 41(2-3), pp. 303-329.
- ENGWALL, O. A 3D vocal tract model for articulatory and visual speech synthesis. *Proceed. Fonetik 98, the Swedish Phonetics Conference*. 1998, pp. 196-199.
- GUIMARÃES, I. *Protocolo de Avaliação Orofacial*. Lisboa: Europaxis, 1995.
- HASKINS LABORATORIES. Measuring and Modeling Speech Production. Disponível em: <http://www.haskins.yale.edu/featured/heads/production.html>.
- HONDA, K. Evolution of Vowel Production Studies and Observation Techniques. *Acoust Sci. & Tech*. 2002, 23 (4), pp. 189-194.
- HONDA, M. Human Speech Production Mechanisms. NTT Technical Review. Disponível em: <http://www.ntt.co.jp/tr/0305/files/ntr200305024.pdf>.
- HONDA, M. Study of Speech Communication Mechanisms based on the Understanding of Human Biological Functions. Disponível em: <http://www.nict.go.jp/publication/shuppan/kihou-journal/journal-vol48no3/toku3-3.pdf>.
- HOOLE, P., WISMULLER, A., LEINSINGER, G., KROOS, C., GEUMANN, A., INOUE, M. Analysis of tongue configuration in multi-speaker, multi-volume MRI data. *Proceed. of the 5th Seminar on Speech Production: Models and Data & CREST - Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling, May 1st to 4th, Kloster Seeon, Bavaria, Germany*. 2000.
- MEDURI, S., BAZZOCCHI, M., ZUIANI, C., FALCONE, B., BERTINO, G., MARIONI, G. Functional MR with use of FLASH sequences in the evaluation of the phono-articulatory tract. Institute of Radiology, University of Udine, Italy. *MAGMA*. 1999, 9 (1-2), pp. 5-15.
- METENS, T., SOQUET, A., DEMOLIN, D. Imagerie par Résonance Magnétique. Disponível em: <http://www.ulb.ac.be/philo/phonolab/mri.html>.
- NELSON, N.W. *Childhood Language Disorders in Context*. New York: Merrill, 1993.

NIETO-CASTANON, A., GUENTHER, F.H. Constructing Speaker-Specific Articulatory Vocal Tract Models for Testing Speech Motor Control Hypotheses. ICPH, San Francisco. 1999, pp. 2271-74.

PARSONS, T. *Voice and Speech Processing*. New York: McGraw-Hill, 1987.

PRITCHARD, D. Vocal Tract Visualisation. EECE593 Winter Project Report, April 9, 2002.

STORY, B.H. Using imaging and modeling techniques to understand the relation between vocal tract shape and acoustic characteristics, Proceed. of the Stockholm Music Acoustics Conference. 2003.

TIFFONY, W.R.; CARRELL, J. *Phonetics, Theory and Application*. 2<sup>nd</sup> Edition. New York: McGraw-Hill, 1987.

WELLS, J.C. SAMPA computer readable phonetic alphabet. In Gibbon, D., Moore, R. and Winski, R. (eds.), 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B.